

UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO

JOSÉ LUÍS VIEIRA SOBRINHO

**Abordagem Não-Linear de Redução de
Dimensionalidade em Duas Etapas Aplicada à
Classificação de Engajamento em Redes Sociais**

GOIÂNIA

2025



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

JOSÉ LUÍS VIEIRA SOBRINHO

3. Título do trabalho

“Abordagem Não-Linear de Redução de Dimensionalidade em Duas Etapas Aplicada à Classificação de Engajamento em Redes Sociais”

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **José Luís Vieira Sobrinho, Discente**, em 28/10/2025, às 08:59, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Flavio Henrique Teles Vieira, Professor do Magistério Superior**, em 29/10/2025, às 23:29, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5738121** e o código CRC **7BDB1260**.

JOSÉ LUÍS VIEIRA SOBRINHO

**Abordagem Não-Linear de Redução de Dimensionalidade
em Duas Etapas Aplicada à Classificação de Engajamento
em Redes Sociais**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação, da Escola de Engenharia Elétrica, Mecânica e de Computação, da Universidade Federal de Goiás (UFG), como requisito para a obtenção do Título de Doutor em Engenharia Elétrica e de Computação.

Área de Concentração: Engenharia Elétrica.
Linha de Pesquisa: Sistemas Inteligentes e Computação Aplicada

Orientador: Professor Doutor Flávio Henrique Teles Vieira

Coorientador: Professor Doutor Alisson Assis Cardoso

GOIÂNIA

2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Sobrinho, José Luís Vieira

Abordagem Não-Linear de Redução de Dimensionalidade em Duas Etapas Aplicada à Classificação de Engajamento em Redes Sociais [manuscrito] / José Luís Vieira Sobrinho. - 2025.

150 f. : il.

Orientador: Prof. Dr. Flávio Henrique Teles Vieira; co-orientador Dr. Alisson Assis Cardoso.

Tese (Doutorado) - Universidade Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de Computação (EMC), Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Goiânia, 2025. Bibliografia.

1. Aprendizado de Máquina. 2. Classificação. 3. Otimização. I. Vieira, Flávio Henrique Teles, orient. II. Título.

CDU 621.3



UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

ATA DE DEFESA DE TESE

Ata Nº 3 da sessão de Defesa de Tese de **JOSÉ LUÍS VIEIRA SOBRINHO** que confere o título de Doutor em **Engenharia Elétrica e de Computação**, na área de concentração em **Engenharia Elétrica**.

Aos **quatorze dias do mês de outubro de dois mil e vinte e cinco**, a partir das **14h**, realizou-se a sessão pública de Defesa de Tese intitulada “**Abordagem Não-Linear de Redução de Dimensionalidade em Duas Etapas Aplicada à Classificação de Engajamento em Redes Sociais**”. Os trabalhos foram instalados pelo Orientador, Professor Doutor **Flávio Henrique Teles Vieira - (EMC/UFG)** Orientador, Professor Doutor **Alisson Assis Cardoso - (EMC/UFG)** coorientador, com a participação dos demais membros da Banca Examinadora: Professor Doutor **Cássio Dener Noronha Vinhal - (EMC/UFG)** Membro Titular Externo, Professor Doutor **Leonardo da Cunha Brito - (EMC/UFG)** Membro Titular Interno, Professor Doutor **Wesley Pacheco Calixto - (EMC/UFG)** Membro Titular Interno e **Prof. Dr. Carlos Galvão Pinheiro Júnior - (EMC/UFG)** Membro Titular Externo: **cuja participação ocorreu através de videoconferência**: <https://meet.google.com/fmf-tqmp-pza>. Durante a argüição os membros da banca **não fizeram** sugestão de alteração do título do **trabalho**. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor Flávio Henrique Teles Vieira, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, Aos quatorze dias do mês de outubro de dois mil e vinte e cinco.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Flavio Henrique Teles Vieira, Professor do Magistério Superior**, em 14/10/2025, às 17:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Alisson Assis Cardoso, Professor do Magistério Superior**, em 14/10/2025, às 17:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cassio Dener Noronha Vinhal, Professor do Magistério Superior**, em 14/10/2025, às 17:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Galvao Pinheiro Junior, Professor do Magistério Superior**, em 14/10/2025, às 17:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leonardo Da Cunha Brito, Professor do Magistério Superior**, em 14/10/2025, às 17:17, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **José Luís Vieira Sobrinho, Discente**, em 14/10/2025, às 17:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Wesley Pacheco Calixto, Usuário Externo**, em 15/10/2025, às 10:48, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5710499** e o código CRC **6620484B**.

Referência: Processo nº 23070.051455/2025-21

SEI nº 5710499

*À memória do meu pai, Oswaldo César Alves da Silva.
Mesmo que o tempo tenha nos separado fisicamente,
sei que você estaria orgulhoso deste momento.
Carrego comigo seus ensinamentos e dedico a você,
com todo meu amor, esta conquista que também é sua.*

Agradecimentos

Primeiramente, agradeço à minha família, base fundamental de todas as minhas conquistas. Seu apoio e dedicação constantes foram essenciais para que mais esta realização se tornasse possível.

Agradeço também ao sistema público de educação superior brasileiro e a cada um de seus integrantes. As melhores oportunidades que tive ao longo dos anos surgiram da minha trajetória acadêmica, e por isso sou profundamente grato.

Meu reconhecimento especial vai para todos os colegas que compartilharam comigo as salas de aula da Universidade Federal de Goiás ao longo dos últimos quinze anos. Foi uma honra, da graduação ao doutorado, fazer parte desta comunidade acadêmica.

Por fim, agradeço ao Prof. Dr. Flávio Henrique Teles Vieira, meu orientador. Sua dedicação e seu comprometimento genuíno com o sucesso de seus alunos são não apenas exemplares mas também amplamente reconhecidos por seus orientandos.

“Se vi mais longe, foi por estar sobre ombros de gigantes”.

Isaac Newton

Resumo

Esta tese apresenta uma metodologia de classificação de dados intitulada NL-TSDR (*Nonlinear Two-Stage Dimensionality Reduction*), cujo objetivo é atenuar os desafios apresentados por conjuntos de dados de alta dimensionalidade e seus inerentes problemas de mascaramento de classes. Este fenômeno ocorre quando a sobreposição de instâncias de diferentes grupos leva à degradação do desempenho dos métodos de classificação tradicionais. O método proposto é composto por duas etapas distintas. Na primeira etapa, o objetivo é extrair características discriminativas maximizando a probabilidade de separação de classes, mitigando efetivamente a sobreposição de classes. A segunda etapa aplica transformações não lineares (funções tangente hiperbólica, gaussiana, sigmoide e laplaciana) para aprimorar a separabilidade dos centros de classes por meio de otimização multiobjetivo que maximiza as distâncias interclasse enquanto minimiza a dispersão intraclasse. A validação experimental foi realizada em conjuntos variados de dados tradicionalmente utilizados na literatura e dados reais de engajamento de redes sociais. O método apresenta melhorias consistentes em todos os cenários avaliados em relação às abordagens consideradas nesta tese, alcançando geralmente acurácias superiores mesmo com significativa redução dimensional. Especificamente, no conjunto de dados de redes sociais, o método alcança melhoria de acurácia de 3,04% em relação aos métodos comparados, simultaneamente obtendo um novo conjunto de dados cinco vezes e meia menor que o original. A análise de variância (ANOVA) com $p = 0,0034$, corroborada pelos testes t pareado ($p = 0,0199$) e de Wilcoxon ($p = 0,0153$), indica diferença estatística significativa ($p \leq 0,05$) entre as acurácias de classificação dos dados originais e dos dados processados pelo método proposto. Os resultados evidenciam que o NL-TSDR fornece uma ferramenta eficaz para abordar o mascaramento de classes em tarefas de classificação de alta dimensionalidade, mantendo eficiência computacional adequada para aplicações do mundo real.

Palavras-chave: Aprendizado de Máquina, Classificação, Otimização.

Abstract

This thesis presents a data classification methodology entitled NL-TSDR (Nonlinear Two-Stage Dimensionality Reduction), aimed at mitigating the challenges posed by high-dimensional datasets and their inherent class masking problems. This phenomenon occurs when overlapping instances from different groups lead to the degradation of traditional classification methods' performance. The proposed method comprises two distinct stages. In the first stage, the objective is to extract discriminative features by maximizing pairwise class separation probability, effectively mitigating class overlap. The second stage applies nonlinear transformations (hyperbolic tangent, Gaussian, sigmoid, and Laplacian functions) to enhance class center separability through multi-objective optimization that maximizes inter-class distances while minimizing intra-class dispersion. Experimental validation was conducted on benchmark datasets and real social media engagement data. The method demonstrates consistent improvements across all evaluated scenarios, achieving superior accuracies even with significant dimensional reduction. Specifically, on the social media dataset, the method achieves a 3.04% accuracy improvement over baseline methods, simultaneously obtaining a dataset that is five and a half times smaller than the original. The analysis of variance (ANOVA) with $p = 0.0034$, corroborated by paired t -test ($p = 0.0199$) and Wilcoxon test ($p = 0.0153$), indicates a statistically significant difference ($p \leq 0.05$) between the classification accuracies of the original data and the data processed by the proposed method. The results demonstrate that NL-TSDR provides an effective tool for addressing class masking in high-dimensional classification tasks while maintaining computational efficiency suitable for real-world applications.

Keywords: Machine Learning, Classification, Optimization.

Lista de ilustrações

Figura 1.1 – Relação entre indicadores linguísticos e decisões de financiamento	21
Figura 1.2 – Redes sociais mais populares no mundo em fevereiro de 2025	28
Figura 1.3 – Número de usuários do Instagram no mundo de 2020 a 2025	29
Figura 1.4 – Distribuição dos usuários do Instagram por idade em abril de 2025	29
Figura 1.5 – Países com maior audiência no Instagram em fevereiro de 2025	30
Figura 2.1 – Os três paradigmas fundamentais do aprendizado de máquina	38
Figura 2.2 – Diferença entre métodos determinísticos e não determinísticos	39
Figura 2.3 – Evolução das fronteiras de decisão em problemas de classificação	42
Figura 2.4 – Representação gráfica do <i>kernel trick</i>	45
Figura 2.5 – Representação esquemática de um neurônio artificial	58
Figura 2.6 – Representação esquemática de um MLP	58
Figura 2.7 – Exemplo do Coeficiente de Correlação de Pearson	63
Figura 4.1 – Dispersão das classes por características x_1 e x_2 ($\rho = -0,4380$).	91
Figura 4.2 – Dispersão das classes por características x_1 e x_3 ($\rho = -0,3158$).	91
Figura 4.3 – Dispersão das classes por características x_1 e x_4 ($\rho = -0,2883$).	92
Figura 4.4 – Dispersão das classes por características x_1 e x_5 ($\rho = -0,2162$).	92
Figura 4.5 – Dispersão das classes por características x_1 e x_2 ($\rho = -0,4237$).	93
Figura 4.6 – Dispersão das classes por características x_2 e x_3 ($\rho = 0,9070$).	93
Figura 4.7 – Dispersão das classes por características x_2 e x_4 ($\rho = 0,9788$).	94
Figura 4.8 – Dispersão das classes por características x_2 e x_5 ($\rho = 0,9650$).	94
Figura 4.9 – Dispersão das classes por características x_2 e x_6 ($\rho = 0,9876$).	95
Figura 4.10–Dispersão das classes por características x_3 e x_4 ($\rho = 0,8927$).	95
Figura 4.11–Dispersão das classes por características x_3 e x_5 ($\rho = 0,9004$).	96
Figura 4.12–Dispersão das classes por características x_3 e x_6 ($\rho = 0,8892$).	96
Figura 4.13–Dispersão das classes por características x_4 e x_5 ($\rho = 0,9950$).	97
Figura 4.14–Dispersão das classes por características x_4 e x_6 ($\rho = 0,9832$).	97
Figura 4.15–Dispersão das classes por características x_5 e x_6 ($\rho = 0,9678$).	98
Figura 4.16–Dispersão das classes por características x_1 e x_2 ($d = 4$ e $\rho = 0,9505$).	99
Figura 4.17–Dispersão das classes por características x_1 e x_3 ($d = 4$ e $\rho = 0,1465$).	99
Figura 4.18–Dispersão das classes por características x_1 e x_4 ($d = 4$ e $\rho = 0,3884$).	100
Figura 4.19–Dispersão das classes por características x_2 e x_3 ($d = 4$ e $\rho = 0,2727$).	100
Figura 4.20–Dispersão das classes por características x_2 e x_4 ($d = 4$ e $\rho = 0,2820$).	101
Figura 4.21–Dispersão das classes por características x_1 e x_2 ($d = 4$ e $\rho = -0,4650$).	101
Figura 4.22–Coeficiente de <i>Silhouette</i> para o <code>crab_dataset</code>	102
Figura 4.23–Classificação do subespaço obtido pelo TSDR.	102
Figura 4.24–Diagrama metodologia experimental	109

Figura 5.1 – Progressão da acurácia da classificação através das etapas	112
Figura 5.2 – Progressão do coeficiente de correlação de Pearson	116
Figura 5.3 – Progressão do coeficiente de <i>Silhouette</i>	116
Figura 5.4 – Relação entre redução de dimensionalidade e melhoria de acurácia . . .	118
Figura 5.5 – Melhorias no desempenho demonstrando melhoria simultânea	119
Figura 5.6 – Comparação da acurácia da classificação entre transformações	119
Figura 5.7 – Melhorias no Coeficiente de <i>Silhouette</i> por transformação	120
Figura 5.8 – Melhorias no Coeficiente de Correlação de Pearson por transformação .	120

Lista de tabelas

Tabela 1.1 – Comparação das abordagens regulatórias para proteção de dados . . .	33
Tabela 1.2 – Comparação das abordagens regulatórias para inteligência artificial . .	34
Tabela 2.1 – Comparação entre tipos de aprendizado	39
Tabela 2.2 – Comparação entre métodos determinísticos e não determinísticos . . .	41
Tabela 2.3 – Comparação entre abordagens lineares e não lineares em classificação .	43
Tabela 2.4 – Comparação entre algoritmos de otimização	49
Tabela 2.5 – Comparação entre os métodos de classificação de referência	51
Tabela 2.6 – Comparação das métricas de qualidade dos dados	61
Tabela 2.7 – Comparação das métricas de desempenho de classificação	61
Tabela 2.8 – Interpretação dos valores do Coeficiente de <i>Silhouette</i>	65
Tabela 3.1 – Comparação entre transformações não lineares	82
Tabela 4.1 – Características dos conjuntos de dados de referência	105
Tabela 5.1 – Acurácias de classificação (%) em conjuntos de dados de referência . .	111
Tabela 5.2 – F1 <i>Scores</i> de classificação (%) em conjuntos de dados de referência . .	114
Tabela 5.3 – Avaliação da redução de dimensionalidade e métricas de qualidade . . .	115
Tabela 5.4 – Tempo de execução (segundos) em conjuntos de dados de referência . .	117
Tabela 5.5 – Análise estatística do método NL-TSDR	122

Lista de abreviaturas e siglas

ANOVA	<i>Analysis of Variance</i>
AUC-ROC	<i>Area Under the Receiver Operating Characteristic Curve</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
CDF	<i>Cumulative Distribution Function</i>
DR	<i>Dimensionality Reduction</i>
GA	<i>Genetic Algorithms</i>
GAN	<i>Generative Adversarial Network</i>
GDPR	<i>General Data Protection Regulation</i>
GNN	<i>Graph Neural Network</i>
GPT	<i>Generative Pre-trained Transformer</i>
IoT	<i>Internet of Things</i>
<i>k</i> -NN	<i>k-Nearest Neighbors</i>
LDA	<i>Linear Discriminant Analysis</i>
MLP	<i>Multilayer Perceptron</i>
MPM	<i>Minimax Probability Machine</i>
MPSP	<i>Maximum Pairwise Separation Probability</i>
NLP	<i>Natural Language Processing</i>
PCA	<i>Principal Component Analysis</i>
PHATE	<i>Potential of Heat-diffusion for Affinity-based Transition Embedding</i>
RMSE	<i>Root Mean Squared Error</i>
RNA	<i>Rede Neural Artificial</i>
SME	<i>Simplex de Nelder-Mead</i>
SQP	<i>Sequential Quadratic Programming</i>

SVM	<i>Support Vector Machine</i>
<i>t</i> -SNE	<i>t-Distributed Stochastic Neighbor Embedding</i>
UMAP	<i>Uniform Manifold Approximation and Projection</i>
VAE	<i>Variational Autoencoder</i>

Lista de símbolos

ρ	<i>Coeficiente de Correlação de Pearson</i>
s	<i>Coeficiente de Silhouette</i>
κ	<i>Medida de Separação de Classes da Etapa 1</i>
α	<i>Probabilidade de Separação de Classes da Etapa 1</i>
\mathbf{W}	<i>Matriz de Transformação da Etapa 1</i>
\vec{y}^i	<i>Centros de Gravidade das Instâncias da Etapa 2</i>
v^k	<i>Dispersão das Instâncias da Etapa 2</i>
Ω	<i>Problema de Otimização da Etapa 2</i>

Sumário

Introdução	15
1 Contextualização e Fundamentação Teórica	19
2 Revisão Bibliográfica	36
2.1 Fundamentos de Aprendizado de Máquina e Classificação	36
2.1.1 Tipos de Aprendizado	37
2.1.2 Métodos Determinísticos <i>versus</i> Não Determinísticos	38
2.2 Abordagens Lineares <i>versus</i> Não Lineares em Classificação	40
2.2.1 Possíveis Limitações dos Métodos Lineares	42
2.2.2 Transformações Não Lineares e Separabilidade	44
2.2.3 Otimização Multiobjetivo em Classificação	46
2.3 Algoritmos de Otimização	47
2.4 Métodos Clássicos de Classificação	50
2.4.1 <i>Naïve Bayes</i>	50
2.4.2 <i>k-Nearest Neighbors</i>	52
2.4.3 Análise de Discriminante Linear	54
2.4.4 <i>Multilayer Perceptron</i>	56
2.5 Métricas de Avaliação	60
2.5.1 Coeficiente de Correlação de Pearson	62
2.5.2 Coeficiente de <i>Silhouette</i>	64
2.5.3 Métricas de Desempenho de Classificação	66
2.5.3.1 Acurácia	66
2.5.3.2 <i>Precision e Recall</i>	67
2.5.3.3 <i>F1 Score</i>	69
2.6 Trabalhos Relacionados	70
3 Método Proposto	75
3.1 Etapa 1 - Maximização da Probabilidade de Separação de Classes	76
3.2 Etapa 2 - Otimização de Espaço Não Linear	78
3.2.1 Funções de Transformação Não Lineares	81
3.2.1.1 Transformação Tangente Hiperbólica	82
3.2.1.2 Transformação Gaussiana	83
3.2.1.3 Transformação Sigmoide	84
3.2.1.4 Transformação Laplaciana	84
3.3 Algoritmos e Implementação	85
4 Metodologia Experimental	89
4.1 Experimento com Visualização de Dados	90
4.2 Métodos de Classificação de Referência	103

4.3	Conjuntos de Dados de Referência	104
4.4	Conjunto de Dados do Instagram	105
4.5	Processo de Avaliação e Parametrização dos Métodos	106
5	Resultados	110
6	Discussão	123
6.1	Discussão dos Resultados Experimentais e Limitações	123
6.2	Direções Futuras	128
	Conclusão	131
	 Referências	 134

Introdução

As tarefas de classificação representam uma das aplicações mais tangíveis do aprendizado de máquina (OSISANWO et al., 2017). Nesse contexto, os algoritmos aprendem a construir modelos preditivos a partir de conjuntos de dados de treinamento, permitindo posteriormente a classificação automática de novos indivíduos com base em suas características. *Naive Bayes*, *k-Nearest Neighbors* (*k*-NN) e *Multilayer Perceptron* (MLP) são exemplos de métodos bem estabelecidos para essa finalidade (NARAYANAN et al., 2017).

Contudo, um dos desafios no campo do aprendizado de máquina é a alta dimensionalidade dos conjuntos de dados, comumente observada nos conjuntos de dados da vida real (THUDUMU et al., 2020). Essa característica, que implica em um grande número de atributos de entrada, pode causar baixo desempenho nos algoritmos que processam esses dados (JIA et al., 2022). A alta dimensionalidade dos conjuntos de dados introduz complexidades computacionais e metodológicas que podem degradar o desempenho dos algoritmos tradicionais de classificação. Este fenômeno, conhecido como maldição da dimensionalidade (BERISHA et al., 2021), pode se manifestar através de diversos problemas, incluindo o aumento exponencial do espaço de busca e a alta dispersão dos dados, constituindo um dos objetos de estudo dessa linha de pesquisa.

Além do potencial desafio da alta dimensionalidade dos conjuntos de dados, os algoritmos de classificação podem enfrentar outras limitações (CHEN et al., 2024). Alguns métodos podem apresentar dificuldades para lidar com distribuição de classes desequilibrada, enquanto outros podem apresentar desempenho superior nos conjuntos de dados de treinamento comparado aos de teste (*overfitting*) (SOLTANZADEH; FEIZI-DERAKHSHI; HASHEMZADEH, 2023). Adicionalmente, alguns algoritmos podem mesclar classes que estão próximas no espaço de entrada, levando indivíduos de diferentes classes a se sobreporem (ABOU-MOUSTAFA; De La Torre; FERRIE, 2015), fenômeno conhecido como mascaramento de classe (*class masking*).

O mascaramento de classes representa um desafio relevante em tarefas de classificação de dados de alta dimensionalidade (SOUZA; LUGHOFFER, 2022). Este problema ocorre quando instâncias de diferentes classes se sobrepõem no espaço de características original, dificultando a separação entre grupos distintos para métodos tradicionais de classificação (VUTTIPITTAYAMONGKOL; ELYAN; PETROVSKI, 2021). Como consequência, pode-se observar degradação na acurácia dos modelos preditivos, comprometendo seu desempenho em aplicações práticas. A relação entre o aumento da dimensionalidade e a intensificação deste fenômeno constitui um aspecto central para a compreensão dos limites dos algoritmos de classificação (LIN; CHEN, 2012).

Diante desses desafios, considerando que conjuntos de dados reais frequentemente envolvem espaços de alta dimensão (STEINBACH; ERTÖZ; KUMAR, 2004), as técnicas de redução de dimensionalidade podem desempenhar um papel relevante para os métodos de classificação. Essas técnicas permitem extrair os atributos mais relevantes, potencialmente contribuindo para a mitigação dos problemas relacionados ao mascaramento de classes e otimizando o uso de recursos computacionais durante o processamento (VELLIANGIRI; ALAGUMUTHUKRISHNAN; Thankumar joseph, 2019).

Dessa maneira, técnicas de redução de dimensionalidade emergem como ferramentas para mitigar os possíveis efeitos adversos da alta dimensionalidade. Abordagens tradicionais como Análise de Componentes Principais ou *Principal Component Analysis* (PCA) (JOLLIFFE; CADIMA, 2016) e Análise Discriminante Linear ou *Linear Discriminant Analysis* (LDA) (XANTHOPOULOS; PARDALOS; TRAFALIS, 2013) têm sido utilizadas ao longo dos anos. Contudo, essas técnicas lineares podem se mostrar inadequadas para capturar a complexidade inerente de dados com estruturas não lineares subjacentes, o que é uma das limitações que este trabalho busca endereçar (SERANI; DIEZ, 2025).

A motivação para o desenvolvimento desta pesquisa surge pelo interesse em métodos que lidam de forma eficaz com conjuntos de dados de alta dimensionalidade em domínios onde a acurácia de classificação é um aspecto potencialmente importante, como no contexto das redes sociais. Este domínio apresenta características específicas (GAO et al., 2017) que ilustram as questões abordadas no presente trabalho.

Plataformas de redes sociais geram diariamente enormes volumes de dados multidimensionais, incluindo interações dos usuários, conteúdo textual, metadados temporais, informações demográficas, entre outros (YADAV et al., 2023). A análise de engajamento nessas plataformas requer o processamento de características de alta dimensionalidade (HEUSINGER; RAAB; SCHLEIF, 2022) que frequentemente apresentam sobreposições significativas entre diferentes categorias de usuários ou tipos de conteúdo.

Alguns métodos de redução de dimensionalidade, quando aplicados a dados complexos como os de redes sociais, podem falhar em preservar as informações discriminativas necessárias para uma classificação eficaz (ARAMBURU; BERLANGA; LANZA-CRUZ, 2024). Isso ocorre porque essas técnicas, em sua maioria lineares, podem não capturar as complexas relações não lineares presentes nos padrões de comportamento dos usuários em redes sociais (ZHAO et al., 2021).

Adicionalmente, a eventual necessidade de processamento em tempo real pode impor restrições computacionais que potencialmente inviabilizam o uso de métodos excessivamente complexos. Dessa maneira, abordagens que exigem uma grande quantidade de recursos computacionais podem ser não recomendadas nesses cenários. Portanto, existe um interesse por metodologias que sejam simultaneamente eficazes na preservação de informações discriminativas e computacionalmente eficientes (MUTAMBIK, 2024).

Neste contexto, o objetivo desta tese consiste em desenvolver e validar uma metodologia de classificação de dados que implementa redução de dimensionalidade em duas etapas, denominada NL-TSDR (*Nonlinear Two-Stage Dimensionality Reduction*). A metodologia proposta visa mitigar o problema de mascaramento de classes, com potencial contribuição para o aprimoramento da acurácia de classificação mediante a redução do espaço dimensional dos dados.

A fundamentação teórica e a revisão bibliográfica desta tese contemplam tanto as bases matemáticas de abordagens clássicas da literatura quanto suas aplicações práticas em diferentes domínios. Este levantamento visa identificar lacunas metodológicas e abordagens que se diferenciam da metodologia NL-TSDR proposta. Adicionalmente, estabelece-se um referencial teórico sobre métricas de avaliação em redução de dimensionalidade e em métodos de classificação, constituindo as bases conceituais para o desenvolvimento e validação da metodologia.

Para atingir este objetivo, investiga-se inicialmente o problema de mascaramento de classes em espaços de alta dimensionalidade, estabelecendo métricas para sua mensuração. Posteriormente, desenvolve-se a primeira etapa do método NL-TSDR, voltada à extração de características discriminativas mediante a maximização da probabilidade de separação de classes.

Em seguida, investiga-se a segunda etapa do método, que envolve a implementação de transformações não lineares utilizando funções tangente hiperbólica, gaussiana, sigmoide e laplaciana para modificar e potencialmente melhorar a separabilidade dos centros de classes. Estas transformações constituem um problema de otimização multiobjetivo que maximiza as distâncias interclasse enquanto minimiza a dispersão intraclasse no espaço transformado.

A validação experimental constitui outro objetivo específico, envolvendo a avaliação do método proposto em conjuntos de dados de referência utilizados na literatura. Complementarmente, busca-se aplicar e avaliar o desempenho do NL-TSDR em dados reais de engajamento de redes sociais, demonstrando sua aplicabilidade prática em cenários do mundo real.

O trabalho inclui a realização de análises comparativas, estabelecendo as vantagens e limitações da abordagem proposta. Por fim, pretende-se investigar a sensibilidade do método aos seus parâmetros e sugerir diretrizes práticas para sua configuração em diferentes domínios de aplicação.

Esta tese está estruturada em cinco capítulos, organizados para apresentar o desenvolvimento teórico, metodológico e experimental da pesquisa. No [Capítulo 1](#) apresenta-se a contextualização e fundamentação teórica, estabelecendo as bases conceituais do trabalho.

No [Capítulo 2](#) aborda-se uma revisão bibliográfica e análise do estado da arte em classificação e redução de dimensionalidade, incluindo técnicas de aprendizado de máquina estabelecidas e métricas de avaliação. No [Capítulo 3](#) descreve-se o método NL-TSDR, sua fundamentação matemática e os algoritmos implementados.

No [Capítulo 4](#) e no [Capítulo 5](#) apresenta-se a validação experimental realizada em conjuntos de dados de referência, especificando a metodologia experimental, métricas utilizadas e análise comparativa com métodos referenciais. Nestes capítulos também se aborda a aplicação do método em dados de engajamento de redes sociais, incluindo aspectos de pré-processamento, implementação e análise de resultados.

No [Capítulo 6](#) sintetizam-se os resultados da pesquisa, examinam-se as limitações identificadas e propõem-se direções para trabalhos futuros. As contribuições desta tese inserem-se na discussão dos métodos de redução de dimensionalidade para classificação de dados de alta dimensionalidade, apresentando fundamentos teóricos e aplicações práticas.

1 Contextualização e Fundamentação Teórica

A classificação é uma tarefa do aprendizado de máquina que consiste em categorizar um conjunto de dados em um número finito e predefinido de classes ou categorias, atribuindo rótulos específicos a cada instância (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006). Seu objetivo é construir um modelo preditivo capaz de aprender padrões e relações a partir de um conjunto de dados rotulados (dados de treinamento) para, posteriormente, categorizar novos registros cujas classes (ou categorias) são desconhecidas (dados de teste) (LEVATIĆ et al., 2024).

O processo de classificação envolve a identificação de características discriminantes nos dados que permitem distinguir os indivíduos entre as possíveis diferentes classes. Existem diferentes técnicas para resolver problemas de classificação, cada uma com suas particularidades e adequações específicas, como por exemplo as técnicas de regressão, as árvores de decisão e as redes neurais artificiais (BLOCKEEL et al., 2023). A escolha do algoritmo mais apropriado depende de diferentes fatores, tais como o tamanho do conjunto de dados, a quantidade de características, a distribuição de classes, a interpretabilidade dos dados, a oferta de recursos computacionais e a natureza do problema específico (BARBIERI; GRISCI; DORN, 2024).

Os métodos de classificação têm aplicabilidade ampla e são fundamentais para resolver problemas complexos em diferentes domínios do conhecimento e setores da economia. No setor financeiro, por exemplo, um banco pode estar interessado em desenvolver modelos para identificar quais clientes apresentam maior probabilidade de inadimplência em empréstimos, utilizando variáveis como histórico de crédito, renda, idade, profissão e comportamento transacional.

Esse mesmo banco também pode empregar técnicas de classificação para detectar transações fraudulentas, aprovação automática de cartões de crédito, ou segmentação de clientes para ofertas personalizadas. Neste caso, o processo do banco pode envolver a coleta e preparação de dados históricos, a seleção de características relevantes (*feature engineering*), o treinamento do modelo com dados rotulados e sua posterior validação e implementação.

Existem aplicações onde se utilizaram dados textuais de aplicações de empréstimo de plataformas online de *peer-to-peer lending* para desenvolver modelos preditivos (NETZER; LEMAIRE; HERZENSTEIN, 2019). Nessas abordagens, empregaram-se técnicas de processamento de linguagem natural (NLP ou *natural language processing*) e aprendizado de máquina para extrair características semânticas e linguísticas dos textos das solicitações, incluindo padrões de escrita, escolha de palavras e estrutura gramatical.

Através dessa análise textual, conseguiram identificar sinais sutis que correlacionam com a probabilidade de inadimplência, demonstrando como informações aparentemente subjetivas podem ser transformadas em preditores quantitativos úteis. No estudo, ilustra-se como a classificação probabilística pode ir além dos dados financeiros tradicionais, incorporando dimensões comportamentais e psicológicas do processo de concessão de crédito, resultando em modelos mais abrangentes para avaliação de risco de crédito.

Na plataforma analisada, os solicitantes são requisitados a elaborar uma narrativa textual detalhada explicando as razões pelas quais necessitam do empréstimo e demonstrando como farão bom uso dos recursos financeiros solicitados. Essa descrição serve como um elemento no processo de tomada de decisão dos potenciais credores, que avaliam não apenas os dados financeiros tradicionais, mas também a credibilidade e confiabilidade transmitidas através da comunicação escrita.

Os pesquisadores descobriram que a linguagem utilizada pelos solicitantes na justificativa de um empréstimo constitui um indicador preditivo que impacta a probabilidade de pagamento, frequentemente superando a relevância de métricas financeiras convencionais como classificações de crédito, histórico bancário e renda declarada. Esta descoberta expande a compreensão sobre avaliação de risco de crédito, demonstrando que elementos aparentemente subjetivos podem fornecer percepções quantitativas valiosas sobre o comportamento futuro dos devedores.

Através de análise computacional dos elementos textuais das solicitações, os pesquisadores identificaram padrões linguísticos distintivos que se correlacionam com os resultados de pagamento. Entre os termos mais significativos, cinco emergiram como indicadores positivos de adimplência: “*debt-free*” (livre de dívidas), “*lower interest rate*” (taxa de juros mais baixa), “*after-tax*” (após impostos), “*minimum payment*” (pagamento mínimo) e “*graduate*” (graduado). Em contrapartida, cinco termos mostraram correlação negativa, associando-se à inadimplência: “*God*” (Deus), “*promise*” (promessa), “*will pay*” (vou pagar), “*thank you*” (obrigado) e “*hospital*”.

As expressões relacionadas à adimplência, como “*lower interest rate*” e “*after-tax*”, sugerem um nível sofisticado de educação financeira, ou seja, que o solicitante possui conhecimento técnico sobre produtos financeiros e compreende conceitos como estruturas de juros e implicações fiscais. Quando candidatos mencionam serem “*debt-free*” ou “*graduate*”, eles fornecem evidências potenciais de conquistas pessoais e estabilidade financeira, indicadores de responsabilidade e capacidade de gestão de recursos. Por outro lado, os termos associados à inadimplência sugerem promessas potencialmente vazias.

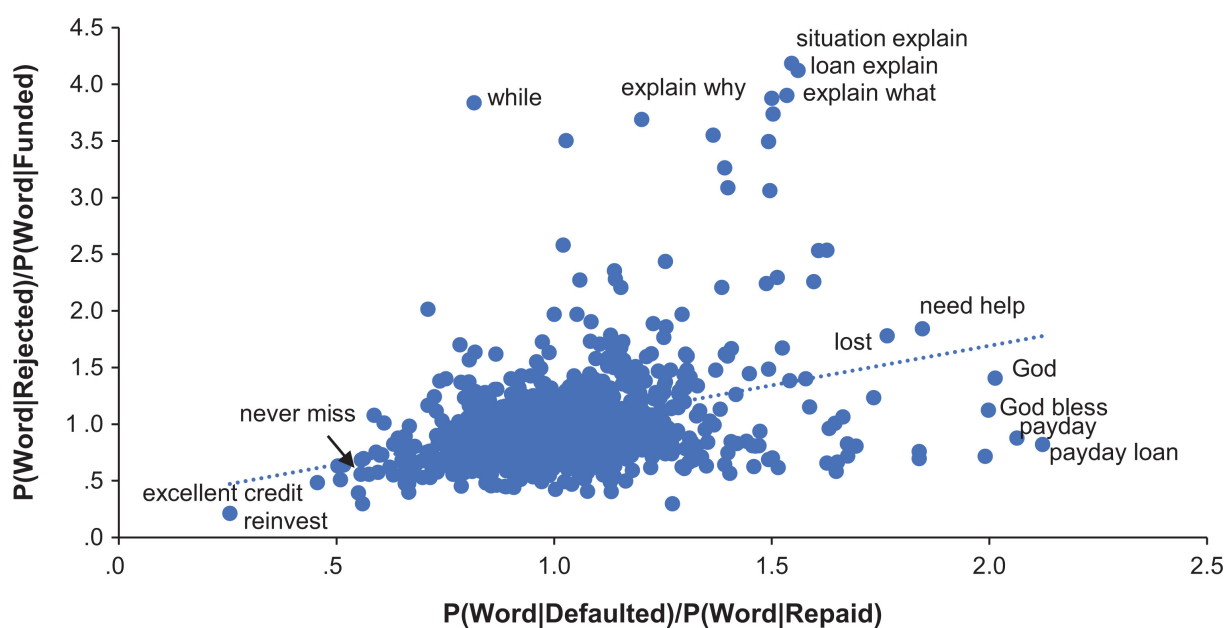
As conclusões do estudo estabelecem diretrizes para avaliação de risco baseada em análise textual: solicitantes que apresentam planos detalhados e específicos para pagamento, demonstrando compreensão dos mecanismos financeiros envolvidos, e que referenciam compromissos honrados no passado, tendem a replicar esse comportamento responsável.

Por outro lado, candidatos que recorrem a promessas genéricas, apelos emocionais ou invocações religiosas apresentam perfis de risco significativamente elevados.

Na [Figura 1.1](#), ilustra-se a relação entre a linguagem utilizada nas solicitações de empréstimo, as decisões dos credores e os resultados de pagamento. O gráfico de dispersão apresenta no eixo horizontal a razão entre a probabilidade de uma palavra aparecer em solicitações inadimplentes *versus* adimplentes, enquanto o eixo vertical mostra a razão entre a probabilidade dessa palavra aparecer em solicitações rejeitadas *versus* financiadas.

Esta visualização permite identificar três padrões: termos localizados na região inferior esquerda (como “*excellent credit*” e “*reinvest*”) associados a menor risco de inadimplência e maior probabilidade de aprovação; palavras posicionadas no extremo direito (como “*God*”, “*payday loan*” e “*need help*”) que apresentam associação com inadimplência; e termos na região superior direita (especialmente variações da palavra “*explain*”) que coincidem com rejeições pelos credores e problemas de pagamento. A linha de tendência pontilhada indica correspondência entre as decisões dos credores e os padrões de inadimplência identificados pela análise computacional.

Figura 1.1 – Relação entre indicadores linguísticos e decisões de financiamento



Fonte: (NETZER; LEMAIRE; HERZENSTEIN, 2019)

Para identificar quais termos linguísticos discriminam entre solicitações de empréstimos inadimplentes e adimplentes, empregou-se um classificador *Naive Bayes* multinomial, uma técnica de aprendizado de máquina probabilística baseada no teorema de Bayes. Apesar da limitação de assumir independência entre os preditores (PAJILA et al., 2023), esta abordagem possibilita a interpretação dos elementos textuais discriminativos.

O estudo identificou três categorias distintas de comportamento: termos que apresentam alta correlação entre baixa probabilidade de financiamento e alta probabilidade de inadimplência (demonstrando alinhamento entre percepção dos credores e realidade), palavras que exercem impacto desproporcional nas decisões de financiamento comparado ao seu real poder preditivo de inadimplência, e termos que efetivamente “enganaram” os credores, resultando em aprovações de empréstimos que posteriormente se tornaram inadimplentes. Esta última categoria representa oportunidades de melhoria para sistemas automatizados de avaliação de crédito, onde algoritmos de classificação podem identificar sinais de risco que credores humanos podem não detectar conscientemente.

Na área da saúde, algoritmos de classificação auxiliam no diagnóstico médico, identificação de doenças através de exames de imagem, e predição de resposta a tratamentos (LI et al., 2023). No marketing digital, são utilizados para classificação de sentimentos em redes sociais, segmentação de público-alvo e sistemas de recomendação (BRANDAO et al., 2025). Em todos esses diferentes domínios, a qualidade e representatividade dos dados de treinamento, bem como a seleção adequada das variáveis preditoras, são aspectos relevantes para o desempenho e confiabilidade do modelo classificador.

A crescente disponibilidade de dados em múltiplas dimensões e formatos tem expandido o escopo e a complexidade dos problemas de classificação contemporâneos (BOUTABA et al., 2018). Em diagnósticos médicos baseados em imagens, por exemplo, uma única tomografia computadorizada pode conter milhões de *pixels*, cada um representando uma característica potencial para o algoritmo de classificação (SHEN; WU; SUK, 2017). Similarmente, na análise de sentimentos em redes sociais, o vocabulário de uma plataforma pode facilmente exceder centenas de milhares de termos únicos, enquanto sistemas de recomendação em *e-commerce* frequentemente lidam com catálogos contendo milhões de produtos e bilhões de interações (ALJUNID et al., 2025).

Esta abundância de dados, embora represente uma oportunidade para criar modelos mais precisos e abrangentes, introduz simultaneamente desafios do aprendizado de máquina moderno, como a maldição da dimensionalidade (*curse of dimensionality*) (MAATEN et al., 2009). Este fenômeno surge quando o número de características ou dimensões dos dados cresce exponencialmente, criando espaços de alta dimensionalidade onde os algoritmos tradicionais de classificação podem enfrentar limitações tanto em termos de desempenho computacional quanto de capacidade de generalização (PENG; GUI; WU, 2025).

À medida que a dimensionalidade aumenta, problemas podem se tornar computacionalmente intratáveis, e padrões identificáveis em espaços de baixa dimensão podem se tornar menos detectáveis (DONOHO et al., 2000). Neste contexto, desenvolvem-se abordagens para mitigar esses efeitos, incluindo técnicas de redução de dimensionalidade e de seleção de características que preservem a informação relevante enquanto reduzem redundâncias nos modelos de aprendizado de máquina (GUYON; ELISSEEFF, 2003).

A maldição da dimensionalidade pode ser quantificada por meio de diversas abordagens que demonstram como a complexidade computacional e os requisitos do conjunto de dados crescem de forma não linear com o aumento da dimensionalidade (DONOHO et al., 2000). Uma das manifestações mais diretas deste fenômeno emerge da análise de densidade em espaços multidimensionais através da discretização do espaço de características. Para compreender este conceito, considere que cada variável ou característica pode ser dividida em faixas de valores (*bins*) para análise estatística, ou seja, por exemplo, dividir a variável “idade” em faixas de 10 anos, ou “renda” em faixas de R\$ 1.000.

Quando se aplica esta discretização simultaneamente a múltiplas dimensões, cria-se uma grade multidimensional onde cada célula representa uma combinação específica de faixas de valores. Ao dividir cada dimensão em k faixas (por exemplo, $k = 10$ faixas por variável), o número total de células ou compartimentos no espaço cresce segundo $N \approx k^d$, onde N representa o número de amostras necessário para que cada célula tenha representação estatística adequada, k o número de subdivisões por dimensão, e d o número de dimensões (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Para ilustrar: com duas variáveis (idade e renda) divididas em 10 faixas cada, obtém-se $10^2 = 100$ células; com três variáveis, $10^3 = 1.000$ células; mas com apenas 20 variáveis, tem-se 10^{20} células. Esta relação exponencial implica que problemas aparentemente similares apresentam requisitos computacionais drasticamente diferentes: enquanto um problema bidimensional pode requerer 100 amostras para adequada representação, um problema com 20 dimensões demandaria aproximadamente 10^{20} amostras.

Espaços de alta dimensionalidade também podem apresentar o fenômeno de concentração de distâncias, onde as distâncias euclidianas entre pontos tendem a convergir para valores similares, tornando métricas baseadas em distância menos discriminativas (BEYER et al., 1999). À medida que o número de características aumenta, a complexidade do problema pode se multiplicar através de aspectos que se reforçam mutuamente (BOUSQUET; ELISSEEFF, 2002): o espaço de hipóteses possíveis expande exponencialmente, tornando a busca por soluções mais custosa computacionalmente, indicando que maiores volumes de dados se tornam necessários para que algoritmos aprendam adequadamente.

A dimensão Vapnik-Chervonenkis, uma medida que determina a capacidade de aprendizado de um algoritmo, cresce linearmente com o número de características ($p + 1$) (VAPNIK, 1999). Isso significa que, mesmo no cenário mais otimista, são necessárias pelo menos $p + 1$ amostras apenas para evitar que o modelo simplesmente “decore” os dados de treinamento (AWASTHI; HAGHTALAB; ZHAO, 2023). Esta relação fundamenta uma regra prática da comunidade de aprendizado de máquina de que $n \gg p$ (onde n representa o número de amostras no conjunto de dados), idealmente com $n \geq 10p$ ou $n \geq 50p$, dependendo da complexidade do algoritmo de aprendizado de máquina empregado (PEREG, 2023).

Quando a condição $n \gg p$ não é satisfeita, configurando o cenário “*small n, large p*”, observam-se problemas como matrizes singulares, *overfitting* e redução da capacidade de generalização (JOHNSTONE, 2001). A presença de múltiplas classes (C) introduz complexidade adicional que se manifesta nos requisitos de dados e na estrutura geométrica do espaço de decisão (JAIN; DUIN; MAO, 2000).

A complexidade aumenta quando se distinguem múltiplas classes (C). A separação de dezenas ou centenas de categorias simultaneamente em espaços de alta dimensionalidade pode requerer, segundo teoremas de aprendizado estatístico, aproximadamente $n \geq C \times p \times \log(p)$ amostras para garantir representação adequada de C classes em p dimensões (SHALEV-SHWARTZ; BEN-DAVID, 2014). Dessa forma, os requisitos de dados podem crescer proporcionalmente: dobrar o número de classes e dobrar o número de características pode resultar na necessidade quatro vezes maior de amostras.

Do ponto de vista geométrico, separar múltiplas classes requer a criação de fronteiras de decisão (as superfícies matemáticas que delimitam onde uma classe termina e outra começa). Para C classes, são necessários $(C - 1) \times p$ parâmetros para definir os hiperplanos de separação. Simultaneamente, essas fronteiras dividem o espaço em regiões distintas, e o número máximo de regiões possíveis (R) pode ser calculado considerando todas as combinações possíveis de classes (C) tomadas de k em k elementos, onde k varia de 0 até p dimensões, segundo $R \leq \sum_{k=0}^p \binom{C}{k}$ (DESHPANDE, 2014).

Este aumento no número de fronteiras resulta em maior complexidade de classificação, indicando que maiores volumes de dados podem ser necessários para manter níveis similares de desempenho à medida que o número de classes aumenta. Quando alta dimensionalidade e múltiplas classes se combinam, a complexidade total pode crescer proporcionalmente a C^p (VERLEYSSEN; FRANÇOIS, 2005), caracterizando cenários computacionalmente desafiadores.

Em aplicações reais, a classificação de textos pode envolver vocabulários de até 10^6 palavras únicas (XUE et al., 2024), o reconhecimento de imagens trabalha com milhões de *pixels* e análises genômicas lidam com milhares de genes para identificar fenótipos (RUNCIE; MUKHERJEE, 2013). Estas características dimensionais sugerem requisitos de dados que podem exceder a disponibilidade dos conjuntos de dados existentes, resultando no fenômeno conhecido como “*sparsity*”, uma dispersão onde a densidade efetiva de dados aproxima-se de zero: $n/(C \times 2^p) \rightarrow 0$. Neste contexto, classes minoritárias apresentam representação reduzida, resultando em modelos com menor capacidade de representar categorias menos frequentes (HE; GARCIA, 2009).

Diante desses desafios, técnicas como redução de dimensionalidade, seleção criteriosa de características, métodos de regularização e estratégias hierárquicas de classificação tornam-se úteis e potencialmente necessárias para viabilizar problemas de aprendizado de máquina em domínios de alta dimensionalidade.

Para compreender como esses desafios se manifestam na prática, pode-se examinar o problema de prever quais *posts* no Instagram terão alto engajamento: uma tarefa que exemplifica tanto a maldição da dimensionalidade quanto problemas correlatos. Cada publicação na plataforma caracteriza-se por centenas de variáveis: características textuais da legenda (sentimento, comprimento, uso de *hashtags*, menções), elementos visuais da imagem (cores dominantes, presença de faces, objetos identificados, qualidade estética), atributos temporais (horário de publicação, dia da semana, sazonalidade), métricas do usuário (número de seguidores, taxa de engajamento histórica, frequência de *posts*) e fatores contextuais (*trending topics*, eventos atuais e mudanças no algoritmo do *feed*). Essa alta dimensionalidade contribui para a complexidade da modelagem computacional, mas um problema emerge quando se busca compreender a distribuição real dos dados de engajamento e a correlação entre as diferentes características observáveis.

Embora as plataformas de redes sociais mantenham sigilo sobre estatísticas detalhadas de distribuição de engajamento, por razões comerciais e proteção de algoritmos proprietários, estudos independentes sugerem que o engajamento segue distribuições *power law*, onde uma pequena fração do conteúdo (aproximadamente 20%) gera a maioria (cerca de 80%) dos resultados (BARABÁSI, 2005; ADAMIC; BUYUKKOKTEN; ADAR, 2003). Essa opacidade demanda que se infiram padrões comportamentais com base em amostras e princípios teóricos estabelecidos, como o princípio de Pareto observado em fenômenos de redes sociais e comunidades online (NEWMAN, 2005).

Essa distribuição desbalanceada resulta no fenômeno conhecido como mascaramento de classes (*class masking*), onde a classe majoritária (*posts* de baixo engajamento) mascara os padrões das classes minoritárias (*posts* de alto engajamento), levando algoritmos de aprendizado de máquina a ignorar as características que distinguem diferentes tipos de conteúdo. Diferentemente da maldição da dimensionalidade, que afeta a eficiência computacional, o mascaramento de classes compromete a capacidade dos modelos de identificar padrões nas categorias de interesse prático.

O mascaramento de classes emerge como uma consequência natural de distribuições desbalanceadas, onde a classe majoritária esconde os padrões das classes minoritárias durante o processo de aprendizado. Matematicamente, este fenômeno pode ocorrer quando a proporção de uma classe majoritária $P(y_{\text{maj}})$ é maior que as demais classes, com razões de desbalanceamento $P(y_{\text{maj}})/P(y_{\text{min}}) > 10$ (HE; GARCIA, 2009). No contexto de classificação de engajamento, isso significa que a abundância de *posts* com baixo engajamento (classe majoritária) pode mascarar os padrões que caracterizam *posts* virais (classe minoritária), levando algoritmos a prever baixo engajamento.

O problema manifesta-se de forma sutil porque algoritmos de aprendizado de máquina podem otimizar métricas que são enviesadas na direção da classe majoritária. Em outras palavras, quando 90% dos dados pertencem à classe majoritária, um algoritmo pode

alcançar 90% de acurácia simplesmente ignorando completamente as classes minoritárias e sempre prevendo a classe mais comum. Este viés não é uma falha de implementação, mas uma consequência matemática: o algoritmo está otimizando corretamente a função objetivo, mas a função não reflete adequadamente os objetivos reais do problema (JAPKOWICZ; STEPHEN, 2002).

As implicações práticas do mascaramento tornam-se evidentes ao examinar métricas de desempenho tradicionais. A acurácia global pode ser enganosamente alta, mas métricas como sensibilidade (*recall*) para classes minoritárias podem frequentemente aproximar-se de zero (JOHNSON; KHOSHGOFTAAR, 2019). Este resultado não indica necessariamente baixa qualidade do modelo, mas sim que o modelo “aprendeu” a estratégia estatisticamente ótima para a função de custo apresentada, estratégia que falha em capturar os fenômenos de real interesse. Em problemas como detecção de conteúdo viral em redes sociais, detecção de fraudes em aplicações de crédito ou diagnóstico médico de doenças raras, essa limitação pode tornar impraticáveis modelos que apresentam uma boa acurácia, uma vez que falham sistematicamente em identificar exatamente os casos que motivaram a construção do sistema preditivo (HAIXIANG et al., 2017). Estas limitações do mascaramento de classes exemplificam a necessidade de investigar esses fenômenos em contextos reais. Plataformas de redes sociais, particularmente o Instagram, representam cenários apropriados para este tipo de investigação.

A relevância acadêmica do Instagram como objeto de estudo transcende sua popularidade como plataforma de compartilhamento de conteúdo, constituindo um microcosmo digital que exemplifica desafios da sociedade contemporânea: desde questões técnicas de aprendizado de máquina até dilemas éticos sobre privacidade, desinformação e influência política. Esta rede social oferece um laboratório para investigar como algoritmos de classificação de engajamento operam em contextos de complexidade dimensional e desbalanceamento de classes, ao mesmo tempo em que levanta questões sobre o papel das plataformas digitais na formação da opinião pública.

A trajetória do Instagram exemplifica a velocidade e escala das transformações digitais contemporâneas. Fundado em 2010 por Kevin Systrom e Mike Krieger, cofundador brasileiro nascido em São Paulo, o aplicativo começou como um projeto universitário focado em simplicidade: compartilhamento de fotos com filtros estéticos. A origem brasileira de um dos fundadores adiciona uma dimensão para o contexto nacional desta pesquisa, demonstrando como a inovação tecnológica transcende fronteiras geográficas, idiomas e culturas. Após 25 mil *downloads* nas primeiras 24 horas de lançamento, a plataforma cresceu, atraindo 1 milhão de usuários em dois meses e alcançando a marca de US\$ 1 bilhão em valor de aquisição pela Meta (então Facebook) em abril de 2012 (Facebook, 2012). Esta transação demonstrou o potencial das redes sociais e marcou o início de uma era de consolidação no setor tecnológico.

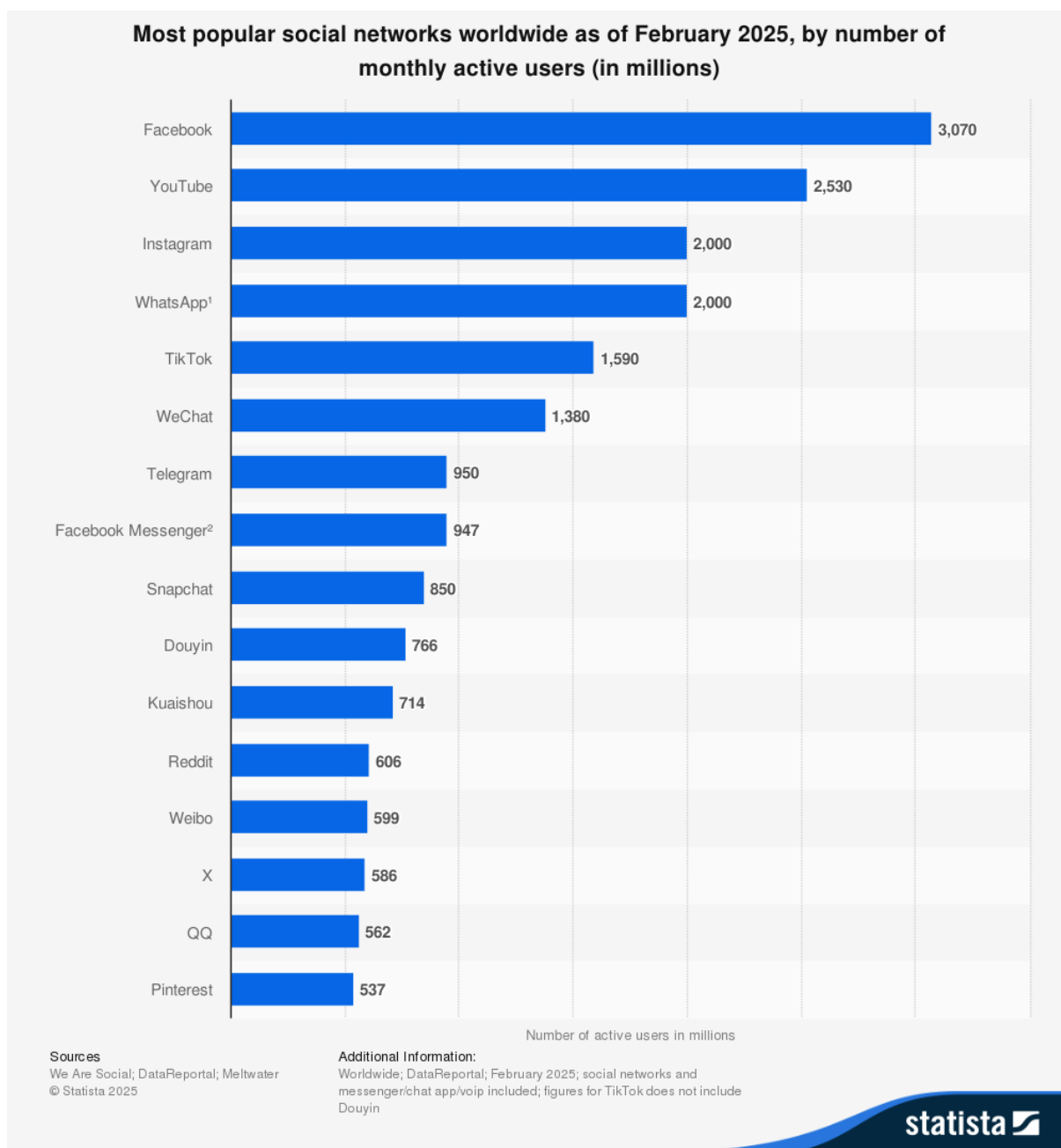
Com 1,5 a 2 bilhões de usuários ativos, o Instagram estabeleceu-se como uma das principais redes sociais. Junto ao WhatsApp, divide a terceira colocação mundial entre as redes sociais em número de usuários, consolidando a presença da Meta no setor, conforme [Figura 1.2](#) e [Figura 1.3](#). A [Figura 1.4](#) revela que o Instagram mantém um perfil jovem em sua base de usuários. A faixa etária de 25 a 34 anos representa o maior segmento da plataforma, com 31,6% dos usuários, seguida pelos usuários de 18 a 24 anos, que correspondem a 29,5%. Juntos, esses dois grupos etários concentram mais de 60% da audiência do Instagram. A participação diminui conforme a idade avança: usuários de 35 a 44 anos representam 16,3%, enquanto aqueles entre 45 e 54 anos correspondem a 9%. As faixas etárias superiores têm presença menor na plataforma, com usuários de 55 a 64 anos representando 4,9% e aqueles com 65 anos ou mais 3,2%. Essa distribuição confirma o posicionamento do Instagram como uma rede social que atrai adultos jovens e *millennials*, refletindo suas características visuais e sua cultura digital.

No Brasil, o alcance do Instagram é expressivo: com 134,6 milhões de usuários, a plataforma alcança 62% da população brasileira ([KEMP, 2024a](#)), posicionando o país como o terceiro maior mercado mundial do Instagram, atrás da Índia e dos Estados Unidos, conforme [Figura 1.5](#). O crescimento brasileiro tem sido acelerado, com um aumento de 18,6% entre 2023 e 2024, refletindo a expansão da conectividade digital no país e a preferência cultural brasileira por plataformas visuais e interativas ([Hootsuite, 2024](#)). Esses números posicionam o Brasil como um mercado consumidor e como um laboratório para estudos sobre comportamento em redes sociais, engajamento de conteúdo e dinâmicas de comunicação digital em contextos culturais específicos.

A influência sociopolítica do Instagram e demais plataformas da Meta tornou-se evidente durante eventos eleitorais de repercussão mundial. O episódio envolvendo a Cambridge Analytica nas eleições presidenciais americanas de 2016 revelou como dados de 87 milhões de usuários do Facebook foram utilizados para *microtargeting* político, demonstrando o potencial das redes sociais para influenciar processos democráticos através de técnicas de disseminação de conteúdo em redes sociais ([CADWALLADR, 2018](#)). Este episódio não se limitou aos Estados Unidos: investigações revelaram que técnicas similares foram empregadas em dezenas de países, incluindo o Brasil. No contexto brasileiro, as eleições de 2018 e 2022 evidenciaram como WhatsApp e Instagram se tornaram veículos para disseminação de desinformação política.

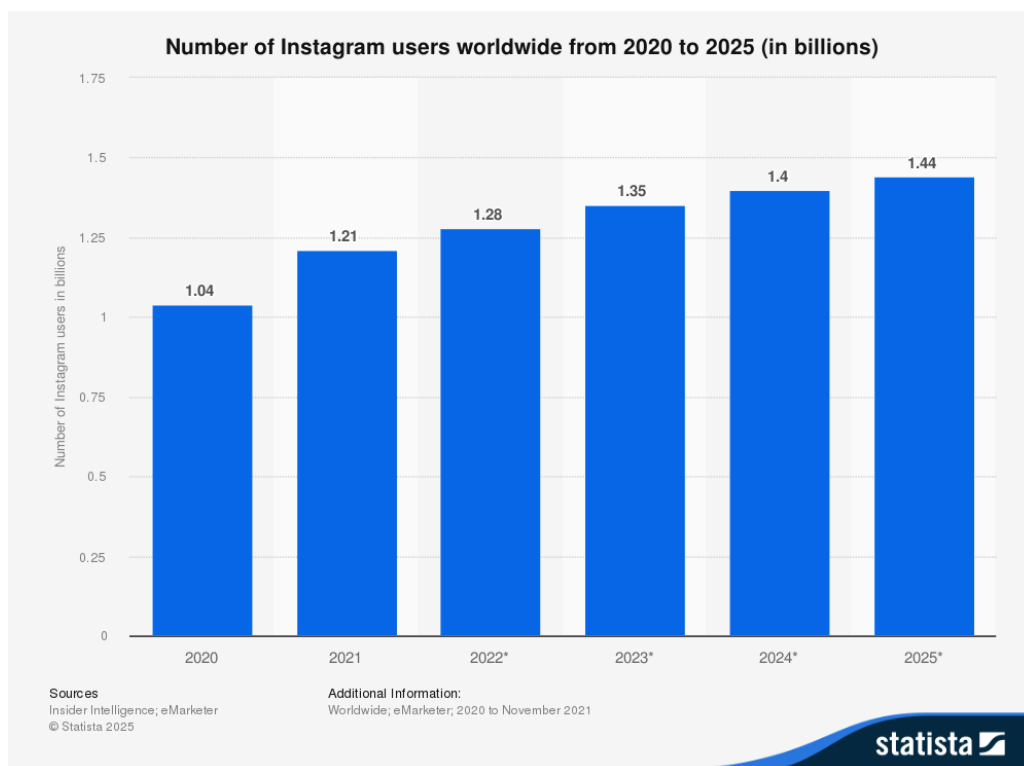
Durante o segundo turno de 2018, foram identificadas 346 *fake news* circulando em apenas 60 dias, uma média de 6 notícias falsas por dia, demonstrando a escala industrial da desinformação em períodos eleitorais ([DOURADO, 2020](#)). As eleições de 2022 intensificaram essa problemática, levando o Tribunal Superior Eleitoral (TSE) a remover mais de 100 postagens com narrativas enganosas e estabelecer parcerias formais com plataformas digitais para combate à desinformação ([Tribunal Superior Eleitoral, 2022](#)).

Figura 1.2 – Redes sociais mais populares no mundo em fevereiro de 2025



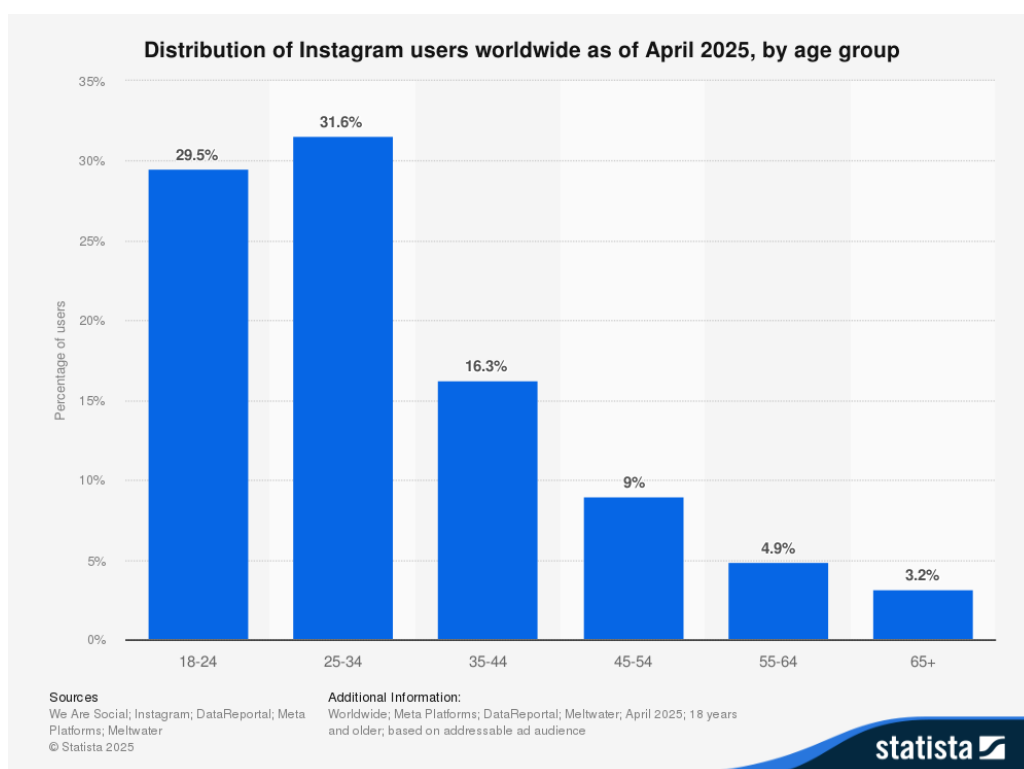
Fonte: (Statista, 2025d)

Figura 1.3 – Número de usuários do Instagram no mundo de 2020 a 2025



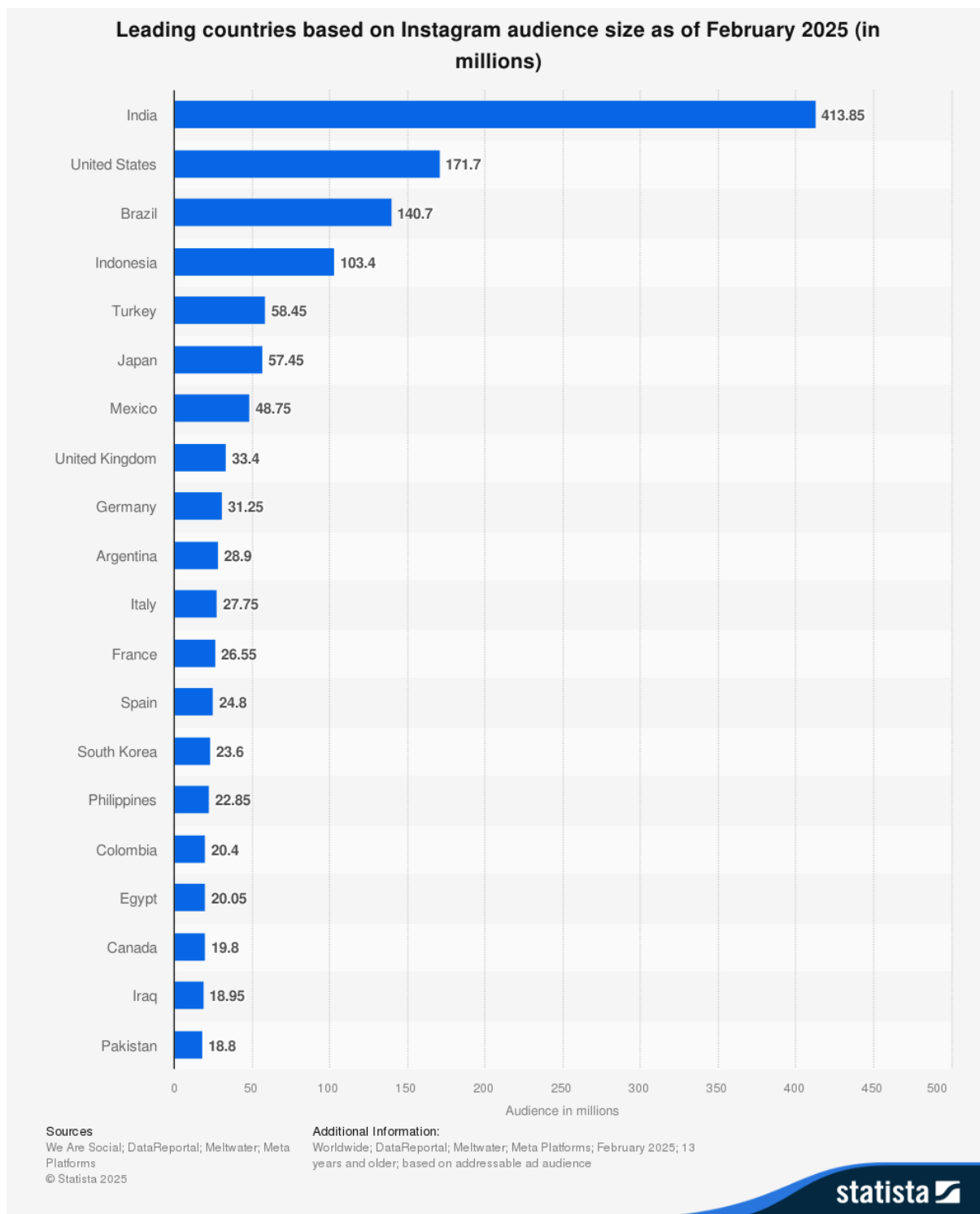
Fonte: (Statista, 2025c)

Figura 1.4 – Distribuição dos usuários do Instagram por idade em abril de 2025



Fonte: (Statista, 2025b)

Figura 1.5 – Países com maior audiência no Instagram em fevereiro de 2025



Fonte: (Statista, 2025a)

No Brasil, a Lei Geral de Proteção de Dados (LGPD), promulgada em 2018 e inspirada no Regulamento Geral sobre a Proteção de Dados (GDPR ou *General Data Protection Regulation*) europeu, estabeleceu princípios fundamentais para tratamento de dados pessoais, criando a Autoridade Nacional de Proteção de Dados (ANPD) como órgão fiscalizador (Brasil, 2018). A aplicação prática dessa regulamentação tornou-se evidente em 2024, quando a ANPD proibiu temporariamente a Meta de utilizar dados de usuários brasileiros para treinamento de inteligência artificial, citando “risco iminente de dano grave e irreparável” aos direitos fundamentais dos titulares, promovendo um intenso debate sobre o tema (ANPD, 2024a).

Embora a proibição tenha sido revogada em agosto de 2024 após a empresa cumprir exigências, o episódio é um exemplo de como diferentes jurisdições estão desenvolvendo abordagens distintas para regulamentação de tecnologias emergentes (ANPD, 2024b). Enquanto a União Europeia adota postura mais restritiva através do GDPR e do recente *AI Act*, os Estados Unidos mantêm abordagem mais permissiva baseada na autorregulação da indústria, criando um mosaico regulatório complexo que empresas globais como a Meta devem navegar (European Union, 2024).

O dilema contemporâneo sobre regulamentação *versus* inovação na era da inteligência artificial exemplifica perfeitamente os desafios de governança tecnológica no século XXI. O Senado brasileiro aprovou em dezembro de 2024 um marco regulatório para inteligência artificial que busca “equilibrar inovação e proteção”, estabelecendo regras para desenvolvimento e uso de sistemas de IA enquanto exclui algoritmos de redes sociais da categoria de “alto risco” (Agência Senado, 2024). Esta decisão reflete tensões entre a necessidade de proteger direitos e a preocupação de não sufocar a inovação tecnológica que posiciona o Brasil competitivamente no cenário global. O governo federal planeja investimentos em inteligência artificial para modernização de serviços públicos, demonstrando como a regulamentação pode coexistir com incentivos à inovação (LNCC, 2024).

O GDPR europeu, que entrou em vigor em maio de 2018, estabeleceu-se como paradigma mundial para proteção de dados pessoais, destacando-se por sua aplicação extraterritorial, multas substanciais de até 4% do faturamento global anual e direitos robustos dos titulares, como o direito ao esquecimento e à portabilidade (European Parliament; Council of the European Union, 2016). Já o *AI Act*, que entrou em vigor em agosto de 2024 com aplicabilidade plena prevista para 2026, representa o primeiro marco regulatório abrangente para inteligência artificial no mundo, adotando uma abordagem baseada em risco que classifica sistemas de IA em quatro categorias: proibidos, alto risco, transparência limitada e risco mínimo (European Union, 2024). Contudo, ambas as regulamentações enfrentam críticas por parte do setor empresarial, especialmente de pequenas e médias empresas, que argumentam que os custos de conformidade e as exigências burocráticas podem inibir a inovação tecnológica e a competitividade no mercado global.

Embora a LGPD tenha se inspirado no GDPR, existem diferenças notáveis. A principal delas reside na criação da ANPD como órgão centralizador, contrastando com o modelo europeu que delega a fiscalização às autoridades nacionais de cada Estado-membro. No tratamento de dados sensíveis, o GDPR adota postura mais restritiva, praticamente proibindo seu uso salvo previsão legal específica, enquanto a LGPD permite flexibilidade mediante consentimento específico. As sanções também diferem em magnitude: o GDPR estabelece multas de até 4% do faturamento, enquanto a LGPD limita-se a 2%, refletindo uma abordagem mais moderada (OLIVEIRA; SILVEIRA; OLIVEIRA, 2025).

O marco regulatório brasileiro para inteligência artificial adota filosofia distinta do *AI Act* europeu, buscando equilíbrio entre proteção de direitos e estímulo à inovação. A principal divergência reside na exclusão dos algoritmos de redes sociais da categoria de “alto risco”, uma decisão controversa que contrasta com a abordagem mais restritiva da Europa. O texto brasileiro enfatiza fortemente a proteção dos direitos autorais, exigindo que empresas de tecnologia informem quais conteúdos protegidos foram utilizados no treinamento de sistemas de IA e garantindo aos criadores o direito de vetar o uso de suas obras (Agência Senado, 2024). Estruturalmente, o Brasil criará o Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA), coordenado pela ANPD, diferindo do modelo europeu que estabelece o *AI Office* como autoridade central.

Os Estados Unidos mantêm abordagem distinta, caracterizada pela fragmentação regulatória e preferência pela autorregulação setorial. Na proteção de dados, o país não possui legislação federal abrangente, operando com aproximadamente 20 leis setoriais federais e mais de 100 legislações estaduais, sendo a *California Consumer Privacy Act* (CCPA) a mais significativa (California State Legislature, 2018). Esta lei difere substancialmente do GDPR ao permitir a “venda” de dados pessoais mediante mecanismo de *opt-out*, contrastando com o consentimento prévio exigido pela regulamentação europeia. No campo da inteligência artificial, a administração Biden havia estabelecido diretrizes através de ordem executiva em 2023, posteriormente revogada pelo presidente Trump em janeiro de 2025, sinalizando retorno à política de mínima intervenção governamental. Atualmente, os estados avançam independentemente com mais de 700 projetos de lei sobre IA, criando um mosaico regulatório complexo que empresas devem navegar.

Neste cenário global de regulamentação tecnológica, o Brasil posiciona-se como uma “terceira via” entre o rigor regulatório europeu e a permissividade americana, assumindo papel de mediador entre diferentes filosofias regulatórias. A estratégia brasileira busca preservar a competitividade da indústria nacional de tecnologia sem comprometer a proteção dos direitos fundamentais dos cidadãos, adaptando os modelos internacionais às especificidades do mercado doméstico. Esta abordagem pragmática reflete-se tanto na LGPD, que incorpora flexibilidades ausentes no GDPR, quanto no marco de IA, que exclui deliberadamente as redes sociais das categorias de maior risco regulatório.

Para sistematizar e facilitar a compreensão das diferentes abordagens regulatórias analisadas, a [Tabela 1.1](#) e [Tabela 1.2](#) apresentam uma análise comparativa estruturada entre Brasil, União Europeia e Estados Unidos nos campos da proteção de dados pessoais e inteligência artificial, respectivamente. Estes quadros sintetizam as principais características de cada modelo regulatório, evidenciando as convergências e divergências entre as estratégias adotadas pelos diferentes atores globais, bem como suas especificidades.

Tabela 1.1 – Comparação das abordagens regulatórias para proteção de dados

Aspecto	Brasil	União Europeia	Estados Unidos
Marco Legal	LGPD (2018)	GDPR (2016)	CCPA e leis estaduais
Autoridade Reguladora	ANPD (órgão centralizador)	Autoridades supervisoras independentes por Estado-membro	Fragmentada (FTC federal; Procurador-Geral estadual)
Abordagem	Moderadamente restritiva; inspirada no GDPR	Rígida e precaucionária; aplicação extraterritorial	Setorial; autorregulação; foco em consumo
Multas Máximas	Até 2% do faturamento ou R\$ 50 milhões	Até 4% do faturamento ou €20 milhões	Variável por estado e setor
Dados Sensíveis	Consentimento específico e destacado	Proibição geral com exceções; consentimento explícito	Proteções variadas; CCPA permite <i>opt-out</i>
Direitos dos Titulares	Acesso, correção, eliminação, portabilidade, oposição	Acesso, retificação, apagamento, portabilidade, oposição	Direito de saber, deletar, <i>opt-out</i> , não discriminação
Alcance Territorial	Dados tratados no Brasil ou de residentes brasileiros	Extraterritorial (oferecimento de bens/serviços na UE)	Principalmente territorial (ex: residentes da Califórnia)
Características Distintivas	Flexibilidade; órgão centralizador nacional	<i>Privacy by design</i> ; direito ao esquecimento	<i>Opt-out</i> ; ausência de regulação federal unificada

Fonte: Elaborado pelo autor com base em ([Brasil, 2018](#)), ([European Parliament; Council of the European Union, 2016](#)) e ([California State Legislature, 2018](#)).

Especialistas alertam que regulamentações excessivamente prescritivas podem criar “barreiras técnicas, legais ou financeiras” que desfavorecem *startups* e pequenas empresas em relação a grandes corporações com recursos para *compliance* ([CHAVES, 2024](#)). Este equilíbrio delicado torna-se ainda mais complexo quando se considera que plataformas como Instagram não são apenas empresas de tecnologia, mas infraestruturas de comunicação que moldam discursos públicos, influenciam eleições e afetam a saúde mental de bilhões de usuários. Consequentemente, a regulamentação dessas plataformas não pode ser vista apenas através da lente econômica ou tecnológica, mas deve incorporar considerações sobre democracia, direitos humanos e bem-estar social, criando um desafio regulatório sem precedentes na história da humanidade.

Tabela 1.2 – Comparação das abordagens regulatórias para inteligência artificial

Aspecto	Brasil	União Europeia	Estados Unidos
Marco Legal	PL 2.338/2023 (em tramitação na Câmara; previsão 2025)	<i>AI Act</i> (2024) - aplicação plena em 2026	Sem regulamentação federal; mais de 700 projetos estaduais
Autoridade Competente	Sistema Nacional de Regulação e Governança de IA (SIA), coordenado pela ANPD	<i>AI Office</i> da Comissão Europeia; autoridades nacionais	Sem autoridade federal; agências estaduais variadas
Classificação de Risco	Sistemas proibidos, alto risco e baixo risco	4 categorias: inaceitável, alto risco, transparência limitada, risco mínimo	Sem classificação federal; critérios variam por estado
Principais Restrições	Pontuação social governamental; manipulação comportamental subliminar	Pontuação social; manipulação subliminar; reconhecimento facial em tempo real	Sem proibições federais específicas
Transparência	Empresas devem informar conteúdos protegidos usados no treinamento; direito de veto dos criadores	Rotulagem de conteúdo gerado por IA; transparência sobre dados de treinamento	Compromissos voluntários; sem obrigatoriedade federal
Sanções	A definir na regulamentação	Até €35 milhões ou 7% do faturamento global	Varia por estado; sem estrutura federal
Filosofia Regulatória	Equilíbrio entre inovação e proteção de direitos; adaptação ao mercado emergente	Precaucionária: proteção de direitos fundamentais e segurança como prioridade	Autorregulação da indústria; competitividade; resistência à intervenção federal

Fonte: Elaborado pelo autor com base em (Agência Senado, 2024), (European Union, 2024) e (OLIVEIRA; SILVEIRA; OLIVEIRA, 2025).

A convergência entre os desafios técnicos da maldição da dimensionalidade e do mascaramento de classes, exemplificada através do contexto de classificação de engajamento no Instagram, revela a complexidade dos problemas contemporâneos de aprendizado de máquina em ambientes de produção real. Este capítulo demonstrou como questões técnicas não podem ser dissociadas de suas implicações sociais, econômicas e políticas. O Instagram, com seus 134,6 milhões de usuários brasileiros e influência documentada em processos democráticos, representa mais que uma plataforma de entretenimento: constitui uma infraestrutura digital onde algoritmos de classificação operam em larga escala, influenciando a visibilidade de conteúdo, a formação de opinião pública e o sucesso econômico de milhões.

A análise apresentada sugere que soluções para problemas de alta dimensionalidade e desbalanceamento de classes podem ser inadequadas quando confrontadas com a realidade de plataformas como Instagram. A necessidade de processar centenas de características por publicação, distinguir padrões entre bilhões de *posts*, e fazê-lo em tempo real para audiências globais, cria um cenário onde limitações algorítmicas podem se tornar obstáculos.

Simultaneamente, as considerações regulatórias, desde a LGPD até dilemas sobre transparência algorítmica e responsabilidade social, introduzem considerações que devem ser incorporadas no desenvolvimento de soluções técnicas. Esta intersecção entre desafios computacionais, requisitos de desempenho e considerações éticas estabelece um problema de pesquisa que possivelmente transcende fronteiras disciplinares tradicionais.

Consequentemente, torna-se evidente a necessidade de investigar as abordagens propostas na literatura para mitigação desses desafios, avaliando sua eficácia técnica e sua viabilidade prática em contextos de redes sociais de grande escala. As técnicas clássicas para redução de dimensionalidade e classificação, balanceamento de classes e otimização de métricas alternativas foram desenvolvidas para domínios com características potencialmente diferentes daquelas encontradas em plataformas digitais contemporâneas.

A revisão dessas metodologias, considerando suas limitações e potenciais adaptações para o contexto de classificação de engajamento em redes sociais, constituirá o foco do próximo capítulo desta investigação. Esta análise da literatura abrangerá tanto as abordagens clássicas de aprendizado de máquina quanto as técnicas de aprendizado profundo (*deep learning*), examinando como cada uma responde aos desafios impostos pela natureza dinâmica, heterogênea e de alta dimensionalidade dos dados de redes sociais. A avaliação comparativa das diferentes abordagens propostas identificará possíveis lacunas metodológicas e oportunidades para desenvolvimentos futuros que possam endereçar as complexidades da classificação de comportamento de usuário em ambientes digitais contemporâneos.

2 Revisão Bibliográfica

O desenvolvimento de métodos para classificação de dados de alta dimensionalidade requer compreensão dos fundamentos teóricos e das limitações dos algoritmos de aprendizado de máquina existentes. Neste capítulo, apresentam-se os conceitos fundamentais, métodos de classificação e métricas de avaliação que constituem a base teórica para o desenvolvimento e validação da metodologia NL-TSDR proposta nesta tese.

A seleção dos tópicos abordados orienta-se pelos desafios enfrentados em tarefas de classificação com conjuntos de dados de alta dimensionalidade, particularmente o fenômeno de mascaramento de classes. Neste contexto, abordam-se as distinções entre abordagens lineares e não lineares, bem como as características dos métodos determinísticos e não determinísticos, propriedades que influenciam a capacidade de separação de classes.

Selecionaram-se os métodos de classificação apresentados neste capítulo como referência por representarem paradigmas distintos de aprendizado e por serem utilizados na literatura para avaliação comparativa de algoritmos. O *Naive Bayes* oferece base conceitual para classificação probabilística, o *k-Nearest Neighbors* (*k*-NN) representa métodos baseados em distâncias, a Análise de Discriminante Linear (LDA ou *Linear Discriminant Analysis*) exemplifica abordagens de separação linear e o Percéptron de Múltiplas Camadas (MLP ou *Multilayer Perceptron*) demonstra a capacidade de modelagem não linear através de redes neurais (CARUANA; NICULESCU-MIZIL, 2006).

As métricas de avaliação discutidas fornecem ferramentas para análise da qualidade dos dados transformados e medidas quantitativas para comparação de desempenho dos algoritmos. O Coeficiente de Correlação de Pearson permite avaliar a preservação de características discriminativas após a redução dimensional, o Coeficiente de *Silhouette* oferece medida da qualidade da separabilidade entre grupos, enquanto as métricas de classificação possibilitam a avaliação da eficácia do método proposto (BAGIROV; ALIGULIYEV; SULTANOVA, 2023).

2.1 Fundamentos de Aprendizado de Máquina e Classificação

O aprendizado de máquina constitui um ramo da inteligência artificial que se concentra no desenvolvimento de algoritmos capazes de identificar padrões em dados e realizar previsões ou decisões sem programação explícita para cada situação específica (RUSSELL; NORVIG, 2009). No contexto de classificação, o objetivo é atribuir rótulos (*labels*) a instâncias de dados com base em características observadas, processo que requer a compreensão de diferentes paradigmas de aprendizado e suas propriedades.

A eficácia de um classificador depende da escolha adequada do paradigma de aprendizado e das características algorítmicas empregadas. A distinção entre métodos supervisionados e não supervisionados determina como o conhecimento é extraído dos dados, enquanto a natureza determinística ou não determinística dos algoritmos influencia a reprodutibilidade dos resultados obtidos (GUNDERSEN; KJENSMO, 2018).

Estes fundamentos embasam o desenvolvimento de métodos de redução dimensional e classificação, uma vez que as transformações propostas devem preservar as propriedades discriminativas dos dados originais e manter compatibilidade com diferentes algoritmos de aprendizado (JOLLIFFE; CADIMA, 2016). Nesta seção, estabelecem-se os conceitos teóricos para contextualizar as contribuições metodológicas apresentadas nesta tese.

2.1.1 Tipos de Aprendizado

O aprendizado supervisionado caracteriza-se pela disponibilidade de um conjunto de dados de treinamento onde cada instância está associada a um rótulo conhecido, permitindo ao algoritmo aprender a relação entre características de entrada e saídas desejadas. Neste paradigma, o processo de treinamento visa minimizar o erro entre as predições do modelo e os valores reais conhecidos, estabelecendo uma função de mapeamento que possa generalizar para novos dados não observados durante o treinamento (BISHOP; BISHOP, 2024).

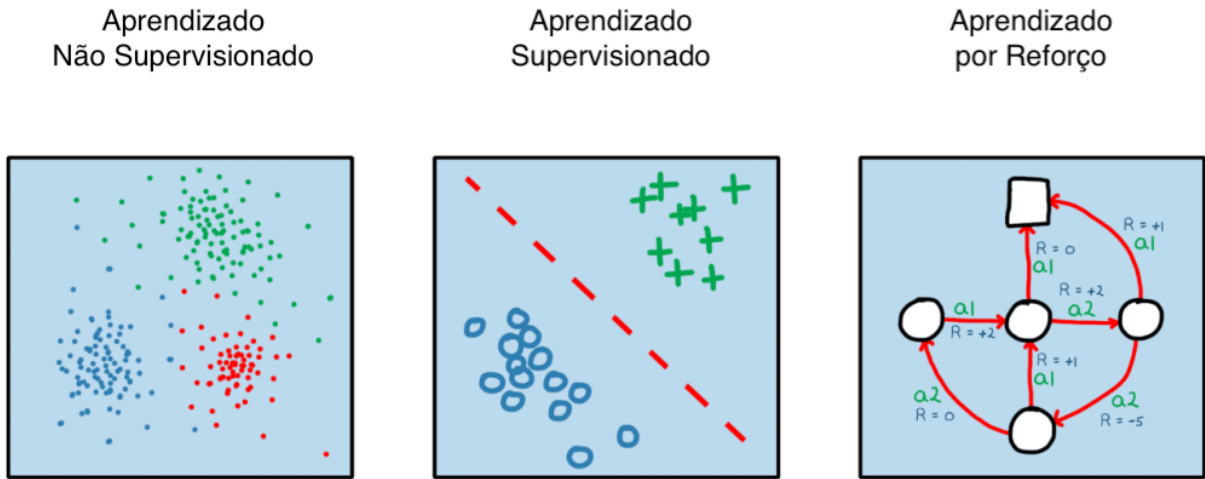
A diversidade de problemas e estruturas de dados encontrados em aplicações reais demanda uma gama variada de abordagens algorítmicas, cada uma adequada a contextos específicos e objetivos distintos. Na Figura 2.1, encontram-se os paradigmas fundamentais do aprendizado de máquina que abordam diferentes tipos de problemas: o aprendizado supervisionado, o aprendizado não supervisionado e o aprendizado por reforço.

Os algoritmos de classificação supervisionada dependem da qualidade e representatividade dos dados rotulados para construir modelos eficazes (DUDA; HART; STORK, 2012). A presença de ruído, desbalanceamento entre classes ou mascaramento de características pode comprometer o desempenho destes métodos, tornando necessárias técnicas de pré-processamento e transformação de dados.

Em contraste, o aprendizado não supervisionado opera sobre conjuntos de dados sem rótulos conhecidos, buscando descobrir estruturas ocultas ou padrões intrínsecos nos dados (JAIN, 2010). Técnicas como clusterização, redução de dimensionalidade e detecção de anomalias exemplificam este paradigma, onde o objetivo é revelar organizações naturais dos dados sem orientação externa explícita.

O aprendizado por reforço fundamenta-se na interação dinâmica entre um agente e seu ambiente, onde o conhecimento é adquirido através de um sistema de recompensas e penalizações (SUTTON; BARTO, 2018). Este paradigma enfrenta o desafio inerente do equilíbrio entre exploração de novas ações e exploração de estratégias já conhecidas.

Figura 2.1 – Os três paradigmas fundamentais do aprendizado de máquina



Fonte: Adaptado pelo autor de (TAN, 2024).

Para melhor compreensão das distinções e complementaridades entre estes paradigmas fundamentais do aprendizado de máquina, a Tabela 2.1 apresenta uma análise comparativa que abrange desde as características dos dados de entrada até as aplicações típicas de cada abordagem. Esta comparação sistemática evidencia não apenas as particularidades metodológicas de cada paradigma, mas também suas vantagens e limitações inerentes, fornecendo subsídios para a compreensão de como diferentes estratégias de aprendizado podem ser combinadas de forma sinérgica em aplicações específicas.

2.1.2 Métodos Determinísticos *versus* Não Determinísticos

A distinção entre métodos determinísticos e não determinísticos influencia a escolha e avaliação de algoritmos de aprendizado de máquina, manifestando-se na reprodutibilidade dos experimentos e na confiabilidade dos resultados obtidos (HENDERSON et al., 2018). Na Figura 2.2, ilustra-se essa diferença conceitual: métodos determinísticos seguem trajetória única, enquanto métodos não determinísticos exploram múltiplas possibilidades.

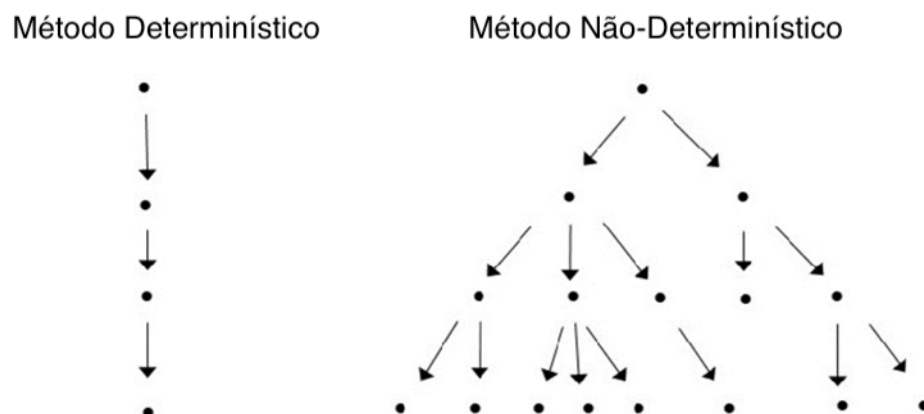
Métodos determinísticos produzem resultados idênticos quando executados múltiplas vezes sobre o mesmo conjunto de dados com os mesmos parâmetros. Algoritmos como a LDA e versões básicas do *Naive Bayes* exemplificam este comportamento, onde a solução é única e determinada pelos dados de entrada e pelos parâmetros do modelo. Esta propriedade assegura a reprodutibilidade experimental e permite a análise comparativa entre abordagens (HAN, 2025).

Tabela 2.1 – Comparação entre tipos de aprendizado

Características	Supervisionado	Não Supervisionado	Por Reforço
Dados	Rotulados	Sem rótulos	Interações com ambiente
Objetivo	Predizer saídas corretas	Descobrir padrões ocultos	Otimizar decisões por recompensas
Técnicas	Classificação, regressão, SVM, redes neurais	Clusterização (<i>k-means</i>), PCA, detecção de anomalias	<i>Q-learning</i> , <i>policy gradient</i> , <i>actor-critic</i>
Vantagens	Alta precisão; interpretabilidade; validação objetiva	Sem necessidade de rotulação; descoberta exploratória	Adaptação dinâmica; aprendizado por experiência
Limitações	Requer dados rotulados; sensível a ruído	Validação subjetiva; resultados ambíguos	Convergência lenta; alta complexidade computacional
Aplicações	Diagnóstico médico, detecção de spam	Segmentação de mercado, recomendação	Robótica, jogos, controle autônomo

Fonte: Elaborado pelo autor com base em (DUDA; HART; STORK, 2012), (JAIN, 2010) e (SUTTON; BARTO, 2018).

Figura 2.2 – Diferença entre métodos determinísticos e não determinísticos



Fonte: Adaptado pelo autor de (MOULINES, 2010).

Os métodos não determinísticos incorporam elementos de aleatoriedade em seus processos de otimização ou inicialização, resultando em soluções diferentes a cada execução. O MLP exemplifica este comportamento através da inicialização aleatória de pesos e do uso de algoritmos de otimização estocásticos (NARKHEDE; BARTAKKE; SUTAONE, 2022). Esta aleatoriedade permite que o algoritmo explore diferentes regiões do espaço de soluções, evitando mínimos locais.

No contexto de conjuntos de dados de alta dimensionalidade, a escolha entre abordagens determinísticas e não determinísticas deve considerar o equilíbrio entre reprodutibilidade e capacidade exploratória. Métodos determinísticos oferecem resultados consistentes mas podem ser limitados por sua natureza fixa, enquanto métodos não determinísticos podem fornecer soluções superiores através de múltiplas execuções, requerendo, contudo, procedimentos adequados de validação estatística (BEYER et al., 1999).

A crescente complexidade dos problemas de aprendizado de máquina tem incentivado o desenvolvimento de abordagens híbridas que combinam elementos determinísticos e não determinísticos para maximizar as vantagens de ambos os paradigmas (ZHOU, 2012). A hibridização permite que componentes determinísticos forneçam estabilidade e reprodutibilidade em etapas críticas do processamento, enquanto elementos estocásticos exploram eficientemente espaços de soluções complexos, relevante em problemas de otimização multiobjetivo onde múltiplos critérios devem ser simultaneamente otimizados.

Para sistematizar as principais diferenças e implicações práticas entre essas duas abordagens, a Tabela 2.2 apresenta uma síntese comparativa dos aspectos mais relevantes que devem ser considerados na seleção de métodos para problemas específicos de aprendizado de máquina. Esta comparação estruturada facilita a compreensão das vantagens e limitações inerentes a cada paradigma e auxilia na tomada de decisões.

A metodologia NL-TSDR incorpora elementos de ambas as abordagens: a primeira etapa de extração de características pode ser implementada de forma determinística, garantindo consistência na identificação de características discriminativas, enquanto a segunda etapa de transformações não lineares pode beneficiar-se de elementos estocásticos na otimização multiobjetivo para melhor explorar o espaço de soluções (MAATEN, 2014).

2.2 Abordagens Lineares *versus* Não Lineares em Classificação

A distinção entre métodos lineares e não lineares em classificação constitui aspecto a ser considerado na seleção de algoritmos para diferentes tipos de problemas e estruturas de dados. Esta distinção mostra-se relevante quando se trabalha com conjuntos de dados de alta dimensionalidade, onde a capacidade de modelar relações complexas entre características influencia o desempenho de um sistema de classificação (YUAN; HO; LIN, 2012).

Tabela 2.2 – Comparação entre métodos determinísticos e não determinísticos

Aspecto	Determinísticos	Não Determinísticos
Reprodutibilidade	Resultados idênticos com mesmos parâmetros	Resultados variam devido à aleatoriedade
Exploração	Trajectoria fixa no espaço de soluções	Múltiplas trajetórias possíveis
Otimização	Tendência a mínimos locais	Maior capacidade de encontrar ótimos globais
Validação	Análise comparativa direta	Requer procedimentos estatísticos
Vantagens	Consistência, reprodutibilidade, simplicidade	Flexibilidade, exploração ampla, soluções superiores
Limitações	Trajectoria fixa limita qualidade	Variabilidade, múltiplas execuções necessárias
Exemplos	LDA, <i>Naive</i> Bayes	MLP com inicialização aleatória, algoritmos genéticos

Fonte: Elaborado pelo autor com base em (HAN, 2025) e (NARKHEDE; BARTAKKE; SUTAONE, 2022).

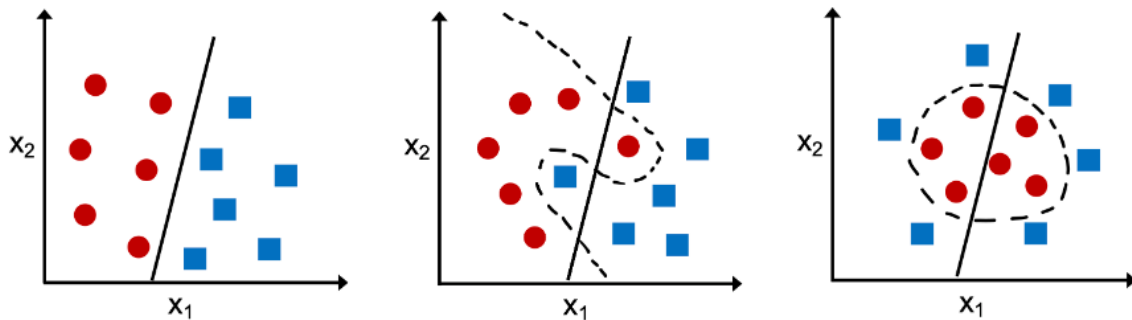
Métodos lineares assumem que as classes podem ser separadas por fronteiras de decisão lineares, uma suposição que potencialmente se mostra inadequada para dados do mundo real. Em contrapartida, abordagens não lineares possuem a capacidade de modelar relações mais complexas, criando fronteiras de decisão complexas que podem adaptar-se melhor às estruturas inerentes dos dados (SCHOLKOPF; SMOLA, 2001).

A Figura 2.3 demonstra essa progressão através de três cenários: inicialmente, uma fronteira linear que pode ser insuficiente para separar adequadamente as classes; em seguida, uma fronteira não linear simples que melhora a capacidade de separação; e, finalmente, fronteiras não lineares complexas capazes de capturar padrões mais sofisticados.

Esta evolução é uma evidência de como a flexibilidade dos modelos permite uma adaptação às características intrínsecas dos conjuntos de dados, embora também introduza desafios relacionados ao *overfitting* e à interpretabilidade dos resultados (LINARDATOS; PAPASTEFANOPOULOS; KOTSIANTIS, 2021). A transição de fronteiras lineares para não lineares representa uma relação custo-benefício no aprendizado de máquina: à medida que se aumenta a capacidade expressiva dos modelos, obtém-se maior poder de representação para capturar relações complexas nos dados, mas simultaneamente eleva-se o risco de memorização de ruídos presentes no conjunto de treinamento.

No contexto específico de conjuntos de dados de alta dimensionalidade afetados pelo mascaramento de classes, a escolha entre abordagens lineares e não lineares adquire importância ainda maior. O fenômeno de mascaramento ocorre quando a sobreposição de instâncias de diferentes classes compromete a capacidade discriminativa dos algoritmos (BLAGUS; LUSA, 2010), situação que pode ser abordada através de transformações não lineares adequadas combinadas com estratégias de otimização multiobjetivo.

Figura 2.3 – Evolução das fronteiras de decisão em problemas de classificação



Fonte: Adaptado pelo autor de (KUMAR, 2020).

Para sistematizar as principais características e diferenças entre essas abordagens, a Tabela 2.3 apresenta uma análise comparativa dos aspectos mais relevantes que distinguem métodos lineares e não lineares em tarefas de classificação. Esta comparação abrange desde aspectos fundamentais como a natureza das fronteiras de decisão até considerações práticas relacionadas ao custo computacional e adequação para diferentes tipos de conjuntos de dados, fornecendo diretrizes para a seleção de algoritmos em contextos específicos.

2.2.1 Possíveis Limitações dos Métodos Lineares

Os métodos lineares de classificação baseiam-se na premissa de que as classes presentes nos dados podem ser separadas por hiperplanos no espaço de características original. Esta abordagem assume que existe uma combinação linear das características de entrada capaz de distinguir entre as diferentes categorias, resultando em fronteiras de decisão lineares.

A Análise de Discriminante Linear exemplifica esta categoria de métodos, buscando encontrar projeções lineares que maximizem a separação entre classes e minimizem a dispersão dentro de cada classe (ZHAO et al., 2024). Versões lineares de algoritmos como o *Support Vector Machine* (SVM) ou regressão logística operam sob princípios semelhantes, procurando identificar hiperplanos de separação através de combinações lineares das características originais.

Os métodos lineares apresentam complexidade computacional reduzida, convergem para soluções e produzem modelos que podem ser interpretados. Métodos lineares apresentam capacidade de generalização quando a suposição de separabilidade linear é válida, evitando problemas de *overfitting* observados em modelos de maior complexidade (YUAN; HO; LIN, 2012).

Tabela 2.3 – Comparação entre abordagens lineares e não lineares em classificação

Aspecto	Métodos Lineares	Métodos Não Lineares
Fronteiras de Decisão	Hiperplanos lineares no espaço de características original	Fronteiras arbitrariamente complexas, adaptáveis à estrutura dos dados
Capacidade de Representação	Limitada a relações lineares entre características	Modelagem de relações complexas e interações não lineares
Complexidade do Modelo	Baixa complexidade, poucos parâmetros	Alta complexidade, maior número de parâmetros
Risco de Overfitting	Menor tendência ao overfitting devido à simplicidade	Maior suscetibilidade ao overfitting, especialmente com dados limitados
Interpretabilidade	Alta interpretabilidade, coeficientes com significado direto	Menor interpretabilidade, relações mais difíceis de compreender
Custo Computacional	Baixo custo computacional para treinamento e predição	Maior custo computacional, especialmente durante o treinamento
Adequação para Alta Dimensionalidade	Pode ser inadequado quando relações são inerentemente não lineares	Melhor capacidade para capturar padrões complexos em espaços de alta dimensão
Robustez a Ruído	Maior robustez devido à simplicidade do modelo	Menor robustez, pode memorizar ruídos nos dados
Generalização	Boa generalização quando suposições lineares são válidas	Generalização dependente de regularização e seleção adequada de hiperparâmetros
Exemplos de Algoritmos	Regressão Logística, SVM Linear, Perceptron	SVM com kernels não lineares, Redes Neurais, Árvores de Decisão
Tratamento de Mascaramento	Limitado pela impossibilidade de criar fronteiras complexas	Potencial superior para resolver sobreposições complexas entre classes

Fonte: Elaborado pelo autor com base em (LINARDATOS; PASTEFANOPOULOS; KOTSIANTIS, 2021).

Entretanto, as limitações dos métodos lineares tornam-se aparentes quando aplicados a problemas onde as relações entre características são intrinsecamente não lineares ou quando as fronteiras de decisão ótimas possuem formas complexas. Em conjuntos de dados de alta dimensionalidade, estas limitações são amplificadas pela maldição da dimensionalidade, onde o aumento do número de características pode paradoxalmente degradar o desempenho de métodos lineares devido à dispersão dos dados no espaço de alta dimensionalidade.

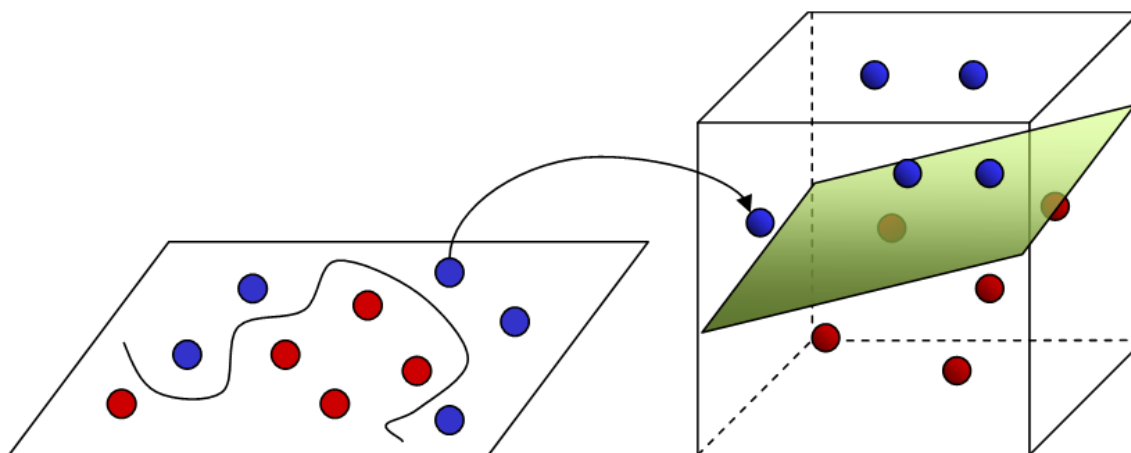
O mascaramento de classes representa uma limitação particularmente severa para métodos lineares, pois a sobreposição de instâncias de diferentes categorias no espaço original pode tornar impossível a identificação de hiperplanos de separação eficazes (WANG et al., 2025). Nestes cenários, transformações do espaço de características tornam-se necessárias para revelar estruturas discriminativas que não são aparentes através de combinações lineares simples das características originais.

2.2.2 Transformações Não Lineares e Separabilidade

As transformações não lineares oferecem uma abordagem poderosa para superar as limitações inerentes aos métodos lineares, permitindo que algoritmos de classificação operem em espaços transformados onde a separabilidade entre classes pode ser significativamente aprimorada. Estas transformações modificam o espaço de representação dos dados através de funções não lineares, potencialmente revelando estruturas discriminativas ocultas que não são aparentes no espaço original.

O princípio por trás das transformações não lineares reside na capacidade de mapear dados de um espaço onde as classes são não linearmente separáveis para um novo espaço onde a separação linear torna-se possível ou onde métodos não lineares podem operar mais efetivamente. Este conceito é exemplificado pelo *kernel trick* utilizado pela SVM (*Support Vector Machine*), onde transformações implícitas para espaços de alta dimensionalidade permitem a identificação de fronteiras de decisão complexas no espaço original (CORTES; VAPNIK, 1995).

Na verdade, *kernel trick* representa uma das inovações mais significativas no desenvolvimento da SVM e baseia-se na substituição dos produtos escalares convencionais por funções *kernel* que calculam implicitamente a similaridade entre pares de pontos de dados em um espaço de características transformado (BOSER; GUYON; VAPNIK, 1992). A elegância desta abordagem, ilustrada na Figura 2.4, reside no fato de que, embora o classificador opere como um hiperplano linear no espaço transformado de alta dimensionalidade, ele pode representar fronteiras de decisão arbitrariamente complexas no espaço de entrada original, superando assim as limitações fundamentais dos métodos de separação linear tradicionais.

Figura 2.4 – Representação gráfica do *kernel trick*

Fonte: Adaptado pelo autor de ([WILIMITIS, 2018](#)).

Funções como tangente hiperbólica, gaussiana, sigmoide e laplaciana representam exemplos comuns de transformações não lineares, cada uma com características específicas que podem ser adequadas para diferentes tipos de estruturas de dados. A tangente hiperbólica oferece propriedades de saturação que podem ser úteis para controlar o intervalo de valores transformados, enquanto funções gaussianas introduzem aspectos de localização que podem revelar agrupamentos naturais nos dados. Funções sigmoides proporcionam transições suaves entre diferentes regiões do espaço, e transformações laplacianas podem capturar relações de decaimento exponencial ([SCHOLKOPF; SMOLA, 2001](#)).

A eficácia das transformações não lineares em problemas de classificação de alta dimensionalidade está relacionada à sua capacidade de reduzir o mascaramento entre classes. Ao modificar as distâncias relativas entre instâncias de diferentes categorias, estas transformações podem amplificar diferenças discriminativas enquanto atenuam variações irrelevantes dentro das classes. Este processo potencialmente resulta em uma melhor separabilidade que se traduz em desempenho superior dos algoritmos de classificação subsequentes ([HOUSTON; GLAVIN; MADDEN, 2020](#)).

A seleção de transformações não lineares requer consideração das características dos dados e dos objetivos do problema de classificação. Diferentes funções podem ser aplicadas para diferentes tipos de distribuições de dados, e a combinação de múltiplas transformações pode produzir resultados distintos daqueles obtidos através de transformações isoladas. Esta flexibilidade demanda metodologias para a seleção e otimização de transformações.

2.2.3 Otimização Multiobjetivo em Classificação

A otimização multiobjetivo em problemas de classificação considera múltiplos critérios conflitantes simultaneamente. Diferentemente da otimização de objetivo único, onde existe uma medida de qualidade a ser maximizada ou minimizada, cenários multiobjetivo requerem abordagens que possam encontrar soluções que representem compromissos entre objetivos conflitantes. No contexto de classificação, os objetivos incluem a maximização das distâncias entre centros de classes (separabilidade interclasse) e a minimização da dispersão das instâncias dentro de cada classe (coesão intraclasse).

Estes objetivos podem ser conflitantes: transformações que aumentam a separação entre classes podem aumentar a dispersão dentro das classes, ou vice-versa. A identificação de soluções que equilibrem estes critérios constitui objetivo central da otimização multiobjetivo em classificação.

Algoritmos evolutivos multiobjetivo, como NSGA-II (*Non-dominated Sorting Genetic Algorithm*) (DEB; JAIN, 2014), SPEA (*Strength Pareto Evolutionary Algorithm*) (ZITZLER; THIELE, 1999) e MOEA/D (*Multi-Objective Evolutionary Algorithm based on Decomposition*) (ZHANG; LI, 2007), representam abordagens estabelecidas para lidar com este tipo de problema. Estes métodos mantêm populações de soluções candidatas e evoluem em direção ao conjunto de soluções ótimas onde nenhuma solução pode ser melhorada em um objetivo sem deteriorar outros objetivos.

A definição adequada das funções objetivo constitui um aspecto relevante na aplicação de otimização multiobjetivo à classificação. Métricas como a razão entre variâncias interclasse e intraclasse, distâncias entre centróides de classes ou medidas baseadas em separabilidade linear podem ser empregadas para quantificar diferentes aspectos da qualidade de separação. A escolha específica destas métricas deve refletir as características do problema e os requisitos de desempenho desejados (JIAO et al., 2024).

A aplicação de otimização multiobjetivo em transformações não lineares para classificação oferece a possibilidade de identificar configurações de parâmetros que maximizam simultaneamente múltiplos critérios de qualidade. Este processo pode envolver a otimização de parâmetros das próprias funções de transformação, pesos associados a diferentes características ou configurações de algoritmos de classificação subsequentes. O resultado é um conjunto de soluções que representam diferentes relações custo-benefício entre os objetivos considerados, permitindo a seleção da solução mais adequada com base em critérios específicos do problema ou preferências do usuário (Akhter et al., 2024).

O NL-TSDR exemplifica esta abordagem ao combinar transformações não lineares com otimização multiobjetivo para simultaneamente maximizar a separabilidade entre classes e minimizar a dispersão intraclasse, resultando em representações de dados que facilitam significativamente o processo de classificação em cenários de alta dimensionalidade.

2.3 Algoritmos de Otimização

A otimização constitui um elemento no desenvolvimento e treinamento de algoritmos de aprendizado de máquina, particularmente em problemas de classificação. Conforme discutido na seção anterior sobre otimização multiobjetivo, a maioria dos classificadores modernos depende da minimização ou maximização de funções objetivo para ajustar seus parâmetros de forma a obter o melhor desempenho possível nos dados de treinamento (NOCEDAL; WRIGHT, 2006). Este processo de otimização pode assumir diferentes formas, desde a minimização de funções de custo quadráticas em regressão linear até a maximização de funções de verossimilhança em modelos probabilísticos ou ainda a minimização de funções de perda mais complexas em redes neurais. Naturalmente, é um processo que pode demandar uma grande quantidade de recursos computacionais.

A escolha do algoritmo de otimização adequado influencia diretamente tanto a qualidade da solução encontrada quanto o tempo computacional necessário para o treinamento. Neste contexto, é essencial compreender as características e limitações dos diferentes métodos disponíveis, bem como suas aplicabilidades específicas aos problemas de classificação. Os algoritmos de otimização podem ser classificados em métodos determinísticos, que seguem uma trajetória bem definida em direção ao ótimo, e métodos estocásticos (não determinísticos), que incorporam elementos aleatórios em sua busca.

Entre os métodos determinísticos, destacam-se aqueles baseados em gradientes e métodos que não requerem informações sobre derivadas. O método de programação quadrática sequencial (SQP ou *Sequential Quadratic Programming*) é implementação para problemas de otimização não linear com restrições (BYRD; HRIBAR; NOCEDAL, 1999). Ele aborda problemas de otimização não linear transformando-os iterativamente em subproblemas de programação quadrática. Esta abordagem é especialmente eficaz quando as restrições do problema são suaves e quando uma boa estimativa inicial está disponível. Em problemas de classificação, isto se traduz em convergência rápida para classificadores onde a função objetivo e as restrições possuem propriedades matemáticas favoráveis, como convexidade local.

Quando as restrições não são um fator limitante, o método Simplex de Nelder-Mead (SME) oferece uma alternativa eficaz baseada. Este método possui a vantagem distinta de não requerer informações sobre gradientes, tornando-o aplicável a funções objetivo que são descontínuas, ruidosas ou para as quais o cálculo de derivadas é computacionalmente proibitivo (LAGARIAS et al., 1998). A simplicidade conceitual deste método, combinada com sua robustez em problemas de baixa a moderada dimensionalidade, o torna uma escolha popular para otimização exploratória e para problemas onde a suavidade da função objetivo não pode ser garantida. Em problemas de classificação, esta característica é valiosa quando a função de custo incorpora métricas discretas ou quando procedimentos de validação cruzada são integrados diretamente na função objetivo.

Entretanto, tanto o SQP quanto o SME compartilham uma potencial limitação: sua natureza determinística os torna suscetíveis a convergir para ótimos locais em vez do ótimo global. Esta limitação pode ser particularmente problemática em problemas de classificação onde múltiplos mínimos locais podem existir, alguns dos quais podem corresponder a soluções de qualidade inferior (AZEVEDO; ROCHA; PEREIRA, 2024).

Para superar esta limitação, métodos de otimização global baseados em metaheurísticas têm ganhado crescente atenção na comunidade de aprendizado de máquina. Entre estes, os algoritmos genéticos (GA ou *genetic algorithms*) representam uma das abordagens mais estabelecidas utilizadas (GOLDBERG, 1989). Eles geralmente são baseados nos princípios da evolução natural, utilizando conceitos como seleção, cruzamento e mutação para explorar o espaço de busca.

O funcionamento do algoritmo genético inicia com a criação de uma população inicial de soluções candidatas, representadas como cromossomos que codificam os parâmetros do problema de otimização (KATOCH; CHAUHAN; KUMAR, 2021). A cada geração, os indivíduos são avaliados através da função objetivo, e aqueles com melhor aptidão têm maior probabilidade de serem selecionados para reprodução. O operador de cruzamento combina características de dois ou mais pais para gerar descendentes, enquanto o operador de mutação introduz variações aleatórias para manter a diversidade genética da população.

Esta abordagem populacional oferece várias vantagens significativas para problemas de classificação. Primeiramente, a natureza estocástica do algoritmo permite explorar múltiplas regiões do espaço de busca simultaneamente, reduzindo substancialmente o risco de convergência prematura para ótimos locais. Segundo, a diversidade mantida na população através dos operadores genéticos facilita a descoberta de soluções inovadoras que métodos determinísticos poderiam não encontrar. Terceiro, algoritmos genéticos são naturalmente adequados para problemas com múltiplos objetivos, uma característica valiosa em classificação onde frequentemente deseja-se otimizar simultaneamente precisão (*precision*), *recall* e generalização (YOKOYAMA; FERRO; SCHULZE, 2024).

A eficácia dos algoritmos genéticos em problemas de classificação tem sido demonstrada em diversos contextos, desde a seleção de características até a otimização de hiperparâmetros em redes neurais. A capacidade de trabalhar com representações discretas, contínuas ou mistas torna esta abordagem particularmente flexível para diferentes tipos de problemas de classificação.

No contexto prático da implementação, a escolha entre estes diferentes algoritmos de otimização deve considerar múltiplos fatores (RIOS; SAHINIDIS, 2013). Para problemas “bem comportados” com funções objetivo suaves e restrições lineares ou convexas, métodos determinísticos como o SQP geralmente oferecem convergência rápida e soluções de alta precisão. A disponibilidade de informações sobre gradientes e a estrutura matemática do problema são fatores decisivos na seleção do algoritmo específico dentro desta categoria.

Quando a função objetivo apresenta características que impedem o uso eficaz de métodos baseados em gradientes, o SME oferece uma alternativa robusta, embora com convergência potencialmente mais lenta (LAGARIAS et al., 1998). Este método é particularmente útil em fases exploratórias da modelagem ou quando uma compreensão preliminar do comportamento da função objetivo está sendo desenvolvida.

Para problemas onde a presença de múltiplos ótimos locais é suspeita ou confirmada, algoritmos genéticos representam a escolha mais apropriada, apesar do maior custo computacional (KATOCH; CHAUHAN; KUMAR, 2021). A capacidade de exploração global compensa este custo adicional quando a qualidade da solução é prioritária sobre a eficiência computacional.

A integração destes diferentes métodos em uma estratégia híbrida também representa uma abordagem promissora (AZEVEDO; ROCHA; PEREIRA, 2024). Por exemplo, um algoritmo genético pode ser utilizado para identificar regiões promissoras do espaço de busca, seguido pela aplicação de um método determinístico para refinamento local da solução. Esta combinação aproveita as vantagens de exploração global dos métodos estocásticos com a precisão e eficiência de convergência local dos métodos determinísticos.

Para facilitar a seleção adequada entre estes algoritmos em aplicações práticas de classificação, a Tabela 2.4 apresenta uma síntese comparativa das principais características dos métodos discutidos. Esta comparação considera aspectos fundamentais como o tipo de busca realizada, requisitos computacionais, capacidade de lidar com restrições, e adequação para diferentes tipos de problemas de otimização encontrados em tarefas de classificação.

Tabela 2.4 – Comparação entre algoritmos de otimização

Característica	SQP	SME	GA
Tipo de busca	Local	Local	Global
Suporte a restrições	Sim	Não	Sim
Complexidade computacional	Moderada	Baixa	Alta
Robustez a ruído	Baixa	Moderada	Alta
Convergência	Rápida	Moderada	Lenta
Adequado para multimodal	Não	Não	Sim
Determinístico	Sim	Sim	Não

Fonte: Elaborado pelo autor com base em (BYRD; HRIBAR; NOCEDAL, 1999), (LAGARIAS et al., 1998) e (GOLDBERG, 1989).

Em suma, a compreensão das características e limitações destes algoritmos de otimização pode orientar o desenvolvimento de soluções de classificação. A escolha do método relaciona-se com o equilíbrio entre tempo de processamento e desempenho da solução obtida. Essa consideração ganha pertinência à medida que os problemas de classificação aumentam em complexidade e escala.

2.4 Métodos Clássicos de Classificação

A seleção de algoritmos de classificação para avaliação comparativa de novos métodos constitui uma etapa no desenvolvimento de pesquisas em aprendizado de máquina. Os métodos apresentados nesta seção foram escolhidos por representarem diferentes paradigmas de classificação e por serem reconhecidos como referências confiáveis na literatura. Cada um possui características distintas em termos de complexidade computacional e desempenho em diferentes tipos de conjuntos de dados.

A organização dos métodos segue uma progressão lógica da simplicidade conceitual para a complexidade algorítmica (UDDIN et al., 2019). Inicia-se com o *Naive Bayes*, que oferece uma base para compreensão da classificação probabilística e serve como ponto de partida para discussões mais avançadas. O k -NN representa a categoria de métodos baseados em distâncias, caracterizados pela sua natureza não paramétrica e simplicidade de implementação. A LDA exemplifica as abordagens de separação linear, demonstrando tanto as vantagens quanto as limitações dos métodos lineares em espaços de alta dimensionalidade. Por fim, o MLP introduz conceitos de modelagem não linear através de redes neurais, estabelecendo uma ponte conceitual para compreender as transformações não lineares empregadas no método NL-TSDR.

Para facilitar a compreensão das características distintivas de cada método e suas implicações no contexto desta pesquisa, a [Tabela 2.5](#) apresenta uma síntese comparativa dos algoritmos selecionados. Esta comparação evidencia as diferenças fundamentais em termos de paradigma de aprendizado, natureza paramétrica, complexidade computacional e principais características operacionais, fornecendo um panorama abrangente que subsidia a análise experimental subsequente.

Cada um destes algoritmos apresenta comportamentos distintos quando confrontado com problemas de alta dimensionalidade e mascaramento de classes, fornecendo assim uma base comparativa para avaliar a eficácia das transformações propostas nesta tese. A análise de suas limitações específicas contribui para justificar a necessidade de desenvolvimento de novos métodos capazes de lidar mais efetivamente com os desafios inerentes à classificação em espaços de alta dimensionalidade.

2.4.1 *Naive Bayes*

O classificador *Naive Bayes* representa um dos métodos clássicos em aprendizado de máquina, baseando-se nos princípios da teoria de probabilidade bayesiana para realizar predições de classe (ALPAYDIN, 2021). Para compreender este algoritmo, é essencial primeiro entender o teorema de Bayes, nomeado em homenagem ao reverendo Thomas Bayes, que estabelece a relação entre probabilidades condicionais conforme a [Equação 2.1](#).

Tabela 2.5 – Comparação entre os métodos de classificação de referência

Método	Paradigma	Natureza	Características
<i>Naive</i> Bayes	Probabilístico	Paramétrico	Assume independência condicional; robusto a overfitting; eficiente com poucos dados
<i>k</i> -NN	Baseado em distâncias	Não paramétrico	Lazy learning; sensível à maldição da dimensionalidade; não requer treinamento
LDA	Separação linear	Paramétrico	Assume distribuições gaussianas; redução de dimensionalidade; limitado a fronteiras lineares
MLP	Redes neurais	Paramétrico	Aproximação universal; capacidade não linear; propenso a overfitting

Fonte: Elaborado pelo autor com base em (ALPAYDIN, 2021).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

Onde:

- $P(A)$ é a probabilidade anterior (*prior*) da hipótese A ;
- $P(B)$ é a probabilidade da evidência;
- $P(A|B)$ é a probabilidade posterior da hipótese dado a evidência;
- $P(B|A)$ é a verossimilhança da evidência dada a hipótese.

O classificador opera sob a premissa de independência condicional entre as características, assumindo que o valor de uma característica particular é independente do valor de qualquer outra característica, dada a classe da instância. Esta suposição “ingênua” (*naive*) confere o nome ao algoritmo, pois raramente é válida na prática. Contudo, esta simplificação permite que o método calcule eficientemente a probabilidade posterior de cada classe possível, selecionando aquela com maior probabilidade (*maximum a posteriori*) como predição final (ALPAYDIN, 2021).

Existem diferentes implementações do classificador adaptadas aos tipos de dados. O *Naive* Bayes gaussiano assume que os valores associados a cada classe seguem distribuição normal. O *Naive* Bayes multinomial é adequado para dados que seguem distribuição multinomial, frequentemente utilizado em processamento de texto. Já o *Naive* Bayes de Bernoulli é aplicado quando as características são binárias.

Uma vantagem do método reside na sua simplicidade conceitual e eficiência computacional. O algoritmo requer apenas uma passagem pelos dados de treinamento para estimar as probabilidades necessárias, resultando em tempos de treinamento baixos mesmo para conjuntos de dados grandes. Adicionalmente, demonstra desempenho surpreendentemente bom em muitas aplicações práticas, servindo como referência para comparação.

A violação da suposição de independência condicional não impede necessariamente o bom desempenho do classificador, pois o método requer apenas que a classe com maior probabilidade posterior seja corretamente identificada, não que as probabilidades estimadas sejam precisas em termos absolutos (DOMINGOS; PAZZANI, 1997).

No contexto de conjuntos de dados de alta dimensionalidade, o *Naive Bayes* apresenta comportamento interessante: embora a suposição de independência possa ser mais severamente violada em espaços complexos, o método pode beneficiar-se da abundância de características informativas, desde que cada uma contribua individualmente para a discriminação entre classes (RENNIE et al., 2003).

Apesar de sua simplicidade, uma limitação do algoritmo é sua potencial incapacidade de aprender relacionamentos entre características, o que pode ser importante em domínios onde tais interações são fundamentais para a classificação. Sua robustez e eficiência fazem do *Naive Bayes* uma ferramenta valiosa tanto para aplicações práticas quanto para o ensino de conceitos probabilísticos em classificação.

2.4.2 *k-Nearest Neighbors*

O *k*-NN opera sob o princípio intuitivo de que instâncias similares tendem a pertencer à mesma classe. Este método baseia-se na proximidade entre pontos de dados no espaço de características, classificando novas instâncias através da análise da classe majoritária entre seus *k* vizinhos mais próximos.

Formalmente, dado um conjunto de treinamento $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ contendo n exemplos de treinamento, onde cada $\mathbf{x}_i \in \mathbb{R}^d$ representa um vetor de características no espaço euclidiano de d dimensões e y_i a classe correspondente, o algoritmo *k*-NN classifica uma nova instância \mathbf{x}_q através de um processo sequencial (ABDALLA; AMER, 2022).

O processo de classificação inicia-se com o cálculo da distância $d(\mathbf{x}_q, \mathbf{x}_i)$ entre a instância de consulta \mathbf{x}_q e cada ponto de treinamento \mathbf{x}_i , utilizando uma métrica de distância apropriada. Na sequência, identifica-se o conjunto $\mathcal{N}_k(\mathbf{x}_q)$ que contém os k pontos de treinamento mais próximos de \mathbf{x}_q . Por fim, a classe predita é determinada através de uma votação majoritária, conforme expressa pela Equação 2.2, onde \mathcal{C} representa o conjunto de classes possíveis e $\mathbb{I}(\cdot)$ é uma função de contagem que retorna 1 quando o vizinho \mathbf{x}_i pertence à classe c e 0 caso contrário, permitindo assim contar quantos dos k vizinhos pertencem a cada classe.

$$\hat{y} = \arg \max_{c \in \mathcal{C}} \sum_{\mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_q)} \mathbb{I}(y_i = c) \quad (2.2)$$

A característica não paramétrica do k -NN constitui uma de suas principais vantagens, uma vez que o algoritmo não impõe suposições restritivas sobre a forma funcional ou distribuição subjacente dos dados. Diferentemente de métodos paramétricos que assumem uma estrutura específica (como a linearidade na regressão linear ou a independência condicional no *Naive Bayes*), o k -NN não define um modelo matemático fixo a priori.

Esta propriedade permite que o método se adapte naturalmente às variações locais presentes no conjunto de treinamento, possibilitando a captura de fronteiras de decisão arbitrariamente complexas e não lineares. Tal flexibilidade torna o k -NN particularmente adequado para problemas de classificação onde as suposições paramétricas tradicionais podem ser inadequadas ou restritivas demais.

A implementação prática do k -NN requer a definição de uma métrica de distância apropriada para quantificar a proximidade entre instâncias (WEINBERGER; SAUL, 2009). A distância euclidiana, utilizada devido à sua simplicidade e interpretabilidade geométrica, é definida pela Equação 2.3, onde d representa a dimensionalidade do espaço de características, e x_{qj} e x_{ij} denotam a j -ésima componente dos vetores \mathbf{x}_q e \mathbf{x}_i , respectivamente.

$$d(\mathbf{x}_q, \mathbf{x}_i) = \sqrt{\sum_{j=1}^d (x_{qj} - x_{ij})^2} \quad (2.3)$$

Embora a distância euclidiana seja amplamente adotada, a escolha da métrica de distância deve considerar as características específicas dos dados e do problema. A distância de Manhattan (ou L_1), calculada como $\sum_{j=1}^d |x_{qj} - x_{ij}|$, mostra-se mais robusta a *outliers* e adequada para dados com características em escalas muito diferentes. A distância de Minkowski generaliza ambas as métricas anteriores através do parâmetro p , onde $p = 1$ corresponde à Manhattan e $p = 2$ à euclidiana. Para dados categóricos, a distância de Hamming conta o número de posições onde os atributos diferem, enquanto a distância de Mahalanobis considera a correlação entre variáveis, sendo particularmente útil quando as características apresentam diferentes variâncias ou são correlacionadas (BEYER et al., 1999).

A seleção adequada da métrica de distância constitui, portanto, um aspecto que pode influenciar significativamente a eficácia do classificador k -NN. Após o cálculo das distâncias e identificação dos k vizinhos mais próximos, a classe da nova instância é determinada pela classe que aparece com maior frequência entre estes vizinhos. Situações de empate podem ser resolvidas através de critérios predefinidos, como a seleção aleatória, ou pela consideração das distâncias relativas dos vizinhos empatados.

A seleção do parâmetro k influencia diretamente o comportamento e desempenho do classificador. Valores pequenos de k (particularmente $k = 1$) tendem a produzir fronteiras de decisão complexas e irregulares, que podem resultar em *overfitting* aos dados de treinamento e baixa capacidade de generalização. Por outro lado, valores grandes de k geram fronteiras de decisão mais suaves e regulares, mas podem suavizar excessivamente a estrutura dos dados, ignorando padrões locais.

Em aplicações envolvendo conjuntos de dados de alta dimensionalidade, o k -NN pode enfrentar desafios significativos. À medida que o número de dimensões cresce, as distâncias entre todos os pares de pontos tendem a se tornar progressivamente mais uniformes, comprometendo a capacidade do algoritmo de distinguir entre vizinhos verdadeiramente próximos e distantes. Este fenômeno pode resultar em degradação substancial do desempenho, especialmente na presença de características irrelevantes ou ruidosas que contribuem para a distorção do espaço de características.

O fenômeno de mascaramento de classes representa outro desafio particular para o k -NN, ocorrendo quando instâncias de diferentes classes se sobrepõem no espaço de características original. Nestes cenários, pontos que são geometricamente próximos podem pertencer a classes distintas, levando o algoritmo a tomar decisões de classificação baseadas em vizinhos semanticamente inadequados. Para mitigar este problema, transformações do espaço de características que melhorem a separabilidade entre classes, como técnicas de redução de dimensionalidade ou aprendizado de métricas, podem proporcionar melhorias significativas no desempenho do algoritmo.

2.4.3 Análise de Discriminante Linear

A LDA é um método de classificação que busca encontrar projeções lineares dos dados originais que maximizem a separação entre classes enquanto minimizando a variação dentro de cada classe. Ela representa uma abordagem estatística para classificação e redução de dimensionalidade, fundamentada na suposição de que os dados seguem distribuições gaussianas multivariadas com matrizes de covariância iguais entre classes. A LDA geralmente produz resultados satisfatórios ao lidar com problemas de classificação reais (XANTHOPOULOS; PARDALOS; TRAFALIS, 2013).

O princípio da LDA baseia-se na otimização de um critério objetivo que busca maximizar a razão entre a variância interclasse e a variância intraclasse. Considere um problema de classificação genérico onde um indivíduo aleatório x pertence a uma das k possíveis classes, cujas densidades de probabilidade sejam representadas por $f(x)$. Dessa maneira, um discriminante tenta dividir o conjunto de dados em k regiões distintas que representam todas as classes. Com essas regiões, a classificação por análise discriminante significa simplesmente que x é alocado para a classe j se x estiver na região j . Para realizar essa alocação, existem duas possibilidades:

- Regra da Máxima Verossimilhança: Ao assumir que cada classe pode ocorrer com igual probabilidade, então x é alocado para a classe j se $j = \operatorname{argmax}_i f_i(x)$;
- Regra Bayesiana: Ao assumir que as probabilidades a priori (π) da classe são conhecidas, então x é alocado para a classe j se $j = \operatorname{argmax}_i \pi_i f_i(x)$.

Ao assumir que os dados são provenientes de distribuição gaussiana multivariada, ou seja, a distribuição de x pode ser caracterizada por sua média (μ) e covariância (Σ), formas explícitas das regras de alocação acima podem ser obtidas. Seguindo a regra Bayesiana, os dados x são classificados para a classe j se ele tiver a maior verossimilhança entre todas as k classes para $i = 1, \dots, k$ seguindo a função discriminante $\delta = \log f_i(x) + \log \pi_i$. Em outras palavras, a função discriminante determina a distância de x até cada classe.

Dessa maneira, o limite que separa quaisquer duas classes ocorre onde duas funções discriminantes têm o mesmo valor. Ao assumir a covariância igual entre as k classes, ou seja, em vez de uma matriz de covariância por classe, mas todas as classes têm a mesma matriz de covariância, então é possível obter a função discriminante conforme a [Equação 2.4 \(HASTIE ROBERT TIBSHIRANI, 2009\)](#).

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{\mu_k^T \Sigma^{-1} \mu_k}{2} + \log \pi_k \quad (2.4)$$

Trata-se de uma função linear em x , assim, a fronteira de decisão entre qualquer par de classes é também uma função linear em x , justificando o nome do método: análise discriminante linear. Sem a suposição de covariância igual, o termo quadrático na probabilidade não se cancela, portanto, a função discriminante resultante é uma função quadrática, que dá origem a um segundo método, a análise discriminante quadrática.

A formulação matemática da LDA envolve a decomposição da matriz de dispersão total em componentes interclasse e intraclasse, seguida pela solução de um problema de autovalores generalizado. As direções discriminantes ótimas correspondem aos autovetores associados aos maiores autovalores da matriz resultante, fornecendo uma base de baixa dimensionalidade onde a separabilidade entre classes é maximizada.

Esta abordagem resulta em hiperplanos de decisão lineares que dividem o espaço de características em regiões correspondentes às diferentes classes. Na verdade, uma característica importante da LDA reside na sua capacidade de redução dimensional intrínseca.

Para um problema com c classes, o método pode produzir no máximo $c - 1$ direções discriminantes, independentemente da dimensionalidade original dos dados. Esta propriedade torna a LDA particularmente valiosa em cenários de alta dimensionalidade, onde a redução para um espaço de menor dimensionalidade pode eliminar características redundantes ou ruidosas ([SHARMA; PALIWAL, 2015](#)).

As vantagens da LDA incluem sua fundamentação estatística sólida, eficiência computacional e capacidade de fornecer uma representação de baixa dimensionalidade otimizada para discriminação entre classes. O método é determinístico, produzindo resultados consistentes e reproduzíveis, e oferece uma interpretação geométrica clara das direções mais discriminativas nos dados. No entanto, as limitações da LDA tornam-se aparentes quando suas suposições fundamentais são violadas.

A suposição de igualdade das matrizes de covariância entre classes raramente é satisfeita na prática, especialmente em conjuntos de dados complexos. Adicionalmente, a natureza linear do método limita sua capacidade de modelar relações não lineares entre características e classes, tornando-o inadequado para problemas onde as fronteiras de decisão ótimas são intrinsecamente não lineares.

Em contextos de alta dimensionalidade, a LDA pode enfrentar problemas de singularidade quando o número de características excede o número de amostras de treinamento, situação conhecida como *small sample size problem*. Este cenário requer técnicas de regularização ou métodos alternativos para garantir estabilidade numérica na solução do problema de autovalores.

2.4.4 *Multilayer Perceptron*

O Perceptron de Múltiplas Camadas ou *Multilayer Perceptron* (MLP) é uma Rede Neural Artificial (RNA) caracterizada por possuir pelo menos uma camada intermediária de neurônios, situada entre a de entrada e a de saída (HAYKIN, 2007). Esta arquitetura representa uma das implementações fundamentais de redes neurais artificiais, capaz de aproximar funções não lineares complexas através da composição de múltiplas camadas de neurônios interconectados. Diferentemente do perceptron simples, o MLP pode resolver problemas de classificação que envolvem funções não lineares, tornando-se particularmente adequado para problemas onde as relações entre características e classes são intrinsecamente não lineares ou onde métodos lineares demonstram limitações significativas.

Uma RNA é um modelo computacional inspirado pelo sistema nervoso central de um animal, capaz de realizar aprendizado de máquina e reconhecimento de padrões. As RNAs implementam modelos matemáticos inspirados na estrutura neural de organismos inteligentes e adquirem conhecimento por meio da experiência. Independentemente do seu tipo, são basicamente representadas por um sistema de unidades de processamento interconectadas que, a partir de um determinado conjunto de valores de entrada, computam valores de saída, simulando o comportamento de redes neurais biológicas. O psiquiatra Warren McCulloch e o matemático Walter Pitts, em 1943, foram os primeiros pesquisadores a descreverem um modelo artificial para um neurônio biológico, descrito no artigo *A Logical Calculus of the Ideas Immanent in Nervous Activity* (MCCULLOCH; PITTS, 1943).

Este modelo pioneiro estabeleceu os princípios fundamentais que evoluíram para as capacidades avançadas do MLP atual, sendo sintetizado da seguinte forma: sinais são apresentados à entrada, cada um multiplicado por um peso que indica sua influência na saída do neurônio; é feita a soma ponderada dos sinais que produz um nível de atividade; e se este nível exceder um certo limite, o neurônio produz uma determinada resposta de saída (HAYKIN, 2007). O modelo é formado por um vetor de entradas onde as sinapses são representadas por pesos numéricos, e a soma ponderada das entradas é submetida a uma função de ativação que determina se essa é maior que um determinado valor, garantindo que o neurônio sempre esteja em um de dois possíveis estados: ativado ou desativado.

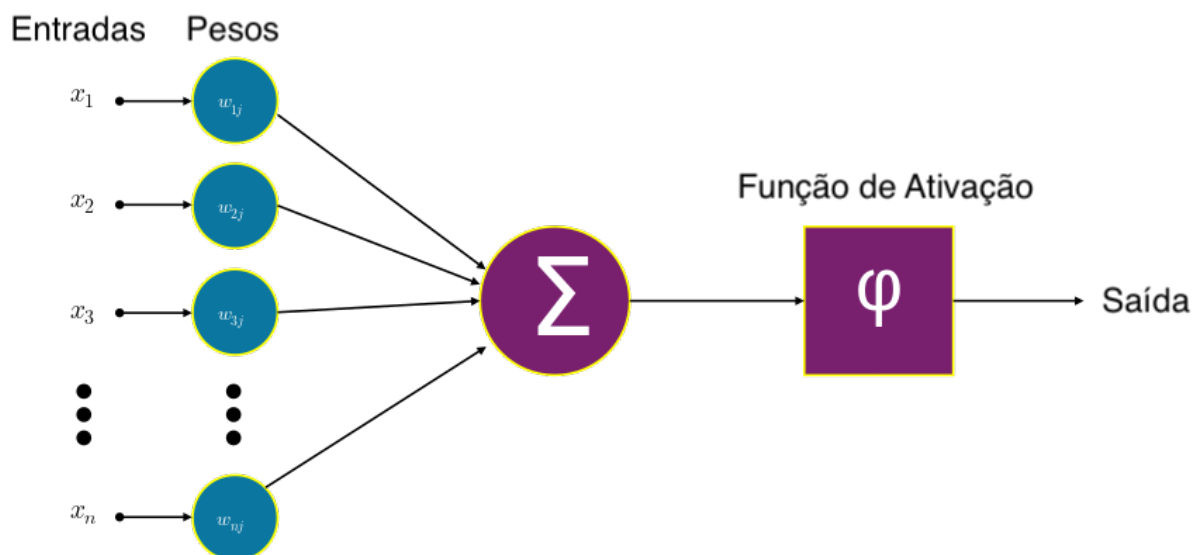
Na Figura 2.5, ilustra-se de forma esquemática a arquitetura do neurônio artificial proposto por McCulloch e Pitts, onde podem ser identificados claramente os componentes fundamentais do modelo. As entradas $I_1, I_2, I_3, \dots, I_N$ representam os sinais de entrada que chegam ao neurônio, simulando os impulsos elétricos recebidos pelos dendritos de um neurônio biológico. Cada entrada é multiplicada pelo seu respectivo peso sináptico $W_1, W_2, W_3, \dots, W_N$, que modula a intensidade do sinal, representando a força da conexão sináptica.

O nódulo central com o símbolo de somatória (Σ) realiza a operação de soma ponderada de todas as entradas multiplicadas pelos seus pesos, calculando o potencial de ativação do neurônio. Este valor resultante é então processado pela função de ativação, representada pelo símbolo em formato de degrau, que implementa o comportamento limiar do modelo: se o valor da soma ponderada exceder um determinado *threshold*, o neurônio é ativado e produz uma saída $y = 1$; caso contrário, permanece inativo com saída $y = 0$, caracterizando assim o comportamento binário que define os dois estados possíveis do neurônio artificial (HAYKIN, 2007).

A arquitetura básica do MLP, representada na Figura 2.6, consiste em uma camada de entrada que recebe as características dos dados, uma ou mais camadas ocultas (*hidden layers*), e uma camada de saída que produz as previsões finais. Cada neurônio em uma camada está conectado a todos os neurônios da camada seguinte, com conexões ponderadas por pesos sinápticos que são ajustados durante o processo de treinamento. Na camada de entrada, os padrões são apresentados à rede; na porção intermediária é realizada a maior parte do processamento; e na camada de saída, o resultado é apresentado.

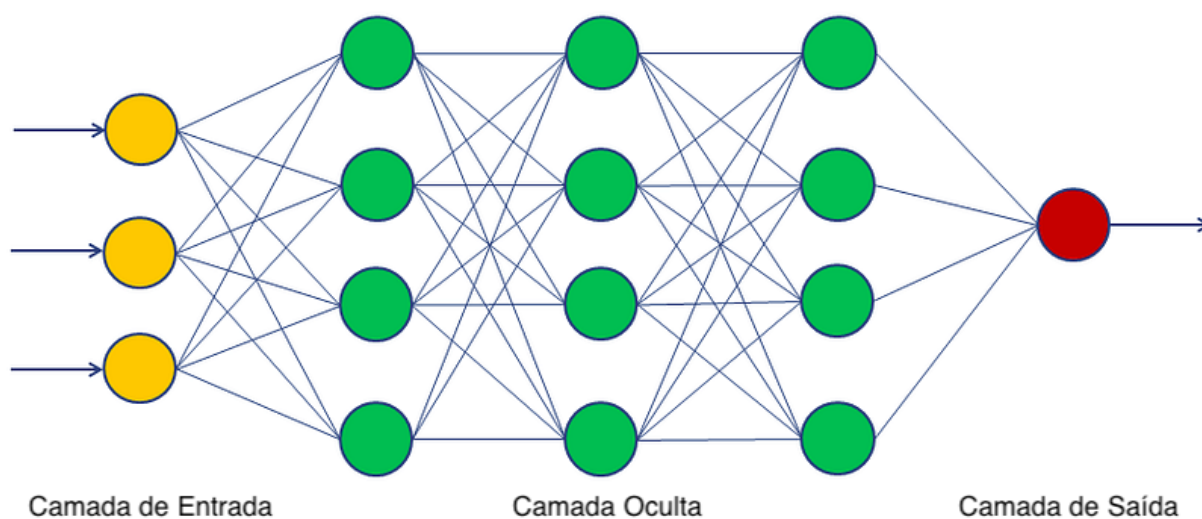
Uma das propriedades do MLP é sua capacidade de aproximação universal, que estabelece teoricamente que uma rede com uma única camada oculta contendo um número suficiente de neurônios pode aproximar qualquer função contínua com precisão arbitrária. Esta propriedade fornece a base teórica para a eficácia do MLP em uma ampla variedade de problemas de classificação, incluindo aqueles com fronteiras de decisão altamente complexas, permitindo que a rede identifique padrões complexos e interações entre características que podem não ser aparentes através de métodos lineares (HAYKIN, 2007).

Figura 2.5 – Representação esquemática de um neurônio artificial



Fonte: Adaptado pelo autor de (SASSON, 2023).

Figura 2.6 – Representação esquemática de um MLP



Fonte: Adaptado pelo auto de (NGO et al., 2022).

As RNAs contam com regras de treinamento onde os pesos de suas conexões são ajustados de acordo com os padrões de referência apresentados, significando que elas aprendem por intermédio de exemplos. A característica de maior destaque de uma RNA é sua habilidade de aprendizado em um determinado ambiente e sua consequente melhora de desempenho, que acontece em função do treinamento, um longo processo iterativo de ajustes aplicado aos seus pesos.

O treinamento do MLP é realizado tipicamente pelo algoritmo da retropropagação (*backpropagation*) (LAWRENCE; GILES, 2000), que utiliza a regra da cadeia do cálculo diferencial para computar os gradientes da função de erro em relação aos pesos da rede. Estes gradientes são então utilizados por algoritmos de otimização, como o gradiente descendente estocástico, para ajustar iterativamente os pesos da rede de forma a minimizar o erro de classificação nos dados de treinamento.

Em sua primeira fase, chamada de propagação adiante (*forward*), sinais de entrada são inseridos na rede e propagados até produzir as respectivas saídas, buscando obter as respostas atuais da rede considerando os pesos sinápticos configurados atualmente. Na segunda fase, denominada de propagação reversa (*backward*), as saídas obtidas são comparadas com as saídas desejadas, os erros em cada sinal são calculados e utilizados para atualizar os pesos sinápticos da rede, até que uma condição de parada seja atingida (LAWRENCE; GILES, 2000).

A natureza não determinística do MLP deriva principalmente da inicialização aleatória dos pesos da rede e do uso de algoritmos de otimização estocásticos. Esta aleatoriedade significa que múltiplas execuções do algoritmo podem produzir modelos com desempenhos ligeiramente diferentes, requerendo técnicas apropriadas de validação estatística para avaliação do método.

Em problemas de alta dimensionalidade, o MLP demonstra tanto vantagens quanto desvantagens. Por um lado, a capacidade de modelagem não linear permite identificar padrões complexos; por outro lado, o grande número de parâmetros da rede pode levar ao *overfitting*, especialmente quando o número de amostras de treinamento é limitado em relação à dimensionalidade dos dados.

As funções de ativação são parte da arquitetura do MLP, determinando como cada neurônio processa e transforma os sinais de entrada em sinais de saída. Entre as funções mais utilizadas, destacam-se a função sigmoide, que mapeia qualquer valor real para o intervalo entre 0 e 1 e foi historicamente uma das primeiras a serem amplamente adotadas; a função tangente hiperbólica (*tanh*), que produz saídas no intervalo entre -1 e 1 e frequentemente demonstra convergência mais rápida que a sigmoide; e a função ReLU (*Rectified Linear Unit*), definida como $f(x) = \max(0, x)$, que se tornou predominante em redes profundas devido à sua simplicidade computacional e capacidade de mitigar o problema do desvanecimento do gradiente (HAYKIN, 2007).

Por se tratar de uma rede neural de múltiplas camadas, a configuração arquitetural do MLP torna-se particularmente sensível às características do conjunto de dados utilizado. A escolha da quantidade adequada de neurônios para cada camada representa um dos aspectos mais desafiadores no projeto de MLPs, sendo frequentemente caracterizada pela ausência de diretrizes definitivas na literatura especializada.

Contudo, algumas heurísticas práticas podem orientar essa decisão (PANCHAL; PANCHAL, 2014): o número de neurônios na camada oculta deve estar situado no intervalo entre o tamanho da camada de entrada e o tamanho da camada de saída; alternativamente, pode-se utilizar aproximadamente $2/3$ do tamanho da camada de entrada acrescido do tamanho da camada de saída; ou ainda, o número de neurônios ocultos deve ser mantido abaixo do dobro do tamanho da camada de entrada. Essas diretrizes, embora não sejam regras absolutas, fornecem pontos de partida úteis para a experimentação, considerando que a configuração ótima frequentemente requer validação empírica através de técnicas como validação cruzada e análise de curvas de aprendizado para encontrar o equilíbrio adequado entre capacidade de representação e generalização.

O mascaramento de classes representa um desafio particular para o MLP, pois a sobreposição de instâncias de diferentes classes pode confundir o processo de otimização, resultando em convergência para mínimos locais subótimos. Nestes cenários, técnicas de pré-processamento que melhorem a separabilidade dos dados podem significativamente beneficiar o desempenho do classificador. A flexibilidade arquitetural do MLP permite diversas configurações em termos de número de camadas ocultas, número de neurônios por camada, e funções de ativação utilizadas. Esta flexibilidade representa tanto uma vantagem, permitindo adaptação a diferentes tipos de problemas, quanto um desafio, requerendo seleção cuidadosa de hiperparâmetros para obtenção de desempenho ótimo, demonstrando a importância de abordagens integradas que combinem arquiteturas robustas com estratégias adequadas de preparação dos dados.

2.5 Métricas de Avaliação

A avaliação adequada de métodos de classificação e transformação de dados requer um conjunto abrangente de métricas que possam quantificar tanto a qualidade das transformações aplicadas aos dados quanto o desempenho dos algoritmos de classificação resultantes.

Esta seção apresenta as principais métricas utilizadas nesta pesquisa, organizadas em duas categorias fundamentais: métricas de qualidade dos dados, que avaliam as propriedades dos conjuntos de dados após transformações, e métricas de desempenho de classificação, que quantificam a eficácia dos algoritmos em tarefas de classificação (SOKOLOVA; LAPALME, 2009).

As métricas de qualidade dos dados assumem particular importância no contexto do aprendizado de máquina, pois permitem avaliar objetivamente se as transformações aplicadas preservam ou melhoram as propriedades discriminativas dos dados originais. Estas métricas fornecem entendimento sobre a estrutura dos dados transformados e sua adequação para tarefas de classificação subsequentes.

Para facilitar a compreensão e comparação das métricas utilizadas, a [Tabela 2.6](#) e a [Tabela 2.7](#) apresentam uma síntese das principais características de cada métrica, incluindo suas faixas de valores e interpretações. Esta organização sistemática permite uma visão consolidada das ferramentas de avaliação empregadas e orienta a interpretação adequada dos resultados obtidos em cada etapa da metodologia proposta.

Tabela 2.6 – Comparação das métricas de qualidade dos dados

Coeficiente	Propósito	Intervalo	Interpretação
Pearson	Mede a força e direção da relação linear entre variáveis	$[-1, 1]$	Valores próximos a ± 1 indicam forte correlação linear
Silhouette	Avalia a qualidade da separação entre <i>clusters</i> e coesão interna	$[-1, 1]$	Valores próximos a 1 indicam boa separação e coesão

Fonte: Elaborado pelo autor com base em (SOKOLOVA; LAPALME, 2009).

Tabela 2.7 – Comparação das métricas de desempenho de classificação

Métrica	Propósito	Intervalo	Interpretação
Acurácia	Proporção total de predições corretas	$[0, 1]$	Valores próximos a 1 indicam alta precisão geral
<i>Precision</i>	Proporção de verdadeiros positivos entre as predições positivas	$[0, 1]$	Valores altos reduzem falsos positivos
<i>Recall</i>	Proporção de verdadeiros positivos identificados corretamente	$[0, 1]$	Valores altos reduzem falsos negativos
F1 <i>Score</i>	Média harmônica entre <i>Precision</i> e <i>Recall</i>	$[0, 1]$	Valores próximos a 1 indicam bom equilíbrio

Fonte: Elaborado pelo autor com base em (SOKOLOVA; LAPALME, 2009).

A seleção das métricas apresentadas foi orientada pelos objetivos específicos da metodologia NL-TSDR, que busca simultaneamente reduzir a dimensionalidade dos dados e melhorar a separabilidade entre classes. Neste contexto, torna-se essencial avaliar tanto a preservação de características discriminativas quanto a qualidade da organização resultante dos dados no espaço transformado.

As métricas de qualidade dos dados desempenham contribuições para a validação de métodos de transformação e redução dimensional, fornecendo medidas quantitativas que permitem avaliar se as modificações aplicadas aos dados originais resultam em representações mais adequadas para classificação. Estas métricas são essenciais para compreender como transformações específicas afetam a estrutura inerente dos dados e sua capacidade discriminativa.

No contexto de conjuntos de dados de alta dimensionalidade, as métricas de qualidade assumem importância ainda maior, pois permitem identificar se as transformações conseguem efetivamente mitigar problemas como mascaramento de classes e redundância de características. A avaliação adequada destas propriedades pode contribuir para a determinação da eficácia de métodos de redução dimensional em preservar ou melhorar a separabilidade entre diferentes categorias de dados.

As duas métricas apresentadas nesta subseção foram especificamente selecionadas por sua complementaridade na avaliação de diferentes aspectos da qualidade dos dados transformados. Enquanto o Coeficiente de Correlação de Pearson quantifica as relações lineares entre características, o Coeficiente de *Silhouette* avalia a qualidade da organização global dos dados em termos de separabilidade e coesão entre grupos.

2.5.1 Coeficiente de Correlação de Pearson

O Coeficiente de Correlação de Pearson constitui uma medida estatística para quantificar a força e direção das relações lineares entre pares de variáveis contínuas (SOKOLOVA; LAPALME, 2009). No contexto de avaliação de métodos de redução dimensional, esta métrica assume importância particular por permitir a análise de como as transformações aplicadas afetam as correlações entre características e da preservação de relações discriminativas nos dados transformados.

A formulação matemática do coeficiente baseia-se na covariância normalizada entre duas variáveis, resultando em um valor que varia entre -1 e +1. O Coeficiente de Correlação de Pearson (ρ) denota o grau de relação entre duas variáveis quantitativas e representa seu grau de correlação ao longo de uma escala de valores situados entre -1 e 1.

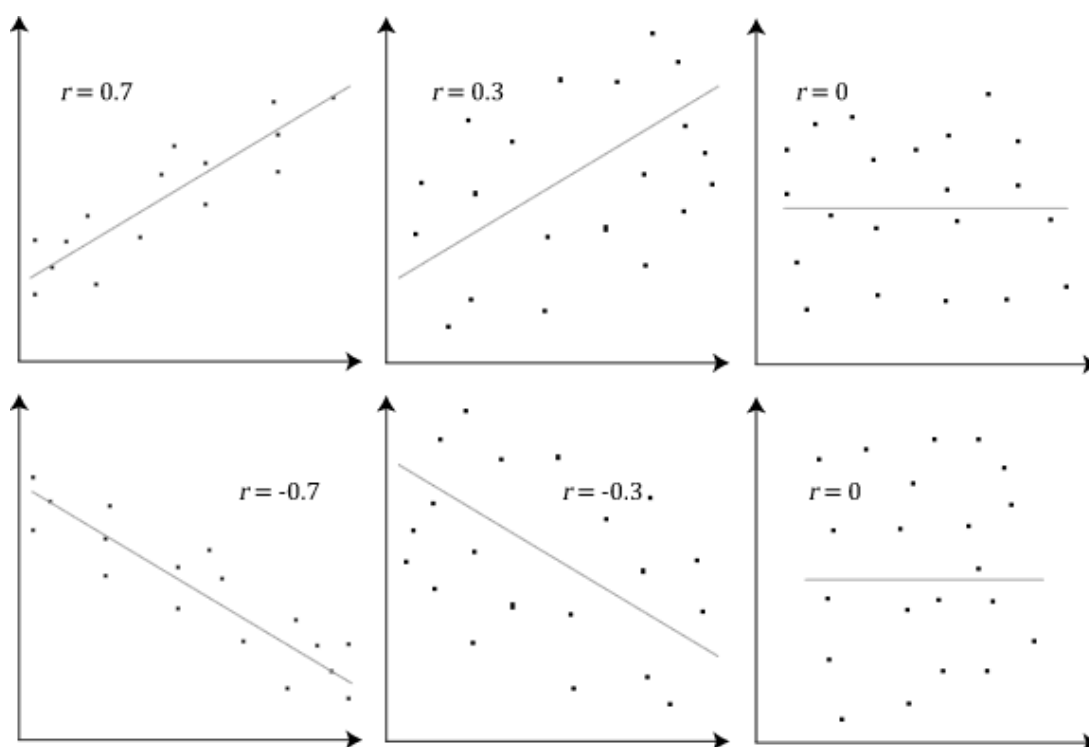
Considerando que duas variáveis A e B possuem, cada, N observações escalares, então $\rho(A, B)$ pode ser definido conforme a Equação 2.5, onde μ e σ são, respectivamente, a média e o desvio padrão de cada variável (BENESTY et al., 2009). Esta normalização torna a métrica independente das unidades de medida das variáveis analisadas, permitindo comparações consistentes entre diferentes conjuntos de dados e escalas.

$$\rho(A, B) = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right) \quad (2.5)$$

Na [Figura 2.7](#), encontram-se as possíveis interpretações do coeficiente, demonstrando como diferentes valores de r correspondem a padrões distintos de dispersão dos dados: correlações fortes ($r = \pm 0.7$) apresentam pontos bem alinhados à reta de tendência, correlações fracas ($r = \pm 0.3$) mostram maior dispersão, e a ausência de correlação ($r = 0$) resulta em distribuição aleatória dos pontos.

Quando o coeficiente de correlação se aproxima de 1, nota-se um aumento no valor de uma variável quando a outra também aumenta, ou seja, há uma relação linear positiva. Quando o coeficiente se aproxima de -1 , também é possível dizer que as variáveis são correlacionadas, mas nesse caso quando o valor de uma variável aumenta o da outra diminui, caracterizando uma correlação negativa ou inversa. Por sua vez, um coeficiente de correlação próximo de zero indica que não há relação linear evidente entre as duas variáveis, embora possam existir relações não lineares não detectadas por esta métrica.

Figura 2.7 – Exemplo do Coeficiente de Correlação de Pearson



Fonte: Adaptado pelo autor de ([INDARJO, 2024](#)).

No contexto específico da metodologia NL-TSDR, o Coeficiente de Correlação de Pearson é utilizado para evidenciar que o processo de redução dimensional resulta na preservação de características mais discriminativas no conjunto de dados transformado. Esta aplicação baseia-se na premissa de que características com maior capacidade discriminativa tendem a apresentar correlações mais fortes e consistentes com os rótulos de classe, enquanto características redundantes ou ruidosas apresentam correlações mais fracas.

A análise das correlações antes e após as transformações permite quantificar objetivamente se o método consegue efetivamente identificar e preservar as características mais relevantes para a tarefa de classificação. Um aumento nas correlações médias ou uma melhor organização das correlações entre características discriminativas e rótulos de classe pode ser interpretado como evidência de que o processo de transformação foi bem-sucedido em eliminar ruído e redundância dos dados originais.

Esta métrica é particularmente valiosa em cenários de alta dimensionalidade, onde a identificação manual de características discriminativas torna-se impraticável. Através da análise sistemática das correlações, é possível verificar se as transformações propostas conseguem efetivamente concentrar a informação discriminativa em um subconjunto menor de características, resultando em representações mais compactas e eficazes para classificação.

A interpretação adequada do Coeficiente de Correlação de Pearson no contexto de redução dimensional requer consideração cuidadosa de suas limitações. A métrica captura apenas relações lineares entre variáveis, podendo não detectar associações não lineares complexas que podem ser relevantes em alguns domínios de aplicação. Adicionalmente, correlações altas não implicam necessariamente causalidade, sendo importante interpretar os resultados em conjunto com outras métricas de qualidade dos dados e conhecimento do domínio. Outras limitações incluem a influência significativa de valores extremos (*outliers*) no cálculo do coeficiente e a maior significância da métrica quando as variáveis seguem distribuições aproximadamente normais. Portanto, embora seja uma ferramenta poderosa para avaliação de métodos de redução dimensional, o Coeficiente de Correlação de Pearson deve ser utilizado em conjunto com outras métricas complementares para uma análise mais abrangente dos resultados obtidos.

2.5.2 Coeficiente de *Silhouette*

O Coeficiente de *Silhouette* representa uma métrica utilizada para avaliar a qualidade de agrupamentos e a separabilidade geral entre grupos de dados, fornecendo uma medida quantitativa da adequação da organização dos dados em *clusters* ou classes. Esta métrica assume particular relevância no contexto de métodos de redução dimensional, onde o objetivo é preservar ou melhorar a estrutura natural de agrupamento dos dados.

O Coeficiente de *Silhouette* é utilizado para mensurar a qualidade de uma técnica de agrupamento (BONNIN, 2017). A formulação baseia-se na comparação entre a coesão interna de cada ponto de dados com seu grupo e a separação deste ponto em relação aos grupos vizinhos. O valor para cada ponto é uma medida de quão semelhante esse ponto é a outros pontos no mesmo *cluster*, em comparação com pontos em outros *clusters*. Para cada ponto, calcula-se a distância média para outros pontos do mesmo grupo (coesão interna) e a distância média para pontos do grupo mais próximo (separação externa).

O Coeficiente de *Silhouette* s_i para o ponto i é definido conforme a [Equação 2.6](#). Na equação, a_i é a distância média do ponto i aos outros pontos no mesmo *cluster* que i , e b_i é a distância média mínima do ponto i a pontos em um *cluster* diferente, minimizada sobre os *clusters*. O coeficiente de silhouette para cada ponto é então definido como a diferença entre estas medidas, normalizada pela maior delas.

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2.6)$$

Como pode ser observado, o Coeficiente de *Silhouette* é calculado ponto a ponto. Todavia, para critérios de análise, é considerado o valor médio das leituras encontradas. Os valores variam de -1 a 1 , onde valores próximos a $+1$ indicam que o ponto está bem agrupado em seu *cluster* e bem separado dos *clusters* vizinhos, valores próximos a zero sugerem que o ponto está próximo à fronteira entre *clusters*, e valores negativos indicam que o ponto pode ter sido atribuído ao *cluster* incorreto. Um valor alto indica que o ponto é bem compatível com seu próprio *cluster* e pouco compatível com os demais. Se a maioria dos pontos tiver um valor alto, a solução de clusterização é apropriada. Se muitos pontos tiverem um valor baixo ou negativo, a solução poderá ter muitos ou poucos *clusters*. A média dos coeficientes individuais fornece uma medida global da qualidade da organização dos dados. O uso desse coeficiente como critério de avaliação de agrupamento é um bom indicador da qualidade da técnica utilizada. De maneira geral, é possível se guiar pelas condições apresentadas na [Tabela 2.8](#).

Tabela 2.8 – Interpretação dos valores do Coeficiente de *Silhouette*

Valor	Interpretação
$s_i = 1$	Os <i>clusters</i> estão bem separados uns dos outros e claramente distinguidos
$s_i = 0$	Os <i>clusters</i> são indiferentes (a distância entre os <i>clusters</i> não é significativa)
$s_i = -1$	Os <i>clusters</i> são atribuídos de maneira errada

Fonte: Elaborado pelo autor com base em ([BONNIN, 2017](#)).

No contexto da metodologia NL-TSDR, o Coeficiente de *Silhouette* é empregado para avaliar como as transformações não lineares afetam a separabilidade entre classes no espaço transformado. Um aumento nos valores do coeficiente após as transformações indica que o método conseguiu efetivamente melhorar a organização dos dados, resultando em classes mais coesas internamente e melhor separadas entre si. A aplicação desta métrica é particularmente valiosa para validar a eficácia da segunda etapa do método NL-TSDR, que emprega otimização multiobjetivo para maximizar distâncias interclasse enquanto minimiza dispersão intraclasse. O Coeficiente de *Silhouette* fornece uma medida direta de quão bem estes objetivos foram alcançados, quantificando objetivamente a melhoria na separabilidade resultante das transformações aplicadas.

A interpretação dos resultados do Coeficiente de *Silhouette* deve considerar tanto os valores absolutos quanto as mudanças relativas em relação aos dados originais. Melhorias significativas no coeficiente após as transformações constituem evidência forte de que o método conseguiu efetivamente reduzir o mascaramento de classes e melhorar a estrutura discriminativa dos dados, facilitando assim o trabalho de algoritmos de classificação subsequentes. Uma potencial vantagem desta métrica reside na sua aplicabilidade independente de suposições sobre a distribuição dos dados ou a natureza dos agrupamentos. Esta flexibilidade torna esta métrica adequada para avaliar uma ampla variedade de transformações e tipos de dados, fornecendo uma medida da qualidade da organização resultante independentemente das características específicas do conjunto de dados analisado.

2.5.3 Métricas de Desempenho de Classificação

As métricas de desempenho de classificação constituem ferramentas fundamentais para avaliação quantitativa da eficácia de algoritmos de aprendizado de máquina, fornecendo medidas objetivas que permitem comparar diferentes métodos e validar melhorias propostas. Estas métricas são essenciais para determinar o sucesso de transformações aplicadas aos dados e para estabelecer referências confiáveis em pesquisas de desenvolvimento de novos algoritmos.

No contexto de conjuntos de dados de alta dimensionalidade e problemas de mascaramento de classes, a seleção adequada de métricas de desempenho torna-se particularmente crítica. Diferentes métricas podem revelar aspectos distintos do desempenho de um classificador, e a interpretação conjunta de múltiplas medidas fornece uma visão mais completa da eficácia do método avaliado. A escolha das métricas deve considerar tanto as características específicas do problema quanto os objetivos da pesquisa.

As três métricas apresentadas nesta subseção representam medidas aceitas na literatura de aprendizado de máquina. Sua aplicação sistemática permite estabelecer comparações entre o método NL-TSDR proposto e os algoritmos de referência, fornecendo evidências quantitativas da eficácia das transformações não lineares em melhorar o desempenho de classificação em cenários de alta dimensionalidade.

2.5.3.1 Acurácia

A acurácia (*accuracy*) representa a métrica de desempenho mais intuitiva e amplamente utilizada em problemas de classificação (SOKOLOVA; LAPALME, 2009), quantificando a proporção de predições corretas em relação ao total de predições realizadas. Esta medida fornece uma visão geral do desempenho do classificador, sendo particularmente útil como primeira aproximação para avaliar a eficácia de diferentes métodos e transformações aplicadas aos dados.

A formulação matemática da acurácia é direta, calculada como a razão entre o número de instâncias corretamente classificadas e o número total de instâncias no conjunto de teste (POWERS, 2020), conforme Equação 2.7.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.7)$$

Nesse sentido, VP representa os verdadeiros positivos, VN os verdadeiros negativos, FP os falsos positivos e FN os falsos negativos. Esta simplicidade conceitual torna a métrica facilmente interpretável e permite comparações diretas entre diferentes algoritmos, sendo frequentemente utilizada como critério primário de seleção de modelos em aplicações práticas.

No contexto da metodologia NL-TSDR, a acurácia serve como indicador principal da eficácia das transformações propostas em melhorar o desempenho de classificação. Melhorias na acurácia após a aplicação das transformações não lineares constituem evidência direta de que o método conseguiu efetivamente reduzir o mascaramento de classes e facilitar a tarefa de discriminação para os algoritmos de classificação subsequentes.

A interpretação da acurácia deve considerar as características específicas do conjunto de dados, particularmente em relação ao balanceamento entre classes. Em cenários onde as classes apresentam distribuições desbalanceadas, a acurácia pode fornecer uma visão distorcida do desempenho real, pois um classificador que simplesmente prediz sempre a classe majoritária pode alcançar acurácia aparentemente alta sem realmente aprender padrões discriminativos úteis.

Esta limitação torna-se especialmente relevante em conjuntos de dados de alta dimensionalidade, onde desbalanceamentos podem ser amplificados. Nestes cenários, a acurácia deve ser interpretada em conjunto com outras métricas que forneçam detalhes sobre o desempenho do classificador em cada classe específica.

Apesar destas limitações, a acurácia é uma métrica relevante, especialmente quando utilizada em conjunto com técnicas adequadas de validação cruzada. Sua simplicidade e interpretabilidade direta fazem dela uma ferramenta valiosa para comunicação de resultados e comparação de métodos em contextos de pesquisa e aplicação prática.

2.5.3.2 *Precision e Recall*

As métricas *Precision* e *Recall* fornecem uma análise mais granular do desempenho de classificação, focando especificamente na capacidade do algoritmo de identificar corretamente instâncias de classes específicas (POWERS, 2020). Estas métricas são particularmente valiosas em cenários onde diferentes tipos de erros possuem custos ou implicações distintas, permitindo uma avaliação mais nuançada da qualidade das predições realizadas.

Precision, também conhecida como valor preditivo positivo, quantifica a proporção de instâncias corretamente identificadas como pertencentes a uma classe específica em relação ao total de instâncias preditas para essa classe (SOKOLOVA; LAPALME, 2009), conforme Equação 2.8.

$$\text{Precision} = \frac{VP}{VP + FP} \quad (2.8)$$

Esta métrica responde à questão: das instâncias que o modelo identificou como pertencentes a esta classe, quantas realmente pertencem a ela? Um valor alto de *Precision* indica que o classificador produz poucos falsos positivos para a classe analisada.

Recall, também denominado sensibilidade ou taxa de verdadeiros positivos, mede a proporção de instâncias de uma classe específica que foram corretamente identificadas pelo classificador em relação ao total de instâncias realmente pertencentes a essa classe (POWERS, 2020), conforme Equação 2.9.

$$\text{Recall} = \frac{VP}{VP + FN} \quad (2.9)$$

Esta métrica aborda a questão: das instâncias que realmente pertencem a esta classe, quantas o modelo conseguiu identificar corretamente? Um valor alto de *Recall* indica que o classificador consegue detectar a maioria das instâncias da classe, produzindo poucos falsos negativos.

A relação entre *Precision* e *Recall* revela um delicado balanço em tarefas de classificação: geralmente, melhorias em uma métrica resultam em degradação da outra (FLACH; KULL, 2015). Classificadores mais conservadores tendem a apresentar alta *Precision* mas baixo *Recall*, enquanto classificadores mais liberais exibem o comportamento oposto. Esta diferença deve ser considerada ao interpretar o desempenho de diferentes métodos.

No contexto de conjuntos de dados de alta dimensionalidade afetados por mascaramento de classes, a análise de *Precision* e *Recall* por classe individual pode revelar padrões sobre como diferentes classes são afetadas pelo fenômeno. Classes com maior sobreposição no espaço original podem apresentar valores particularmente baixos de *Precision* e *Recall*, indicando maior dificuldade de discriminação (BEYER et al., 1999).

A aplicação das transformações NL-TSDR deve idealmente resultar em melhorias tanto de *Precision* quanto de *Recall* para todas as classes, indicando que o método conseguiu efetivamente reduzir o mascaramento sem favorecer especificamente algumas classes em detrimento de outras. A análise detalhada destas métricas por classe permitem avaliar a eficácia das transformações aplicadas.

A interpretação adequada de *Precision* e *Recall* requer consideração do contexto específico da aplicação. Em algumas situações, pode ser preferível otimizar *Recall* (minimizar falsos negativos) mesmo às custas de *Precision*, enquanto em outras aplicações o oposto pode ser verdadeiro. Esta flexibilidade interpretativa torna estas métricas particularmente valiosas para avaliação de métodos em diferentes contextos de aplicação.

2.5.3.3 F1 Score

O *F1 Score* representa uma métrica composta que combina *Precision* e *Recall* em uma única medida (POWERS, 2020), fornecendo uma avaliação equilibrada do desempenho de classificação que considera simultaneamente a capacidade do algoritmo de identificar corretamente instâncias de uma classe e de evitar classificações incorretas.

Esta métrica é especialmente valiosa quando se busca uma medida sintética que capture ambos os aspectos do desempenho de classificação (SOKOLOVA; LAPALME, 2009). A formulação matemática do *F1 Score* baseia-se na média harmônica entre *Precision* e *Recall*, conforme Equação 2.10.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (2.10)$$

O *F1 Score* resulta em um valor que é naturalmente mais influenciado pelo menor dos dois valores (*Precision* e *Recall*). Esta característica significa que um *F1 Score* alto requer que ambos os valores sejam razoavelmente altos, penalizando situações onde uma métrica é alta às custas da degradação significativa da outra (CHICCO; JURMAN, 2020).

A propriedade de média harmônica do *F1 Score* torna esta métrica potencialmente adequada para cenários onde se busca manter um equilíbrio entre *Precision* e *Recall*, evitando soluções extremas que otimizem apenas uma das métricas. Esta característica pode ser relevante em problemas de classificação balanceada, onde não há preferência clara por minimizar especificamente falsos positivos ou falsos negativos.

No contexto da metodologia NL-TSDR, o *F1 Score* fornece uma medida sintética da melhoria geral no desempenho de classificação resultante das transformações aplicadas. Melhorias consistentes no *F1 Score* ao longo de diferentes classes e algoritmos de classificação constituem evidência da eficácia do método em abordar problemas de mascaramento de classes sem criar novos desequilíbrios ou vieses.

A análise do *F1 Score* por classe individual pode revelar como diferentes categorias são afetadas pelas transformações propostas. Classes que originalmente apresentavam baixo *F1 Score* devido ao mascaramento podem mostrar melhorias substanciais após a aplicação do método NL-TSDR, enquanto classes já bem separadas devem manter ou melhorar ligeiramente seus valores, indicando que as transformações não degradam a discriminabilidade existente.

A utilização do *F1 Score* como métrica primária de comparação entre métodos é particularmente apropriada em cenários de pesquisa onde o objetivo é demonstrar melhorias gerais no desempenho de classificação. Sua natureza sintética permite comparações diretas entre diferentes abordagens, facilitando a identificação de métodos superiores e a quantificação das melhorias alcançadas.

Uma extensão do *F1 Score* em problemas multiclasse envolve o cálculo de médias ponderadas ou não ponderadas dos valores individuais por classe (OPITZ; BURST, 2021). Esta variação é representada pela Equação 2.11 e pela Equação 2.12, onde \mathcal{C} representa o conjunto de classes e $|\mathcal{C}|$ sua cardinalidade.

$$\text{F1 Score}_{\text{macro}} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \text{F1 Score}_c \quad (2.11)$$

$$\text{F1 Score}_{\text{micro}} = \frac{2 \cdot \sum_c VP_c}{2 \cdot \sum_c VP_c + \sum_c FP_c + \sum_c FN_c} \quad (2.12)$$

A interpretação adequada do *F1 Score* deve considerar suas limitações, particularmente em cenários com classes muito desbalanceadas onde a métrica pode não refletir adequadamente o desempenho em classes minoritárias. Nestes casos, a análise complementar de *Precision* e *Recall* individuais, bem como consideração de outras métricas como AUC-ROC (*Area Under the Receiver Operating Characteristic Curve*) (FAWCETT, 2006), pode fornecer uma avaliação mais completa do desempenho do classificador.

2.6 Trabalhos Relacionados

Diferentes abordagens na literatura abordam o problema de redução de dimensionalidade supervisionada com estratégias metodológicas distintas do método combinado proposto. Uma categoria são os métodos baseados em *Neighborhood Components Analysis* (NCA) (GOLDBERGER et al., 2004), que diferem ao otimizar diretamente a acurácia de classificação de um algoritmo k -NN estocástico. Enquanto a Etapa 1 do método proposto maximiza probabilidades de separação de classes com base na *Minimax Probability Machine*, o NCA utiliza uma função *softmax* sobre distâncias euclidianas para definir probabilidades de vizinhança, buscando uma transformação linear que maximize a acurácia de validação cruzada. Esta abordagem contrasta com a estratégia de duas etapas proposta, pois integra diretamente a informação de classes na otimização da métrica de distância, sem utilizar uma etapa subsequente de otimização de distâncias.

Outra classe são os métodos baseados em *Maximum Variance Unfolding* (MVU) e suas extensões supervisionadas (WEINBERGER; SHA; SAUL, 2004). O MVU tradicional utiliza programação semidefinida para desdobrar variedades preservando distâncias locais entre vizinhos próximos enquanto maximiza a variância global dos dados transformados.

As variantes supervisionadas, como o *Colored MVU* (SONG et al., 2007), incorporam informação de classes através do critério de independência de Hilbert-Schmidt (HSIC), maximizando a dependência entre os indivíduos e seus rótulos. Esta abordagem difere do método proposto pois utiliza programação semidefinida e não emprega explicitamente o conceito de probabilidade de separação como medida de separabilidade.

Os métodos baseados em *Local Fisher Discriminant Analysis* (LFDA) (SUGIYAMA, 2007) representam uma terceira categoria que combina a análise discriminante de Fisher com preservação de estrutura local dos dados. O LFDA considera as matrizes de dispersão intra e interclasses, sendo efetivo para dados onde amostras de uma mesma classe formam múltiplos agrupamentos, de forma semelhante à Etapa 2 do método proposto. Todavia, o LFDA mantém a formulação clássica baseada em matrizes de dispersão, incorporando pesos baseados em afinidade local. Esta diferença distingue as abordagens, pois o LFDA preserva a estrutura matemática do Fisher clássico enquanto adapta sua aplicação local.

Uma outra categoria são os métodos semissupervisionados que exploram tanto dados rotulados quanto não rotulados. O *Semi-supervised Local Fisher Discriminant Analysis* (SELF) (SUGIYAMA et al., 2010) exemplifica esta abordagem, combinando LFDA para separar amostras rotuladas com PCA para preservar a estrutura global de amostras não rotuladas. Esta estratégia contrasta com o método proposto que opera puramente no regime supervisionado, utilizando exclusivamente informação de classes rotuladas. A incorporação de dados não rotulados no SELF permite uma melhor estimativa da estrutura da variedade subjacente, especialmente quando o número de amostras rotuladas é limitado, representando uma direção metodológica ortogonal à abordagem de duas etapas proposta.

A explosão de dados digitais nas últimas décadas transformou os desafios de classificação e análise preditiva. Segundo estimativas recentes, a humanidade gera aproximadamente 402,74 milhões de terabytes de dados diariamente (Statista, 2024), com o volume global de dados crescendo de apenas 2 zettabytes em 2010 para aproximadamente 147 zettabytes em 2024 (Rivery, 2024). Este crescimento exponencial não se limita apenas ao volume, mas também à complexidade e dimensionalidade dos conjuntos de dados, criando novos paradigmas para métodos de aprendizado de máquina.

A última década testemunhou uma revolução na forma como se lida com dados de alta dimensionalidade, especialmente com o surgimento de métodos híbridos que integram redução de dimensionalidade com objetivos específicos de classificação. O Alinhamento Local de Recursos Profundos (*Local Deep-Feature Alignment*) exemplifica essa tendência emergente, construindo vizinhanças locais para cada amostra e extraindo características por meio de codificadores neurais profundos. Este método demonstra melhorias consistentes de 8-12% em tarefas de classificação de imagens quando comparado a abordagens tradicionais, evidenciando a superioridade de técnicas que consideram simultaneamente a estrutura local e global dos dados (ZHANG; YU; TAO, 2018).

Paralelamente, métodos baseados em grafos, como *t*-SNE (*t-Distributed Stochastic Neighbor Embedding*) e UMAP (*Uniform Manifold Approximation and Projection*), ganharam popularidade em análise exploratória de dados, oferecendo visualizações interpretáveis de estruturas complexas em espaços de alta dimensão. Variantes supervisionadas desses métodos, como o *Supervised UMAP* e o *Parametric t-SNE*, mostraram resultados promissores em preservar separabilidade de classes durante a redução dimensional, alcançando taxas de preservação de estrutura local superiores a 85% mesmo em reduções drásticas de dimensionalidade (SAINBURG; MCINNES; GENTNER, 2021; KOBAK; BERENS, 2019).

Uma tendência particularmente relevante é o desenvolvimento de técnicas de redução de dimensionalidade específicas para domínios. Em bioinformática, métodos como PHATE (*Potential of Heat-diffusion for Affinity-based Transition Embedding*) foram especificamente desenvolvidos para lidar com dados de *single-cell RNA sequencing*, demonstrando capacidade superior de preservar trajetórias biológicas complexas em comparação com métodos generalistas (MOON et al., 2019). Em análise de imagens médicas, técnicas de redução de dimensionalidade baseadas em VAE (*Variational Autoencoder*) têm mostrado eficácia notável na identificação de biomarcadores sutis, com alguns estudos reportando aumentos de até 23% na precisão diagnóstica (WAY; GREENE, 2018).

O campo de NLP também experimentou avanços significativos com o desenvolvimento de *embeddings* contextuais. Modelos como BERT (*Bidirectional Encoder Representations from Transformers*), GPT (*Generative Pre-trained Transformer*) e suas variações revolucionaram a representação de texto, criando representações de alta dimensionalidade que capturam nuances semânticas e sintáticas anteriormente inacessíveis. Trabalhos recentes demonstram que técnicas de redução de dimensionalidade podem manter até 95% da informação semântica original enquanto reduzem significativamente os custos computacionais (ROGERS; KOVALEVA; RUMSHISKY, 2021; QIU et al., 2020).

As redes sociais evoluíram para ecossistemas com mais de 4,8 bilhões de usuários ativos globalmente gerando conteúdo multimodal em escala (KEMP, 2024b). Este volume de dados, que inclui texto, imagens, vídeos, metadados temporais, grafos sociais e padrões comportamentais, apresenta oportunidades únicas para análise preditiva, mas também desafios computacionais consideráveis que requerem abordagens metodológicas inovadoras.

A proliferação de desinformação tornou-se uma preocupação global, com estudos indicando que notícias falsas se espalham seis vezes mais rapidamente que informações verdadeiras e alcançam significativamente mais pessoas (VOSOUGHI; ROY; ARAL, 2018). Trabalhos recentes em classificação automática de *fake news* demonstram a superioridade de abordagens que combinam análise textual, visual e metadados. Pesquisadores desenvolveram uma metodologia que incorpora características temporais, padrões de engajamento e análise de redes sociais, alcançando acurácia superior a 92% na detecção de desinformação, comparado aos 78% de métodos baseados em análise textual (LIU; WU, 2018).

Particularmente interessante é o trabalho que propõe uma abordagem baseada em grafos neurais para detecção de *fake news*, considerando não apenas o conteúdo das publicações, mas também os padrões de propagação através das redes sociais (ZHOU; ZAFARANI, 2019). Seus resultados indicam que a incorporação de informações de rede pode melhorar a precisão de detecção em até 15%, especialmente para notícias que se espalham rapidamente antes que verificações factuais possam ser realizadas.

A capacidade de prever o sucesso de conteúdo em redes sociais possui implicações significativas para marketing digital, jornalismo e compreensão de dinâmicas sociais. Trabalhos pioneiros nesta área demonstraram a viabilidade de estimar popularidade de publicações utilizando características extraídas de redes de amizade e contexto de conteúdo (SOBRINHO, 2019). Outras pesquisas expandiram esses mesmos conceitos, com Khosla et al. (KHOSLA; SARMA; HAMID, 2014) demonstrando que características visuais de imagens podem ser preditivas de sua popularidade, alcançando correlações de até 0,81 entre características extraídas por redes neurais convolucionais e métricas de engajamento.

Estudos mais recentes têm explorado a predição de popularidade em plataformas específicas com características únicas. Para o YouTube, por exemplo, Vallet et al. (VALLET et al., 2015) desenvolveram um modelo que considera características de conteúdo, *timing* de publicação, uso de *hashtags* e correlação com eventos externos. De forma similar, para LinkedIn, pesquisadores identificaram que características profissionais do usuário e relevância do conteúdo para *trends* da indústria são preditores mais significativos que métricas tradicionais de engajamento (USERA; DURHAM, 2025).

Uma linha de pesquisa emergente foca na análise de sentimento multimodal em redes sociais, combinando texto, imagens e até características de áudio em vídeos para compreender melhor o impacto emocional do conteúdo. Zadeh et al. (ZADEH et al., 2017) propuseram o uso de *Tensor Fusion Networks* para integrar essas modalidades diversas, demonstrando melhorias substanciais na acurácia de classificação de sentimento comparado a abordagens unimodais.

O campo de análise de influenciadores também tem se beneficiado significativamente de avanços em redução de dimensionalidade e classificação. Trabalhos recentes identificaram que a "autenticidade percebida" de influenciadores pode ser quantificada através de análise de padrões linguísticos, consistência visual e padrões de engajamento (LIU; ZHENG, 2024).

Desenvolvimentos contemporâneos também exploram a aplicação de técnicas de *federated learning* em análise de redes sociais, permitindo o treinamento de modelos sobre dados distribuídos sem comprometer a privacidade dos usuários. Pesquisadores (MEHTA; KUMAR, 2024) demonstraram que abordagens federadas podem manter desempenho comparável a métodos centralizados enquanto preservam privacidade, uma consideração potencialmente importante dado o cenário regulatório atual.

Uma área particularmente promissora é a aplicação de *Graph Neural Networks* (GNNs) para análise de redes sociais. Hamilton et al. (HAMILTON; YING; LESKOVEC, 2017) desenvolveram GraphSAGE, que permite a geração de *embeddings* para nós não vistos durante o treinamento, facilitando a análise de redes sociais dinâmicas onde novos usuários e conexões surgem constantemente. Trabalhos subsequentes expandiram essa abordagem para incluir características temporais, permitindo a modelagem da evolução de comunidades e influência ao longo do tempo (KAZEMI et al., 2020).

Desenvolvimentos recentes apontam para três direções principais que estão moldando o futuro da área. Primeiro, a integração crescente de técnicas de aprendizado profundo com métodos tradicionais de redução de dimensionalidade está gerando abordagens híbridas mais robustas. Segundo, o desenvolvimento de métricas de qualidade de características específicas para domínios está permitindo avaliações mais precisas da eficácia de diferentes técnicas. Terceiro, a criação de pipelines adaptativos que ajustam estratégias de redução dimensional baseadas em características dos dados e objetivos de classificação está emergindo como uma área de pesquisa promissora.

Especificamente no contexto de aprendizado profundo, VAEs e *Generative Adversarial Networks* (GANs) têm mostrado capacidade notável não apenas para redução de dimensionalidade, mas também para geração de dados sintéticos que preservam características estatísticas dos conjuntos originais. Trabalhos recentes demonstram que dados sintéticos gerados por essas técnicas podem enriquecer conjuntos de treinamento, resultando em melhorias de desempenho em tarefas de classificação (WANG et al., 2019).

A convergência dessas tendências sugere que futuras pesquisas devem focar em abordagens holísticas que considerem simultaneamente a estrutura dos dados, objetivos de classificação e restrições computacionais, movendo-se além da aplicação isolada de técnicas individuais em direção a sistemas integrados de análise preditiva. Particularmente relevante é a necessidade de desenvolver ferramentas que possam automaticamente selecionar e configurar técnicas de redução de dimensionalidade baseadas nas características específicas dos dados de entrada, uma área que pouco explorada na literatura atual.

3 Método Proposto

A predição de engajamento em mídias sociais apresenta desafios únicos em sistemas de informação inteligentes, particularmente o problema de mascaramento de classes que ocorre quando espaços de características de alta dimensionalidade obscurecem as fronteiras discriminativas entre categorias de engajamento. As técnicas tradicionais de redução de dimensionalidade linear frequentemente falham em capturar as relações complexas e não lineares inerentes aos dados de mídias sociais, onde as interações dos usuários exibem distribuições multimodais e fronteiras de decisão irregulares.

Para abordar esses desafios, propomos um método de Redução de Dimensionalidade Não Linear em Duas Etapas (NL-TSDR ou *Nonlinear Two-Stage Dimensionality Reduction*) que combina sistematicamente a otimização de separação probabilística com transformações não lineares adaptativas. O método opera sob a premissa de que espaços de características ótimos para classificação requerem tanto a maximização da separação probabilística entre classes de engajamento quanto adaptações não lineares que capturem os padrões complexos de interação característicos de dados de alta dimensionalidade.

O método proposto fundamenta-se na observação de que os dados de mídias sociais frequentemente apresentam estruturas não lineares complexas que não podem ser adequadamente modeladas por técnicas lineares convencionais. A abordagem em duas etapas permite uma decomposição do problema de redução de dimensionalidade em duas fases complementares: primeiro, a otimização da separabilidade entre classes em um espaço probabilístico; segundo, a aplicação de transformações não lineares para capturar padrões mais complexos nos dados.

Considere um conjunto de dados com n características (*features*) e c classes (*labels*). Na Etapa 1, o conjunto de dados é reduzido para d dimensões, que é um parâmetro ajustável otimizado para máxima precisão do método. Esse subconjunto reduzido de d dimensões então passa pela Etapa 2, empregando otimização multiobjetivo para transformá-lo em um subconjunto final de l dimensões, equivalente ao número de classes no conjunto de dados original ($l = c$). Os subconjuntos reduzidos finais são então classificados usando tanto um discriminador proposto embarcado na segunda etapa quanto métodos de classificação bem estabelecidos.

Nas seções subsequentes, serão detalhadas as duas etapas do método ([seção 3.1](#) e [seção 3.2](#)), assim como as diferentes transformações não lineares aplicadas ao cálculo dos centros de classe ([subseção 3.2.1](#)). Ao final deste capítulo, são também apresentadas uma visão geral dos algoritmo e de sua respectiva implementação ([seção 3.3](#)).

3.1 Etapa 1 - Maximização da Probabilidade de Separação de Classes

A redução de dimensionalidade, ou *dimensionality reduction* (DR), pode ser realizada por meio da maximização da probabilidade de separação das instâncias par a par, conhecida como *Dimension Reduction via Maximizing Pairwise Separation Probability* (DR-MSP) (YANG et al., 2019). Esta técnica aborda o problema de mascaramento de classes ao focar em separar classes que são naturalmente próximas no espaço original. Por sua vez, a probabilidade de separação é baseada principalmente na *Minimax Probability Machine* (MPM), que maximiza a probabilidade de classificação correta dos pontos de dados futuros (LANCKRIET et al., 2003).

Sejam \mathbf{x} e \mathbf{y} vetores aleatórios em um problema de classificação binária, onde cada um corresponde a dados de indivíduos de duas classes diferentes. Suponha que as médias e matrizes de covariância dessas duas classes sejam, respectivamente, $(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ e $(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$, com $\mathbf{x}, \boldsymbol{\mu}_x, \mathbf{y}, \boldsymbol{\mu}_y \in \mathbb{R}^D$ ($\boldsymbol{\mu}_x \neq \boldsymbol{\mu}_y$) e $\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y \in \mathbb{R}^{D \times D}$. O objetivo então é maximizar a probabilidade de separação α de que duas classes estejam em lados opostos de um hiperplano \mathbf{w} , onde $\mathbf{w} \in \mathbb{R}^D$, $\mathbf{w} \neq \mathbf{0}$, e $b \in \mathbb{R}$.

Este hiperplano ótimo separa as duas classes com uma probabilidade máxima em relação a todas as distribuições possíveis com as médias e matrizes de covariância dadas. Este hiperplano ótimo \mathbf{w} e a correspondente probabilidade de separação α de duas classes no subespaço \mathbf{w} podem ser derivados a partir da medida de separação κ , conforme descrito na Equação 3.1 e Equação 3.2 (YANG et al., 2019).

$$\kappa_{ij}(\mathbf{w}) = \frac{|\mathbf{w}^T(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)|}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma}_x \mathbf{w}} + \sqrt{\mathbf{w}^T \boldsymbol{\Sigma}_y \mathbf{w}}} \quad (3.1)$$

$$\alpha(\mathbf{w}) = \frac{\kappa(\mathbf{w})^2}{1 + \kappa(\mathbf{w})^2} \quad (3.2)$$

Dessa forma, κ pode representar a razão sinal-ruído entre as classes, uma vez que o numerador representa a dispersão entre as médias projetadas das classes, e o denominador representa a dispersão combinada (desvio padrão) das classes. Com as médias $(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$ e as matrizes de covariância $(\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y)$ das duas classes, pode-se calcular a probabilidade de separação correspondente α para quantificar a separabilidade da classe entre elas em um determinado subespaço unidimensional \mathbf{w} .

Agora, seja um conjunto de dados com C classes e D dimensões, cuja distribuição condicional da i -ésima classe é dada por $p(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, onde mais uma vez $\boldsymbol{\mu}_i$ e $\boldsymbol{\Sigma}_i$ sejam respectivamente a média e a covariância da i -ésima classe. Dessa maneira, a probabilidade

de separação da i -ésima e da j -ésima classes no subespaço \mathbf{w} pode ser derivada conforme Equação 3.3 e Equação 3.4 (YANG et al., 2019).

$$\kappa_{ij}(\mathbf{w}) = \frac{|\mathbf{w}^T(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)|}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma}_i \mathbf{w}} + \sqrt{\mathbf{w}^T \boldsymbol{\Sigma}_j \mathbf{w}}} \quad (3.3)$$

$$\alpha_{ij}(\mathbf{w}) = \frac{\kappa_{ij}(\mathbf{w})^2}{1 + \kappa_{ij}(\mathbf{w})^2} \quad (3.4)$$

Dessa forma, o objetivo é encontrar o subespaço unidimensional $\mathbf{w} \in \mathbb{R}^D$ onde a soma das probabilidades de separação de todos os pares de classes é maximizada, o que pode ser apresentado conforme a Equação 3.5 (YANG et al., 2019). Sendo um problema não convexo, é possível então utilizar um método otimização baseado em gradiente, conforme a Equação 3.6.

$$\max_{\|\mathbf{w}\|=1} J(\mathbf{w}) = \sum_{1 \leq i < j \leq C} \alpha_{ij}(\mathbf{w}) \quad (3.5)$$

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \gamma^{(t)} \frac{\partial J}{\partial \mathbf{w}} \quad (3.6)$$

Como \mathbf{w} é um subespaço unidimensional, pode-se utilizar o processo recursivo de Gram-Schmidt (DUKES, 2014) para encontrar múltiplos subespaços unidimensionais \mathbf{w}_r de forma a encontrar um espaço multidimensional de d dimensões $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_r, \dots, \mathbf{w}_d)$ onde d é a dimensão-alvo. Após o cálculo da primeira direção (\mathbf{w}_1) que maximiza todos os pares de classes, calcula-se uma segunda direção (\mathbf{w}_2) ortogonal à(s) anterior(es) que maximiza a separação deste espaço, e assim sucessivamente. Uma vez que a matriz de transformação final $\mathbf{W} \in \mathbb{R}^{D \times d}$ é encontrada, o conjunto de dados original pode ser transformado conforme $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ (onde o i -ésimo elemento de $y_i \in \mathbb{R}^d$), criando um subconjunto reduzido das D dimensões originais para a dimensão-alvo d .

Considere o seguinte exemplo, onde deve-se transformar um conjunto de dados original de 5 classes e mil dimensões em um novo subconjunto reduzido de 10 dimensões. A Etapa 1 então:

- Calcula as médias e as covariâncias para cada classe;
- Encontra o primeiro subespaço unidimensional \mathbf{w}_1 que maximiza a probabilidade de separação de todos os 10 pares de classes;
- Encontra o segundo subespaço unidimensional \mathbf{w}_2 ortogonal à \mathbf{w}_1 que maximiza a separação deste espaço;
- Continua até que os 10 espaços ortogonais sejam encontrados;

- Constrói a matriz $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{10}) \in \mathbb{R}^{1000 \times 10}$;
- Calcula o subconjunto de dados resultante $\mathbf{y} = \mathbf{W}^T \mathbf{x}$ encontrando assim uma nova representação com 10 dimensões.

O método de redução de dimensionalidade descrito está fortemente relacionado aos seus homólogos, especialmente a Análise Discriminante Linear. Descrevendo de uma maneira genérica, a LDA visa maximizar a distância entre as médias das diferentes classes e minimizar a variação dentro de cada uma delas (IZENMAN, 2009). Comparado à LDA, ele usa a probabilidade de separação como medida de separabilidade de classe, e tal mecanismo o obriga a colocar mais esforço em pares de classes que são menos discriminantes, o que pode evitar a sobreposição de pares de classes próximos no subespaço obtido, onde o problema de mascaramento de classes é atenuado. Uma vez obtido o subconjunto ótimo, os dados são encaminhados para a segunda etapa.

O processo de otimização na primeira etapa pode ser interpretado geometricamente como a busca por direções no espaço de características que maximizam a separação entre todas as classes simultaneamente. Diferentemente de técnicas como PCA, que maximizam a variância total, ou LDA, que maximiza a razão entre variância interclasse e intraclasse, a abordagem proposta baseia-se em princípios probabilísticos que oferecem garantias teóricas mais robustas sobre a qualidade da separação obtida.

A complexidade computacional da Etapa 1 pode ser estimada, no pior caso, como $O(d(lD^2C^2 + 2D^2))$ (YANG et al., 2019). O termo d representa a dimensionalidade alvo do subespaço que se deseja obter. A complexidade é proporcional a d porque o algoritmo encontra o subespaço final recursivamente, calculando d subespaços unidimensionais. Já l é o número de iterações necessárias para o método do gradiente convergir, otimizando cada subespaço unidimensional. O termo D é a dimensionalidade original dos dados de entrada, e o D^2 surge porque o cálculo do gradiente envolve operações com matrizes e vetores de dimensão $D \times D$ ou $D \times 1$. Por fim, C representa o número de classes nos dados. O termo C^2 surge porque a função objetivo dada pela Equação 3.5 é a soma das probabilidades de separação de todos os pares de classes, o que leva a $\frac{C(C-1)}{2}$ termos.

3.2 Etapa 2 - Otimização de Espaço Não Linear

A segunda etapa introduz transformações não lineares em uma otimização multiobjetivo para aprimorar a separação de *clusters* enquanto preserva a estrutura local. Construindo sobre o subespaço probabilisticamente otimizado da Etapa 1, esta etapa aborda as relações não lineares características dos padrões de engajamento em mídias sociais. A motivação para a inclusão de transformações não lineares surge da observação de que os dados de mídias sociais frequentemente exibem padrões complexos que não podem

ser adequadamente capturados por projeções lineares. Estes padrões incluem relações de saturação em métricas de engajamento, comportamentos de limiar em interações de usuários, e distribuições multimodais resultantes de diferentes segmentos de audiência.

A Etapa 2 visa encontrar uma transformação do conjunto de dados de entrada, neste caso o subespaço ótimo entregue pelo método de redução de dimensionalidade, para um novo conjunto de dados onde as instâncias da mesma classe estão o mais próximas possível umas das outras, enquanto os centros de gravidade dessas classes são o mais longe possível um do outro. Esse método é basicamente formulado como um problema de otimização multiobjetivo, que é convertido diretamente em um problema de otimização (minimização) de um único objetivo.

Considere que $X_{m_k \times n}^k$ inclua todas as instâncias da classe k (um *subset* de S_k), onde cada linha corresponde a uma instância. Transforma-se cada linha dessa matriz pela função F de modo que obtém-se $Y_{m_k \times p}^k$. Define-se \vec{a}^k , um vetor de p dimensões, como o centro de gravidade de todas as m_k instâncias em $Y_{m_k \times p}^k$ conforme a [Equação 3.7](#).

$$\vec{a}^k = \frac{1}{m_k} \sum_{i=1}^{m_k} \vec{y}^i \quad (3.7)$$

O vetor \vec{y}^i é a i -ésima linha de $Y_{m_k \times p}^k$. Também define-se o escalar v^k como a norma dos autovalores da matriz de covariância de $Y_{m_k \times p}^k$ conforme a [Equação 3.8](#).

$$v^k = \left\| \text{Eig}(\text{Cov}(Y_{m_k \times p}^k)) \right\| \quad (3.8)$$

Determina-se $\text{Cov}(\cdot)$ como o operador de covariância e $\text{Eig}(\cdot)$ o cálculo dos autovalores da matriz de entrada. O valor de v^k indica o quanto as instâncias da classe k estão distribuídas ao redor de seu centro por meio das suas mais importantes direções (autovetores). O objetivo do método é adaptar a transformada F de forma que v^k seja minimizado para todo k enquanto as distâncias entre os centros de gravidade seja maximizada. Isso pode ser formulado como um problema de otimização multiobjetivo conforme a [Equação 3.9](#) (BONYADI; TIENG; REUTENS, 2019).

$$\Omega = \begin{cases} \max \|\vec{a}^i - \vec{a}^j\| & \text{para todo } j > i \\ \min v^i & \text{para todo } i \end{cases} \quad i, j \in \{1, \dots, c\} \quad (3.9)$$

O problema contém c objetivos de minimização e $\frac{c(c-1)}{2}$ objetivos de maximização. Com os seguintes critérios, é possível convertê-lo em uma otimização de um único objetivo.

1. Suponha que $A_i(x)$ é uma função e $A_i(x) > 0$ para todo i e x . Uma solução que minimiza $A_i(x)$ está na verdadeira frente de Pareto do problema de otimização multiobjetivo: "minimize $A_i(x)$ para todo i ".

2. Suponha que $A_i(x)$ é uma função e $A_i(x) > 0$ para todo i e x . Uma solução que maximiza $A_i(x)$ está na verdadeira frente de Pareto do problema de otimização: "maximize $A_i(x)$ para todo i ".

As provas para ambas as observações são elementares e podem ser feitas por contradição. Usando as observações 1 e 2, o problema de otimização multiobjetivo definido em [Equação 3.9](#) pode ser transformado em um problema de otimização com um único objetivo, conforme definido pela [Equação 3.10](#) (BONYADI; TIENG; REUTENS, 2019).

$$\Omega = \frac{\gamma + \sum_{k=1}^c v^k}{\left(\prod_i^c \prod_{j=i+1}^c \|\vec{a}^i - \vec{a}^j\|\right)^{\frac{1}{c(c-1)}}} \quad (3.10)$$

A variável γ é uma constante positiva (nos experimentos $\gamma = 1$) utilizada para garantir que, entre todas as possíveis soluções onde $\sum_{k=1}^c v^k = 0$, aquela que maximiza as distâncias entre os centros de gravidade ($\prod_i^c \prod_{j=i+1}^c \|\vec{a}^i - \vec{a}^j\|$) seja priorizada.

O regulador $\frac{1}{c(c-1)}$ é utilizado para balancear a taxa de crescimento entre numerador e denominador. Isso garante que a importância da distinção entre os centros de gravidade seja equilibrada com a importância de minimizar a propagação de instâncias dentro de cada classe.

O produto (em vez de uma simples soma ponderada) obriga o otimizador a encontrar soluções nas quais os centros de classe estão espalhados. Este é um ponto relevante, pois, caso contrário, o algoritmo de otimização pode encontrar uma solução indesejável que mapeie alguns dos centros próximos uns dos outros enquanto afasta os demais centros dos demais.

Após resolver Ω , uma função discriminante, determinada pela [Equação 3.11](#), é utilizada para determinar a qual classe um indivíduo pertence, avaliando sua distância ao centro de gravidade de todas as classes disponíveis. Dessa maneira, quanto menor o valor, maior a probabilidade de a instância pertencer a uma classe específica.

$$f_k(\vec{y}) = \frac{D_k}{\sum_{j=1}^c D_j} \quad (3.11)$$

A função discriminante, definida por $f_k(\vec{y})$, onde $D_k = \|\vec{y} - \vec{a}^k\|$, pode ser interpretada como a probabilidade de $\vec{y} \in S_k$, uma vez que $f_k(\vec{y}) \in [0, 1]$. Em outras palavras, quanto menor o valor de $f_k(\vec{y})$, maior a probabilidade do indivíduo y de pertencer à classe k . Desse ponto em diante, os resultados são convertidos de valores generativos para discriminativos, e as acurácias podem ser calculadas. O uso desta função discriminante não é obrigatório, o que garante a flexibilidade necessária para que este método funcione em conjunto com outros.

A relação entre maximizar a probabilidade de separação na Etapa 1 e o desempenho de classificação aprimorado manifesta-se através do poder discriminativo aprimorado no espaço reduzido. Ao otimizar separações par a par, a Etapa 1 cria uma fundação onde as transformações não lineares subsequentes da Etapa 2 podem estabelecer fronteiras de decisão mais efetivamente, levando a melhorias de precisão empiricamente observadas.

A complexidade computacional da Etapa 2 é estimada, sobretudo, pelo custo da avaliação da função objetivo Ω combinado com a eficiência do método de otimização selecionado (BONYADI; TIENG; REUTENS, 2019). Ao empregar o método Nelder-Mead, que é a base do `fminsearch` no MATLAB, a complexidade temporal *Big O* do algoritmo é estimada em $O(mn_v^3)$. Nesta notação, o fator m representa o número de instâncias no conjunto de treinamento. Por sua vez, o termo n_v é o número total de variáveis que estão sendo otimizadas, definido como o produto $n \times p$, onde n é o número de *features* originais e p é a dimensão do espaço transformado. O fator dominante é o n_v^3 , que surge da multiplicação do custo de uma única avaliação da função ($O(mn_v)$) pelo número de iterações necessárias para o Nelder-Mead convergir ($O(n_v^2)$) (LAGARIAS et al., 1998).

3.2.1 Funções de Transformação Não Lineares

Para capturar melhor diversas relações não lineares nos dados, investiga-se quatro funções de transformação aplicadas aos cálculos dos centros de *clusters*. Cada transformação é selecionada para abordar expectativas teóricas específicas sobre características dos dados e seu impacto na formação e separabilidade de *clusters*.

A seleção das funções de transformação baseia-se em propriedades matemáticas bem estabelecidas e sua adequação para diferentes tipos de distribuições de dados encontradas em aplicações de mídias sociais. Cada transformação é projetada para capturar aspectos específicos da estrutura dos dados que podem não ser adequadamente modelados por transformações lineares (QIU; SAPIRO, 2015).

A Tabela [Tabela 3.1](#) apresenta uma síntese comparativa das quatro transformações não lineares propostas, destacando suas características matemáticas fundamentais, parâmetros de controle e domínios de aplicação adequados. Esta comparação permite identificar a transformação mais apropriada para diferentes cenários de dados, considerando as propriedades estatísticas das distribuições subjacentes e os objetivos específicos da análise de *clusters*. A diversidade de comportamentos matemáticos oferecida por essas transformações proporciona flexibilidade na modelagem de diferentes tipos de não linearidades presentes em dados de mídias sociais, desde distribuições com caudas pesadas até padrões com comportamento de saturação.

Tabela 3.1 – Comparação entre transformações não lineares

Transformação	Intervalo	Parâmetros	Domínios	Vantagens
Tangente Hiperbólica	$[-1, 1]$	β (taxa de compressão)	Distribuições simétricas com presença de <i>outliers</i>	Comprime valores extremos preservando ordenação relativa dos dados
Gaussiana	$[0, 1]$	μ (média), σ (desvio padrão)	Distribuições aproximadamente normais com escalas variáveis	Normalização robusta adequada para características com diferentes unidades
Sigmoide	$[0, 1]$	α (nitidez da transição)	Dados com efeitos de limiar ou comportamento de saturação	Captura não linearidades inerentes com interpretação probabilística natural
Laplaciana	$[0, 1]$	μ (centro), λ (escala)	Distribuições multimodais com <i>outliers</i> significativos	Robustez estatística com aparaamento suave de valores extremos

Fonte: Elaborado pelo autor com base em (QIU; SAPIRO, 2015).

3.2.1.1 Transformação Tangente Hiperbólica

A primeira transformação é a transformação tangente hiperbólica, projetada para dados exibindo distribuições simétricas ao redor de centros de *clusters* com potenciais *outliers*, descrita na Equação 3.12 (KINGMA; BA, 2017).

$$\vec{a}^k = \frac{1}{m_k} \sum_{i=1}^{m_k} \tanh(\beta \vec{y}^i) \quad (3.12)$$

Esta transformação comprime valores extremos enquanto preserva a ordenação relativa, levando a centros de *clusters* mais robustos quando os dados contêm *outliers*. O parâmetro β controla a taxa de compressão: valores maiores criam transições mais abruptas, efetivamente filtrando ruído, enquanto valores menores mantêm relações mais lineares (GLOROT; BENGIO, 2010).

A função tangente hiperbólica possui propriedades matemáticas desejáveis para processamento de dados. Sua natureza sigmoide garante que valores de entrada muito grandes ou muito pequenos sejam mapeados para um intervalo limitado $[-1, 1]$, efetivamente controlando a influência de *outliers* extremos. Além disso, a função mantém a monotonicidade, preservando a ordenação relativa dos dados originais.

Espera-se que esta transformação funcione bem quando as características exibam distribuições *power law* com caudas pesadas. Em tais cenários, a compressão de valores

extremos pode revelar padrões sutis que seriam obscurecidos pela dominância de *outliers*. O parâmetro β oferece controle fino sobre o grau de não linearidade introduzido, permitindo adaptação às características específicas do conjunto de dados.

Do ponto de vista da teoria da informação, a transformação tangente hiperbólica pode ser vista como um mecanismo de equalização que redistribui a “atenção” do algoritmo de forma mais uniforme entre diferentes regiões do espaço de características, potencialmente melhorando a capacidade de discriminação em regiões onde os dados são menos densos.

3.2.1.2 Transformação Gaussiana

Por outro lado, a transformação gaussiana tem como alvo dados com distribuições aproximadamente normais dentro de *clusters*, mas escalas variando entre diferentes categorias de engajamento. Esta transformação é denotada pela Equação 3.13, onde Φ representa a função de distribuição cumulativa normal padrão (ZHU et al., 2021).

$$\vec{d}^k = \frac{1}{m_k} \sum_{i=1}^{m_k} \Phi \left(\frac{\vec{y}^i - \mu}{\sigma} \right) \quad (3.13)$$

O uso da transformação gaussiana normaliza centros de *clusters* para uma escala comum enquanto preserva o posicionamento relativo de instâncias dentro de cada cluster. Ao mapear para probabilidades cumulativas, aborda o desafio de comparar métricas com intervalos vastamente diferentes. Antecipa-se desempenho superior quando padrões seguem distribuições aproximadamente gaussianas dentro de cada classe. Esta transformação é potencialmente útil quando há de se lidar com características que têm unidades ou escalas muito diferentes.

Por exemplo, métricas de engajamento podem variar de contagens de curtidas (potencialmente milhares) a taxas de engajamento (proporções entre 0 e 1). A transformação gaussiana mapeia todas essas características para o intervalo $[0, 1]$ usando suas respectivas distribuições empíricas.

A escolha dos parâmetros μ e σ é crítica para o desempenho da transformação. Estes podem ser estimados a partir dos dados usando a média e desvio padrão amostrais, ou podem ser determinados através de técnicas de estimação que são menos sensíveis a *outliers*. Do ponto de vista estatístico, esta transformação implementa uma forma de padronização que é mais robusta que a normalização *z-score* tradicional, especialmente quando as distribuições subjacentes não são perfeitamente gaussianas. A função CDF (*Cumulative Distribution Function*) gaussiana fornece um mapeamento suave e monótono que preserva a informação ordinal enquanto normaliza a escala.

3.2.1.3 Transformação Sigmoidal

A transformação sigmoide, descrita pela [Equação 3.14](#), é especificamente escolhida para dados onde exibem efeitos de limiar ou comportamento de saturação.

$$\vec{d}^k = \frac{1}{m_k} \sum_{i=1}^{m_k} \frac{1}{1 + e^{-\alpha \vec{y}^i}} \quad (3.14)$$

O parâmetro α determina a nitidez da transição: valores mais altos enfatizam distinções similares a binárias, enquanto valores mais baixos preservam transições graduais. Esta transformação captura cenários onde dados exibem retornos diminuídos ou efeitos de platô. Deve ter desempenho ótimo quando classes exibem padrões de crescimento logístico ([MURPHY, 2022](#)).

A função sigmoide é utilizada em redes neurais e aprendizado de máquina devido às suas propriedades matemáticas favoráveis. Sua derivada tem uma forma fechada simples, facilitando otimização baseada em gradiente. Além disso, a função mapeia todo o espaço real para o intervalo $(0, 1)$, fornecendo uma interpretação probabilística natural.

Em contextos de mídias sociais, muitas métricas de engajamento exibem comportamento sigmoide natural. Por exemplo, a probabilidade de um usuário interagir com conteúdo pode seguir uma curva sigmoide em relação a fatores como tempo desde a publicação ou relevância do conteúdo. A transformação sigmoide pode, portanto, capturar essas não linearidades inerentes nos dados.

O parâmetro α controla a “inclinação” da curva sigmoide. Valores pequenos de α resultam em uma transformação quase linear, enquanto valores grandes criam uma função degrau aproximada. Esta flexibilidade permite que a transformação se adapte a diferentes graus de não linearidade presentes nos dados, desde transições suaves até mudanças abruptas de comportamento.

3.2.1.4 Transformação Laplaciana

Finalmente, a transformação laplaciana, representada pela [Equação 3.15](#), aborda cenários com distribuições de engajamento multimodais e presença significativa de *outliers*, onde sgn é a função sinal e λ é um parâmetro de escala.

$$\vec{d}^k = \frac{1}{m_k} \sum_{i=1}^{m_k} \frac{1}{2} + \frac{1}{2} \text{sgn}(\vec{y}^i - \mu) \left(1 - e^{-\frac{|\vec{y}^i - \mu|}{\lambda}} \right) \quad (3.15)$$

O uso desta transformação fornece robustez contra anomalias enquanto mantém sensibilidade ao comportamento mediano dentro de cada classe. O decaimento exponencial controlado pelo parâmetro λ garante que valores extremamente altos ou baixos contribuam proporcionalmente em vez de predominantemente para os centros de *clusters*. Espera-se

desempenho superior quando dados contêm padrões mistos com transições abruptas (KOTZ; KOZUBOWSKI; PODGORSKI, 2012; XU; TIAN, 2015).

A distribuição laplaciana é conhecida por sua robustez estatística, sendo menos sensível a *outliers* que a distribuição gaussiana. Esta propriedade é transferida para a transformação, que efetivamente implementa uma forma de “aparamento suave” de valores extremos. Diferentemente de métodos de aparamento tradicionais que removem *outliers* completamente, a transformação os inclui mas com peso reduzido.

A forma da transformação laplaciana é assimétrica ao redor de μ , permitindo tratamento diferenciado de *outliers* positivos e negativos. Esta característica é particularmente útil em dados de mídias sociais, onde *outliers* em direções diferentes podem ter significados distintos (por exemplo, engajamento excepcionalmente alto ou baixo).

O parâmetro λ controla a "tolerância" da transformação a desvios em relação ao centro μ . Valores pequenos de λ resultam em maior sensibilidade a outliers, enquanto valores grandes proporcionam maior robustez. A seleção apropriada deste parâmetro permite equilibrar entre sensibilidade a padrões genuínos nos dados e robustez contra ruído ou anomalias.

3.3 Algoritmos e Implementação

A implementação do método NL-TSDR foi desenvolvida em MATLAB, aproveitando suas capacidades otimizadas para computação matricial e algoritmos de otimização não linear. Na Etapa 1, emprega-se o SQP para a otimização de separação probabilística, beneficiando-se de sua robustez em problemas de otimização não linear com restrições através de métodos como programação quadrática sequencial e pontos interiores, conforme detalhado na [seção 2.3](#). Esta escolha é particularmente adequada dado que a otimização probabilística frequentemente requer a imposição de restrições nos parâmetros para garantir propriedades matemáticas desejáveis.

Para a Etapa 2, utiliza-se o método Simplex de Nelder-Mead, que se mostra eficaz para a otimização de transformações não lineares devido à sua capacidade de lidar com funções objetivo que podem apresentar descontinuidades ou ruído inerente às transformações aplicadas, conforme também detalhado na [seção 2.3](#). A combinação destes dois algoritmos determinísticos oferece um equilíbrio entre eficiência computacional e qualidade das soluções, embora seja necessário considerar estratégias de múltiplas inicializações para mitigar o risco de convergência para ótimos locais.

O método completo NL-TSDR é formalizado no Algoritmo 1, que integra tanto as etapas de otimização de separação probabilística quanto de transformação de espaço não linear. A expectativa teórica é que diferentes conjuntos de dados favorecerão dife-

rentes transformações baseadas em suas características de distribuição de engajamento subjacentes.

A seleção de parâmetros para transformações não lineares é realizada através de busca em grade dentro de intervalos teoricamente motivados. Para cada transformação, são estabelecidos limites de parâmetros baseados nas características dos dados e emprega-se validação cruzada para identificar valores ótimos que maximizam a função objetivo.

O processo de seleção de parâmetros é crítico para o sucesso do método. Para a transformação tangente hiperbólica, o parâmetro β é tipicamente explorado no intervalo entre 0, 1 e 10, onde valores baixos preservam mais linearidade e valores altos introduzem maior não linearidade. Para a transformação gaussiana, μ e σ são frequentemente estimados a partir dos dados empíricos, mas podem ser ajustados para otimizar o desempenho.

A transformação sigmoide requer cuidado especial na seleção de α , pois valores muito altos podem causar saturação prematura, enquanto valores muito baixos podem não capturar adequadamente as não linearidades presentes. Tipicamente, explora-se α no intervalo entre 0, 1 e 5. Para a transformação laplaciana, λ é selecionado baseado na escala dos dados, frequentemente como uma fração do desvio padrão das características.

A relação de complexidade computacional entre as etapas favorece eficiência: a Etapa 1 reduz a dimensionalidade de n para d (onde $d \ll n$), permitindo que a Etapa 2 opere em espaços de características significativamente menores. Embora a análise formal de complexidade requeira avaliação específica do conjunto de dados, observação empírica confirma que saídas menores da Etapa 1 facilitam convergência na otimização da Etapa 2.

A integração entre etapas é direta: o subespaço otimizado da Etapa 1 serve diretamente como entrada para a Etapa 2, mantendo relações entre características enquanto permite adaptação não linear. Esta arquitetura modular facilita experimentação com diferentes combinações de técnicas de otimização e funções de transformação.

A flexibilidade do método permite várias estratégias de implementação. A Etapa 1 pode ser implementada usando diferentes algoritmos de otimização, desde métodos de gradiente simples até técnicas mais sofisticadas como otimização quasi-Newton ou algoritmos evolutivos. Similarmente, a Etapa 2 pode empregar diferentes abordagens para otimização multiobjetivo, cada uma com suas próprias vantagens em termos de qualidade da solução e eficiência computacional.

Um aspecto potencialmente importante é a validação da qualidade das transformações obtidas. Além das métricas de desempenho de classificação, é útil examinar propriedades como preservação de vizinhança local, manutenção de estruturas de *cluster*, e interpretabilidade das dimensões resultantes. Estas análises podem fornecer informações valiosas sobre quando e por que certas transformações são mais eficazes para tipos específicos de dados.

O método proposto também oferece oportunidades para extensões futuras. Por exemplo, transformações adaptativas que variam localmente baseadas nas características dos dados, incorporação de conhecimento de domínio através de restrições adicionais na otimização, ou desenvolvimento de métricas de qualidade mais sofisticadas que considerem tanto separabilidade quanto interpretabilidade.

Em termos de aplicabilidade prática, o método NL-TSDR pode ser particularmente adequado para cenários onde a interpretabilidade das dimensões reduzidas é importante, onde os dados exibem estruturas não lineares complexas, e onde métodos tradicionais de redução de dimensionalidade apresentam limitações. A natureza modular do método permite adaptação a diferentes domínios de aplicação com modificações mínimas.

Algoritmo 1 Redução de Dimensionalidade Não Linear em Duas Etapas (NL-TSDR)**Entrada:** Conjunto de dados $X \in \mathbb{R}^{n \times D}$, número de classes c , dimensão intermediária d **Saída:** Dados transformados $Z \in \mathbb{R}^{n \times c}$, função discriminante f 1: **Etapa 1: Otimização de Separação Probabilística**2: Calcular médias de classe μ_i e covariâncias Σ_i para $i = 1, \dots, c$ 3: Inicializar direção de projeção $w \in \mathbb{R}^D$ aleatoriamente4: Normalizar $w \leftarrow w/\|w\|$ 5: **repita**6: Calcular probabilidades de separação $\alpha_{ij}(w)$ para todos os pares (i, j) 7: Calcular função objetivo $J(w) = \sum_{i < j} \alpha_{ij}(w)$ 8: Calcular gradiente $\nabla_w J(w)$ 9: Atualizar $w \leftarrow w + \eta \nabla_w J(w)$ 10: Normalizar $w \leftarrow w/\|w\|$ 11: **até** convergência ou máximo de iterações12: Projetar dados: $Y = Xw^*$ onde w^* é a solução ótima13: Selecionar d melhores direções repetindo o processo ortogonalmente14: **Etapa 2: Otimização de Espaço Não Linear**15: Selecionar função de transformação T (tangente hiperbólica, gaussiana, sigmoide, ou laplaciana)

16: Otimizar parâmetros da transformação via validação cruzada

17: **para** cada classe $k = 1, \dots, c$ **faça**18: Aplicar transformação: $Y_{transformado}^k = T(Y^k, \theta_k)$ 19: Calcular centróide transformado: $\vec{a}^k = \frac{1}{m_k} \sum_{i=1}^{m_k} \vec{y}_{transformado}^i$ 20: Calcular dispersão intraclasse: $v^k = \|\text{Eig}(\text{Cov}(Y_{transformado}^k))\|$ 21: **fim para**

22: Resolver otimização multiobjetivo:

23: $\max \sum_{i < j} \|\vec{a}^i - \vec{a}^j\|$ e $\min \sum_k v^k$ 24: Obter matriz de transformação final $W_{final} \in \mathbb{R}^{d \times c}$ 25: Computar dados finais: $Z = Y \cdot W_{final}$ 26: **Construir Discriminador**27: **para** cada nova amostra \vec{y} **faça**28: Calcular distâncias: $D_k = \|\vec{y} - \vec{a}^k\|$ para $k = 1, \dots, c$ 29: Computar probabilidades: $f_k(\vec{y}) = \frac{D_k}{\sum_{j=1}^c D_j}$ 30: **fim para****retorna** Z, f

4 Metodologia Experimental

Este capítulo apresenta a estrutura experimental desenvolvida para validar a eficácia do método NL-TSDR proposto nesta tese. Toda a implementação experimental foi desenvolvida utilizando MATLAB R2023a, aproveitando suas funcionalidades nativas para processamento de dados, visualização e implementação de algoritmos de aprendizado de máquina. A escolha do MATLAB se justifica pela robusta *Deep Learning Toolbox* e pelos conjuntos de dados de referência já incorporados ao ambiente, facilitando a reprodutibilidade dos experimentos e a comparação com métodos estabelecidos na literatura.

A avaliação experimental é estruturada em cinco etapas complementares, cada uma projetada para examinar diferentes aspectos do desempenho do método proposto. Inicialmente, aborda-se um experimento de visualização de dados (seção 4.1), utilizando o conjunto de dados `crab_dataset` para demonstrar graficamente os efeitos da aplicação do método. Esta abordagem visual permite uma compreensão intuitiva de como o NL-TSDR transforma o espaço de características, melhorando a separabilidade entre classes através da redução de dimensionalidade e otimização das diferenças de distribuição.

Em seguida, é abordada a definição dos métodos de referência (seção 4.2), estabelecendo um conjunto abrangente de algoritmos de classificação que servirão como *baseline* para comparação. Estes métodos incluem técnicas determinísticas e não determinísticas, representando diferentes paradigmas de aprendizado de máquina, desde abordagens clássicas como k -NN e até métodos mais sofisticados como RNAs.

A terceira etapa (seção 4.3) envolve a escolha criteriosa dos conjuntos de dados de referência, selecionando bases de dados que representam desafios diversos no reconhecimento de padrões. Estas bases variam desde problemas balanceados de baixa dimensionalidade até cenários desbalanceados de alta dimensionalidade, permitindo uma avaliação sistemática da robustez do método em diferentes condições experimentais.

Posteriormente, incorpora-se o conjuntos de dados reais do Instagram (seção 4.4), coletados em 2019, para demonstrar a aplicabilidade do método em cenários práticos de análise de redes sociais. Este conjunto de dados apresenta características heterogêneas e relações não lineares complexas, proporcionando um teste realístico da eficácia do NL-TSDR em problemas do mundo real.

Finalmente, processo de avaliação e parametrização dos métodos (seção 4.5) é detalhado, incluindo as métricas empregadas, os protocolos experimentais adotados e a estrutura metodológica ilustrada no diagrama experimental. Esta abordagem sistemática garante que a avaliação seja abrangente, rigorosa e capaz de fornecer conclusões confiáveis sobre o desempenho do método proposto em diferentes contextos de aplicação.

4.1 Experimento com Visualização de Dados

É possível avaliar a eficácia do método proposto visualizando seus efeitos em um determinado conjunto de dados. Neste capítulo, é utilizado inicialmente o conjunto de dados de gênero de caranguejo (`crab_dataset`), disponível no repositório de aprendizado de máquina da UCI (*University of California, Irvine*) e incorporado aos conjuntos de dados de amostra da *Deep Learning Toolbox* do MATLAB.

Nesse *dataset*, existem 200 indivíduos, distribuídos igualmente em 2 classes (masculino ou feminino), caracterizados por 6 atributos físicos (espécie, lábio frontal, largura traseira, comprimento, largura e profundidade) diferentes. É um conjunto de dados simples, que permite avaliar graficamente a distribuição dos indivíduos de acordo com suas respectivas classes.

Um bom ponto de partida é avaliar a correlação entre cada uma das características (*features*), visualizando como elas influenciam (ou não) na classificação dos indivíduos (*labels*). Como no conjunto de dados em questão existem 6 atributos diferentes, é possível realizar quinze comparações distintas entre elas, considerando que essa quantidade é determinada pela seguinte equação, onde q é a quantidade de comparações e n o número de atributos, assumido que $n \geq 2$.

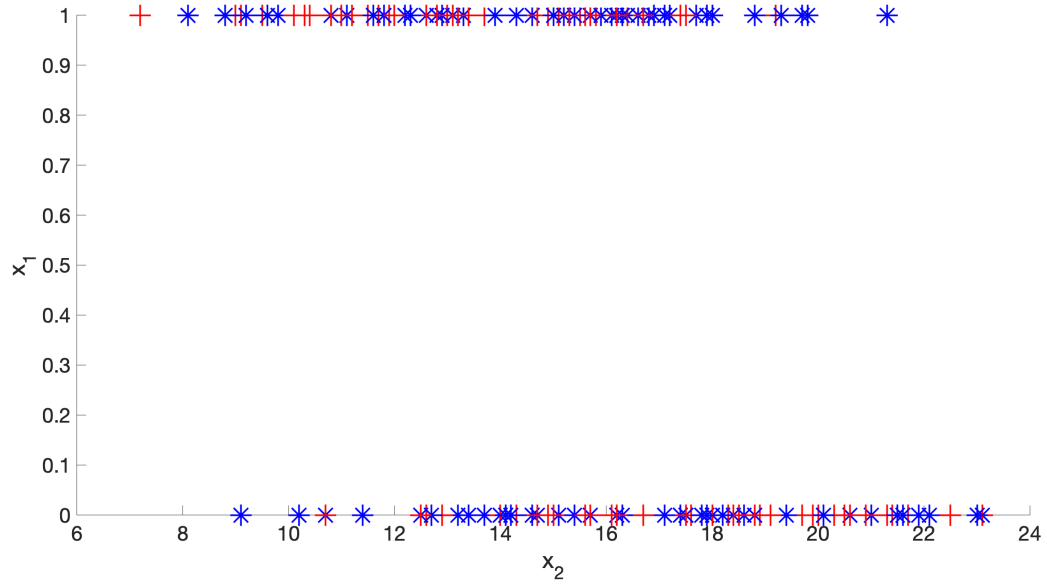
$$q = \sum_{i=2}^{n-1} |1 - n| \quad (4.1)$$

Da [Figura 4.1](#) à [Figura 4.15](#) estão as quinze comparações, de acordo com a quantidade determinada pela [Equação 4.1](#). Imediatamente é possível notar como algumas das características possuem uma acentuada correlação, ao passo em que outras parecem não desempenhar a menor influência na determinação da classe do indivíduo. Interessante também é notar como em alguns casos os indivíduos de diferentes classes são mais separáveis uns dos outros, ao passo em que em outras situações nota-se uma maior sobreposição entre esses indivíduos.

Dessa maneira, para cada comparação, é calculado o Coeficiente de Correlação de Pearson (ρ), que denota o grau de relação entre duas variáveis quantitativas e representa seu grau de correlação ao longo de uma escala de valores situados entre -1 e 1 , conforme detalhado na [subseção 2.5.1](#).

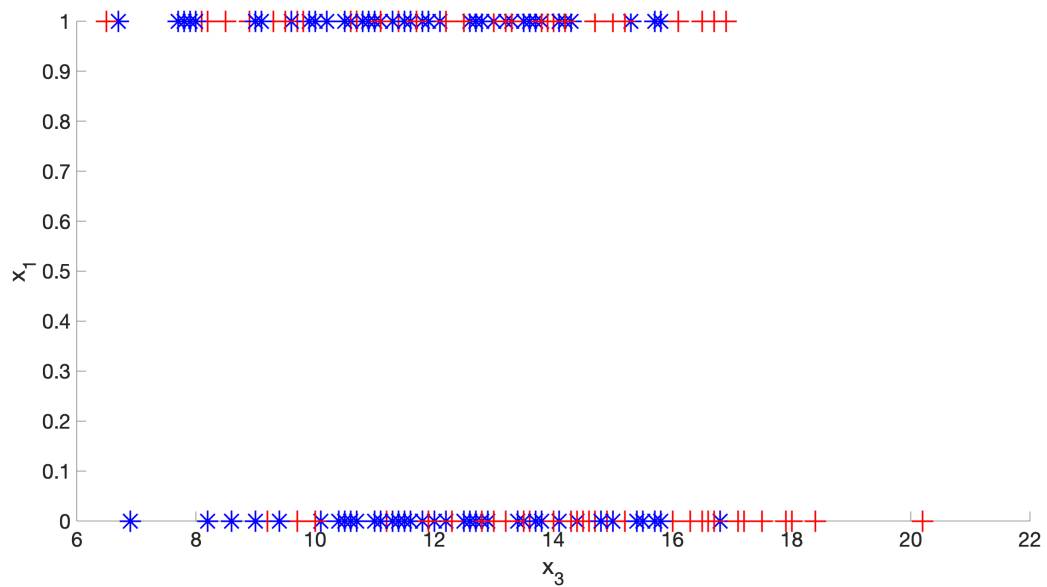
Pautado pelo método proposto, o primeiro passo é aplicar a redução de dimensionalidade, de forma a garantir que as melhores (mais relevantes) características do conjunto de dados sejam extraídas. Dessa maneira, conforme explanado na [seção 3.1](#), é preciso determinar em quantas dimensões o *dataset* deverá ser reduzido. Arbitrariamente, uma vez que a ideia é experimentar com a visualização de dados, determina-se um valor de $d = 4$, ou seja, o subespaço ideal encontrado terá 4 *features*.

Figura 4.1 – Dispersão das classes por características x_1 e x_2 ($\rho = -0,4380$).



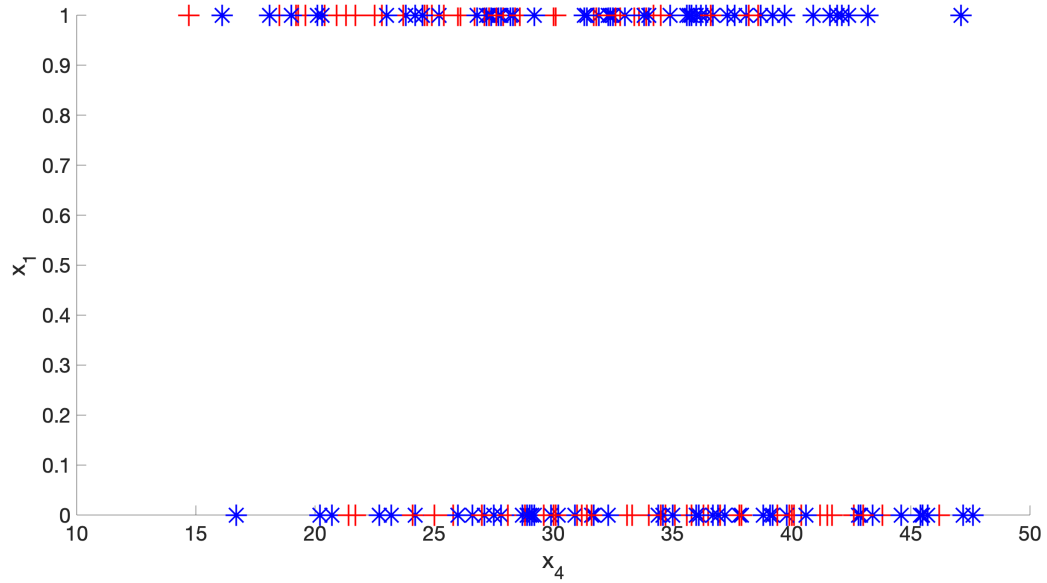
Fonte: Elaborado pelo autor.

Figura 4.2 – Dispersão das classes por características x_1 e x_3 ($\rho = -0,3158$).



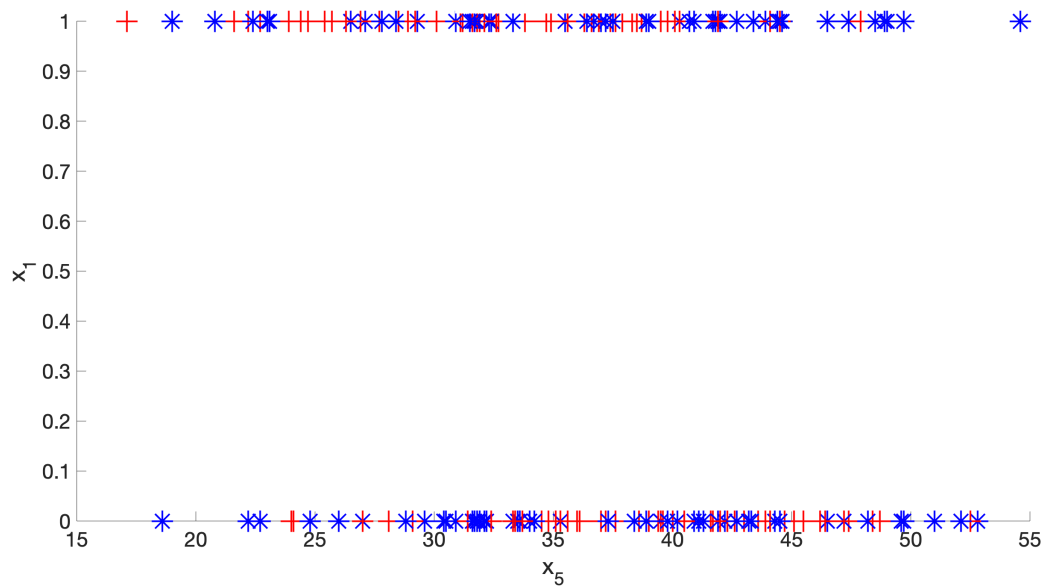
Fonte: Elaborado pelo autor.

Figura 4.3 – Dispersão das classes por características x_1 e x_4 ($\rho = -0,2883$).



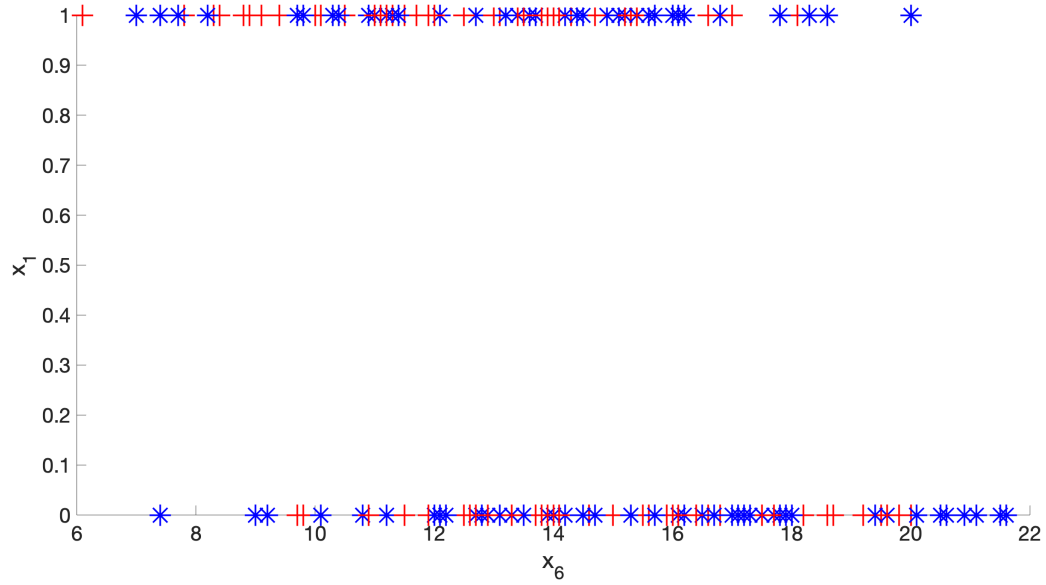
Fonte: Elaborado pelo autor.

Figura 4.4 – Dispersão das classes por características x_1 e x_5 ($\rho = -0,2162$).



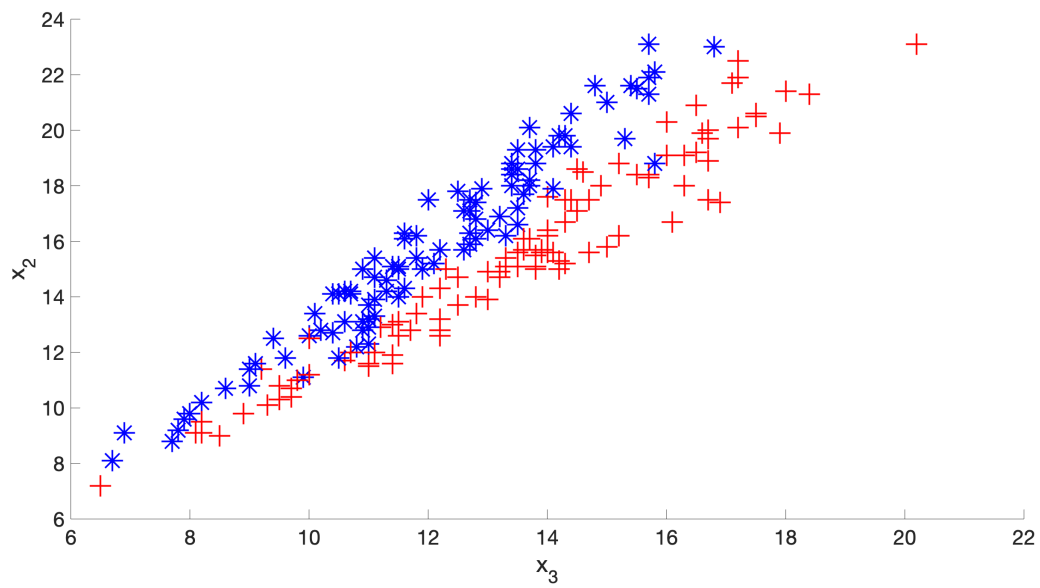
Fonte: Elaborado pelo autor.

Figura 4.5 – Dispersão das classes por características x_1 e x_2 ($\rho = -0,4237$).



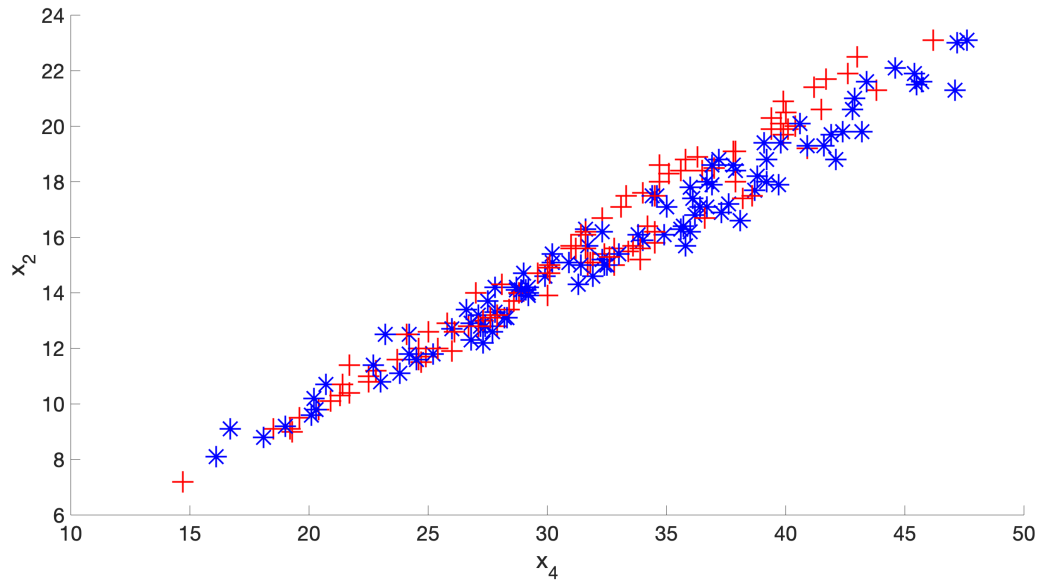
Fonte: Elaborado pelo autor.

Figura 4.6 – Dispersão das classes por características x_2 e x_3 ($\rho = 0,9070$).



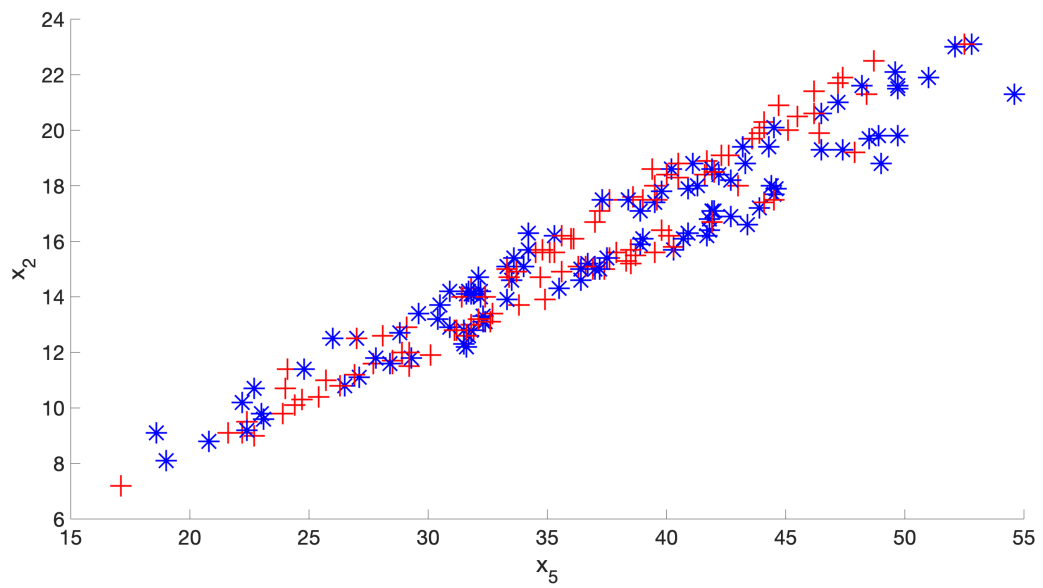
Fonte: Elaborado pelo autor.

Figura 4.7 – Dispersão das classes por características x_2 e x_4 ($\rho = 0,9788$).



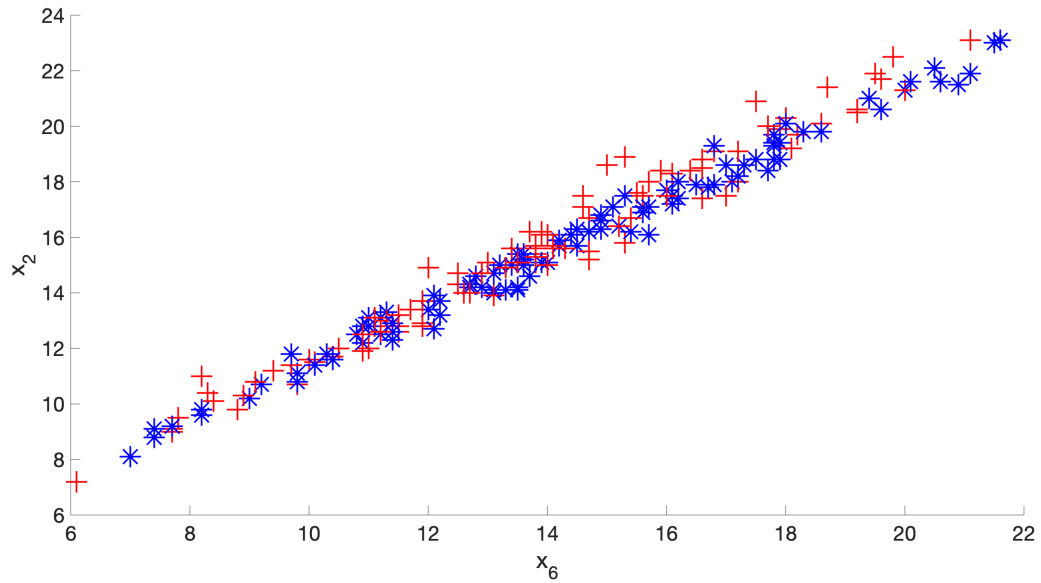
Fonte: Elaborado pelo autor.

Figura 4.8 – Dispersão das classes por características x_2 e x_5 ($\rho = 0,9650$).



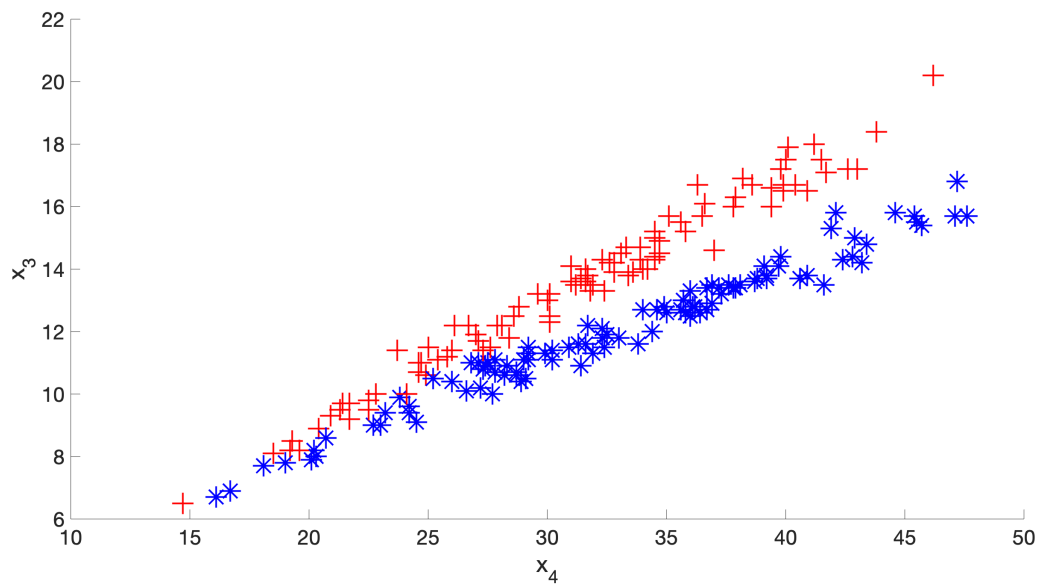
Fonte: Elaborado pelo autor.

Figura 4.9 – Dispersão das classes por características x_2 e x_6 ($\rho = 0,9876$).



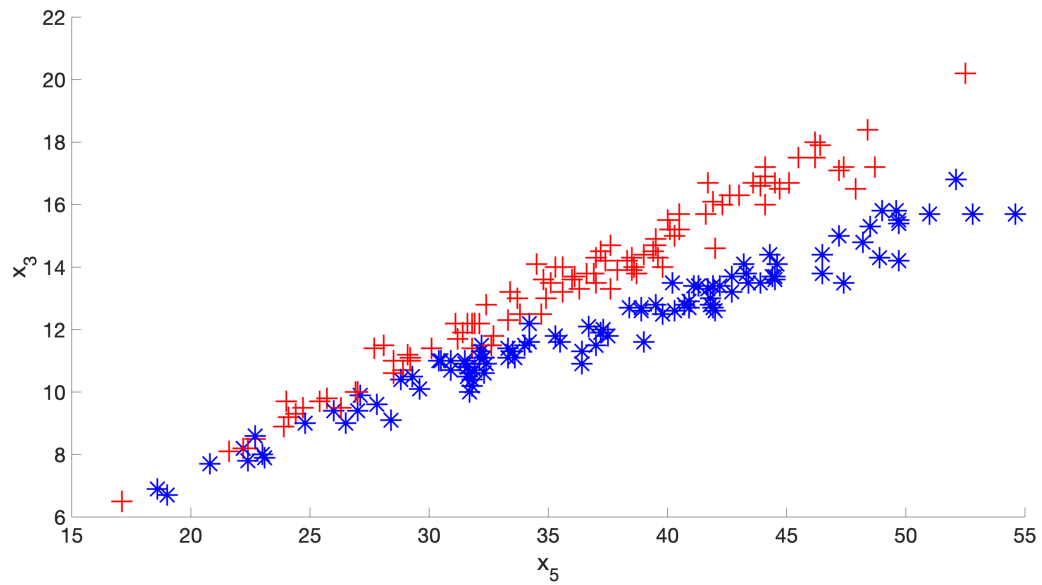
Fonte: Elaborado pelo autor.

Figura 4.10 – Dispersão das classes por características x_3 e x_4 ($\rho = 0,8927$).



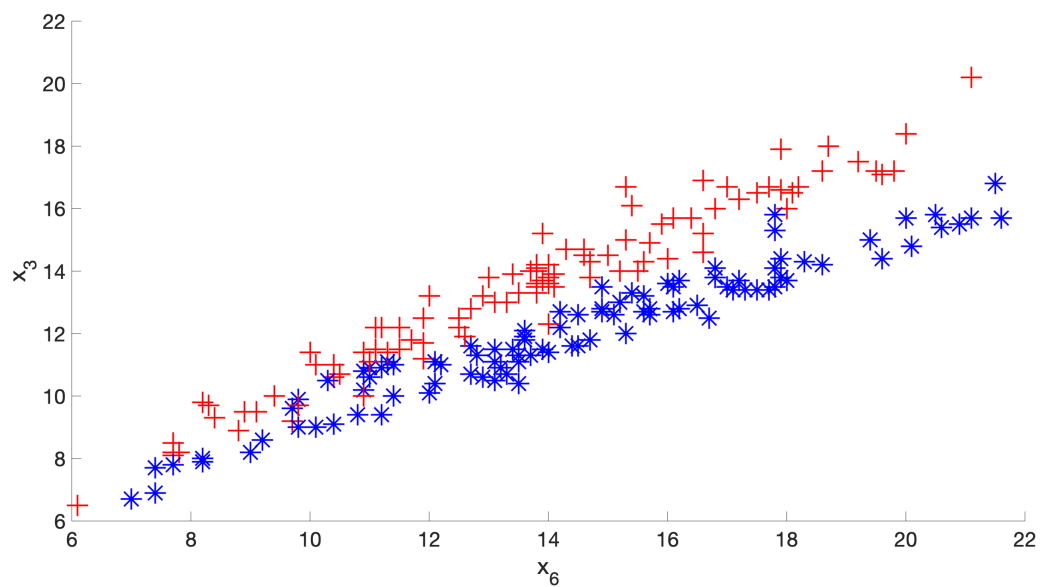
Fonte: Elaborado pelo autor.

Figura 4.11 – Dispersão das classes por características x_3 e x_5 ($\rho = 0,9004$).



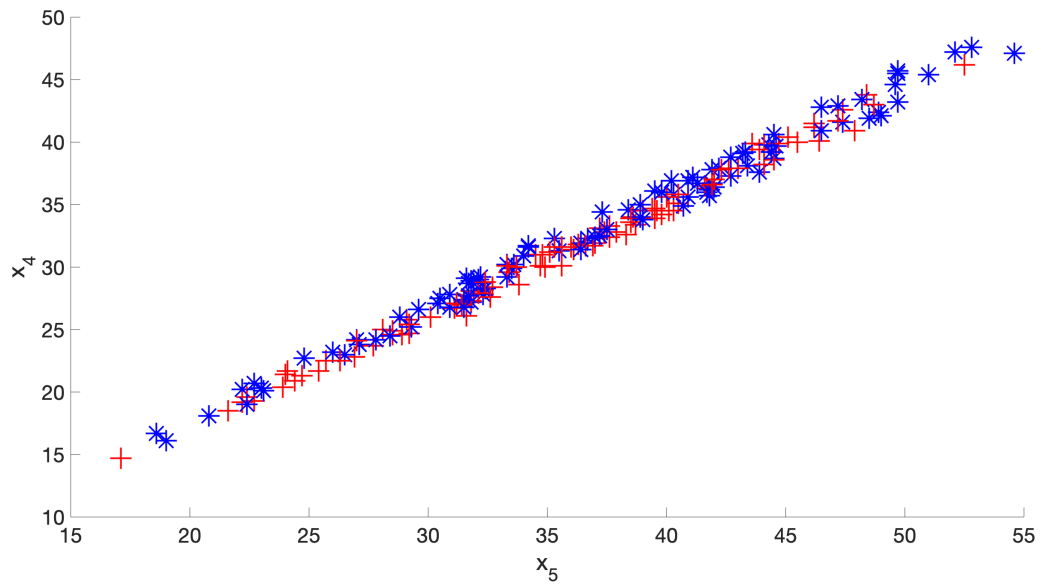
Fonte: Elaborado pelo autor.

Figura 4.12 – Dispersão das classes por características x_3 e x_6 ($\rho = 0,8892$).



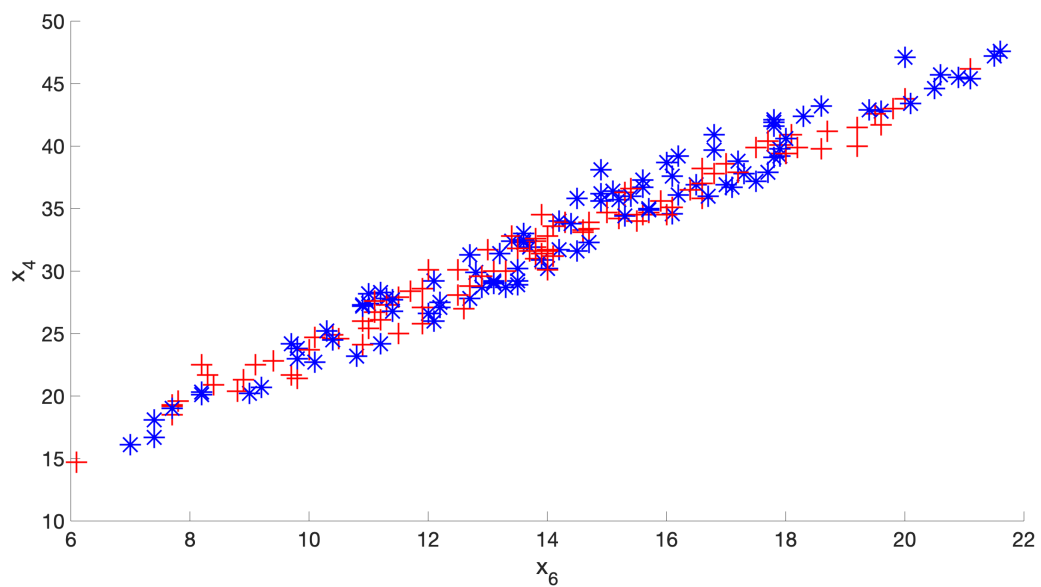
Fonte: Elaborado pelo autor.

Figura 4.13 – Dispersão das classes por características x_4 e x_5 ($\rho = 0,9950$).

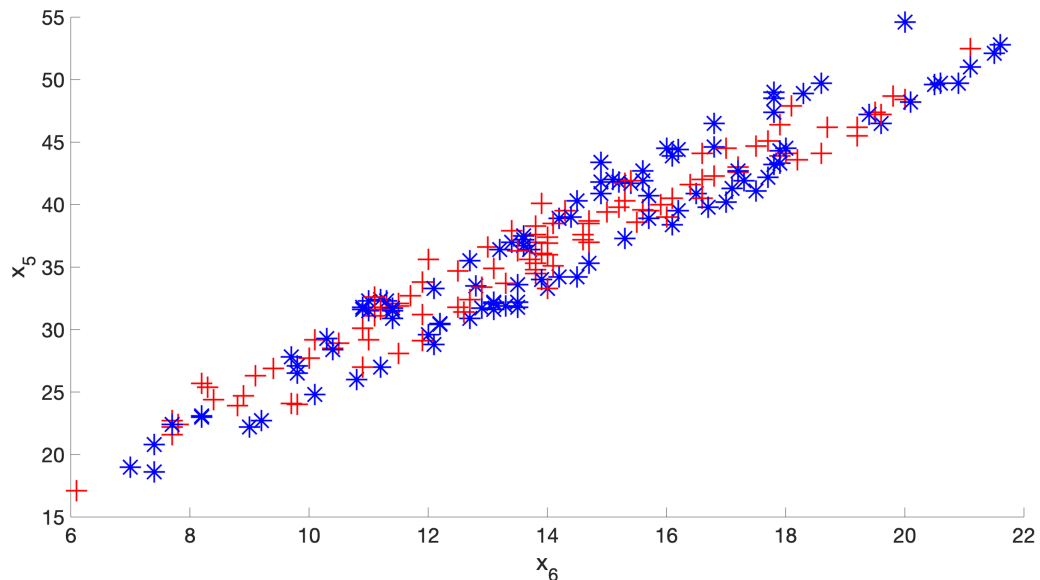


Fonte: Elaborado pelo autor.

Figura 4.14 – Dispersão das classes por características x_4 e x_6 ($\rho = 0,9832$).



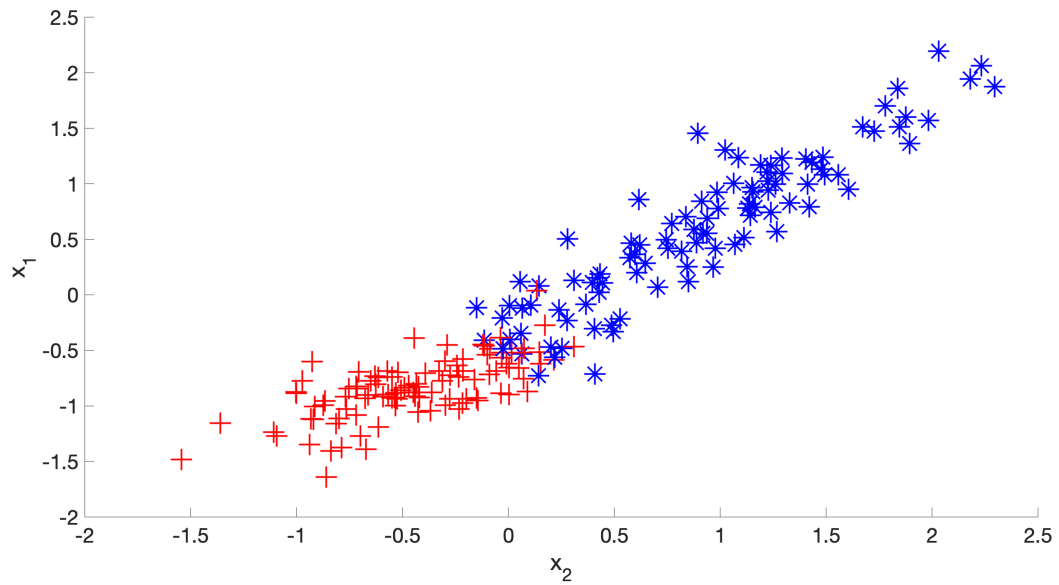
Fonte: Elaborado pelo autor.

Figura 4.15 – Dispersão das classes por características x_5 e x_6 ($\rho = 0,9678$).

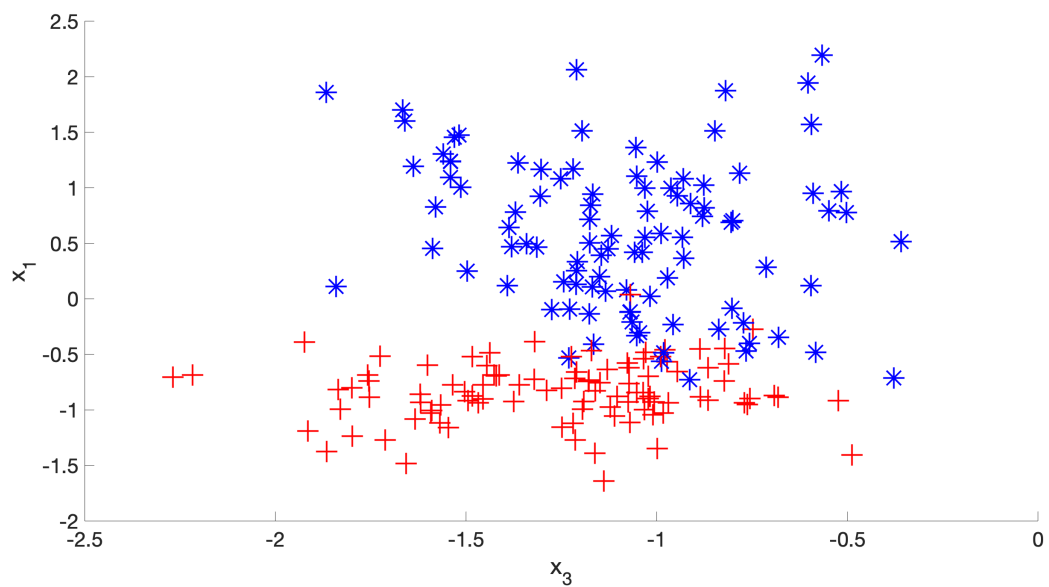
Fonte: Elaborado pelo autor.

Se o conjunto de dados foi reduzido em 4 dimensões, de acordo com a [Equação 4.1](#) existem 6 possíveis comparações. Conforme pode ser observado da [Figura 4.16](#) à [Figura 4.21](#), o método de redução de dimensionalidade faz um bom trabalho ao selecionar as características mais relevantes para a determinação da classe de um indivíduo. É notável que os pontos nos gráficos estão muito mais separáveis, com algumas poucas sobreposições entre eles. Dessa maneira, o algoritmo de classificação receberá um conjunto de dados que teoricamente possui um elevado grau de distinção entre seus indivíduos quando comparado ao conjunto original, aumentando suas chances de apresentar uma maior acurácia.

Assim, conforme explanado na [seção 3.2](#), a etapa de otimização irá encontrar uma função de transformação para o subconjunto de dados ótimo encontrado após a redução de dimensionalidade que maximiza a distância entre os centros de gravidade de indivíduos de diferentes classes, ao mesmo tempo que minimiza a dispersão de instâncias dentro de uma mesma classe. Na [Figura 4.23](#) é possível avaliar graficamente a distribuição de classes para os indivíduos considerando o conjunto de dados $Y_{m_k \times p}^k$, assim como é possível notar visualmente a maior separabilidade dos indivíduos. Ainda que a visualização por meio de gráficos ofereça um modo tangível de avaliar como o algoritmo manipula o conjunto de dados, é importante utilizar um indicador que ofereça uma métrica no contexto de separabilidade e clusterização dos dados. Para isso, existe o Coeficiente de *Silhouette*, utilizado para mensurar a qualidade de uma determinada técnica de agrupamento, conforme detalhado em [subseção 2.5.2](#).

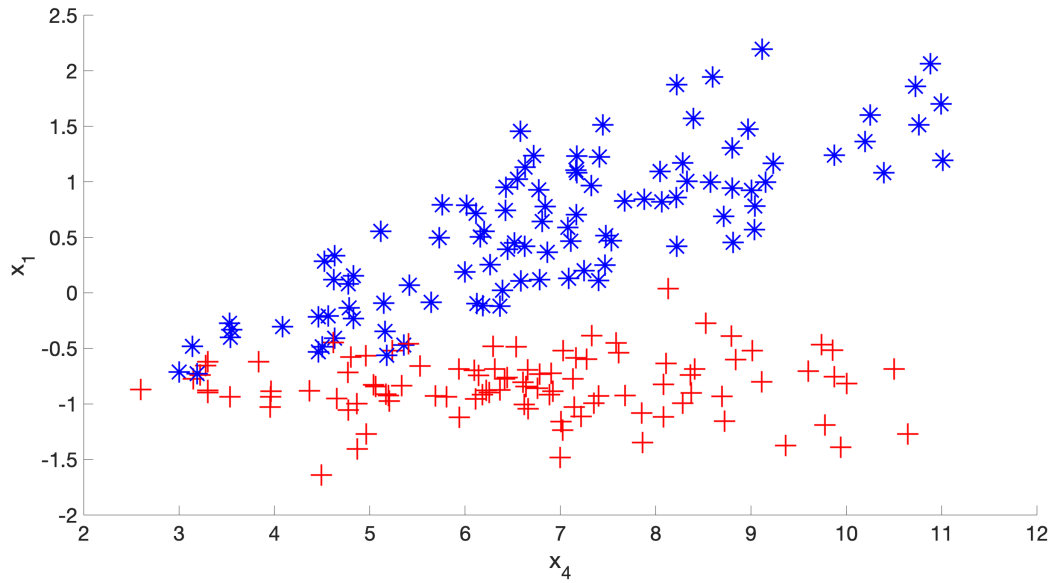
Figura 4.16 – Dispersão das classes por características x_1 e x_2 ($d = 4$ e $\rho = 0,9505$).

Fonte: Elaborado pelo autor.

Figura 4.17 – Dispersão das classes por características x_1 e x_3 ($d = 4$ e $\rho = 0,1465$).

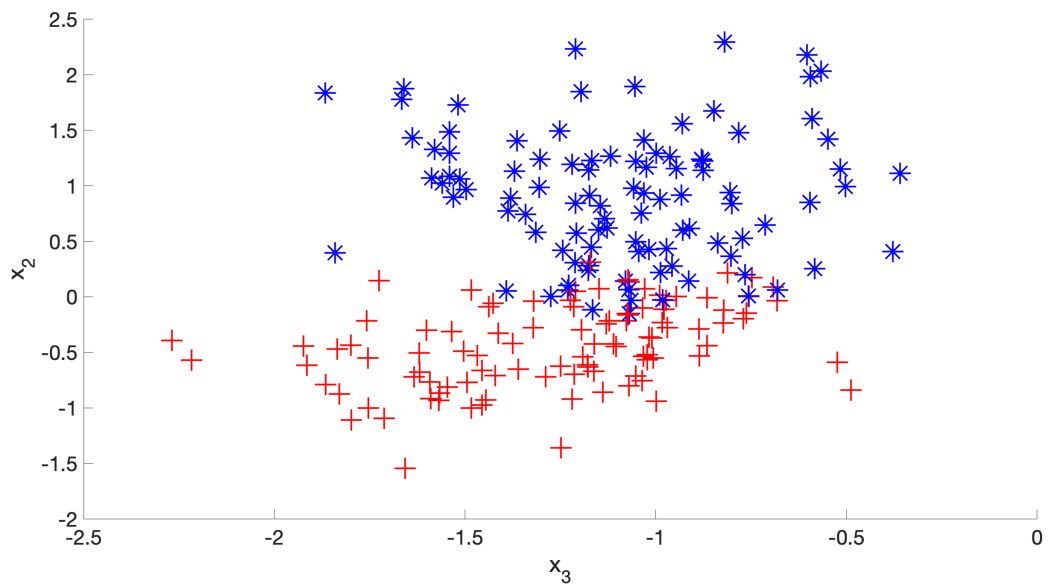
Fonte: Elaborado pelo autor.

Figura 4.18 – Dispersão das classes por características x_1 e x_4 ($d = 4$ e $\rho = 0,3884$).



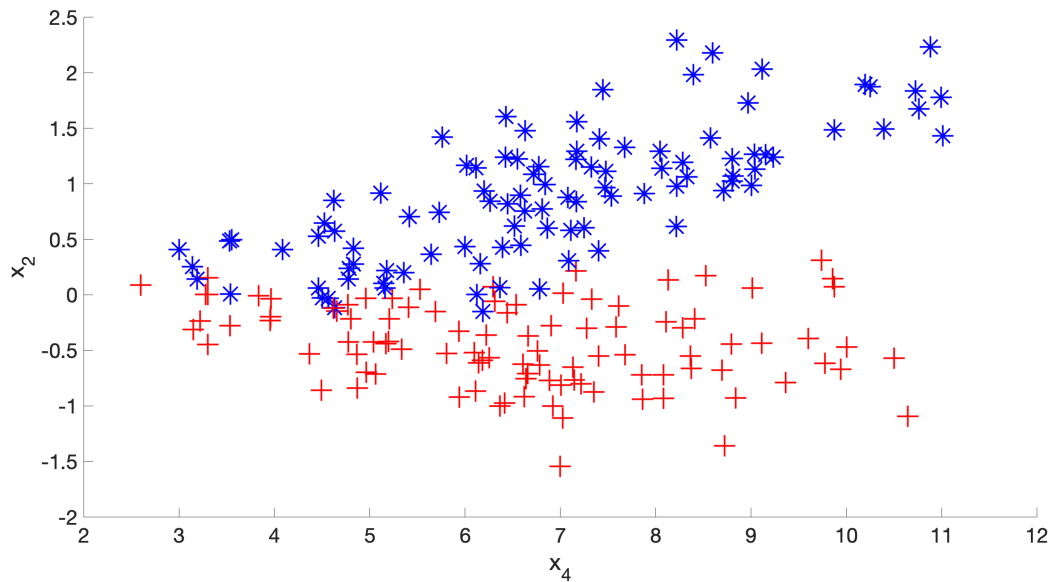
Fonte: Elaborado pelo autor.

Figura 4.19 – Dispersão das classes por características x_2 e x_3 ($d = 4$ e $\rho = 0,2727$).



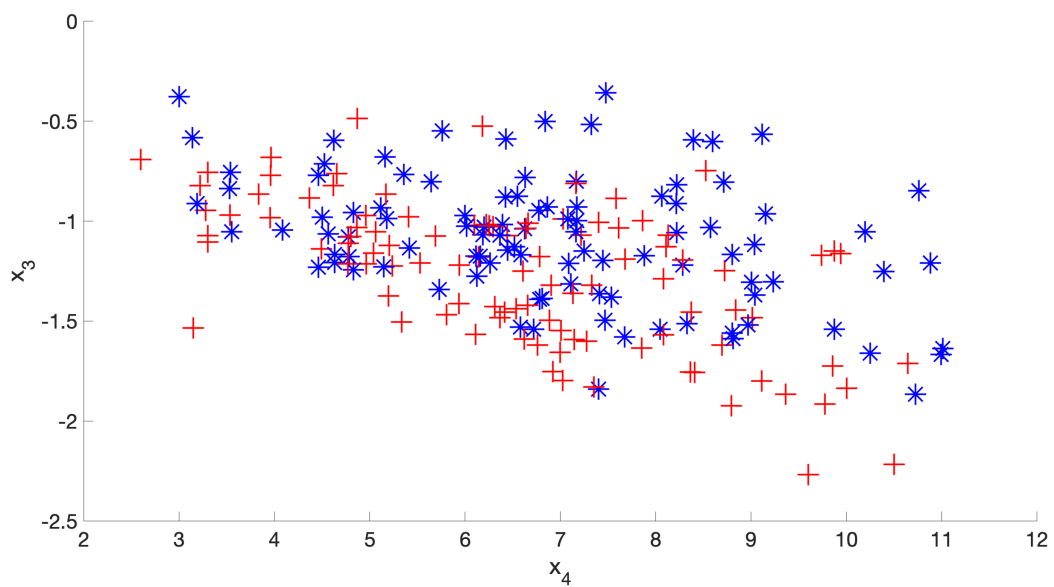
Fonte: Elaborado pelo autor.

Figura 4.20 – Dispersão das classes por características x_2 e x_4 ($d = 4$ e $\rho = 0,2820$).



Fonte: Elaborado pelo autor.

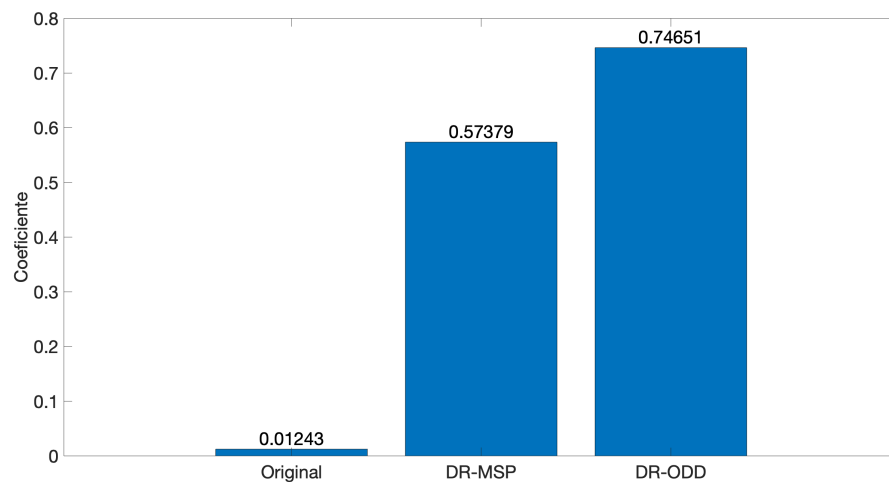
Figura 4.21 – Dispersão das classes por características x_1 e x_2 ($d = 4$ e $\rho = -0,4650$).



Fonte: Elaborado pelo autor.

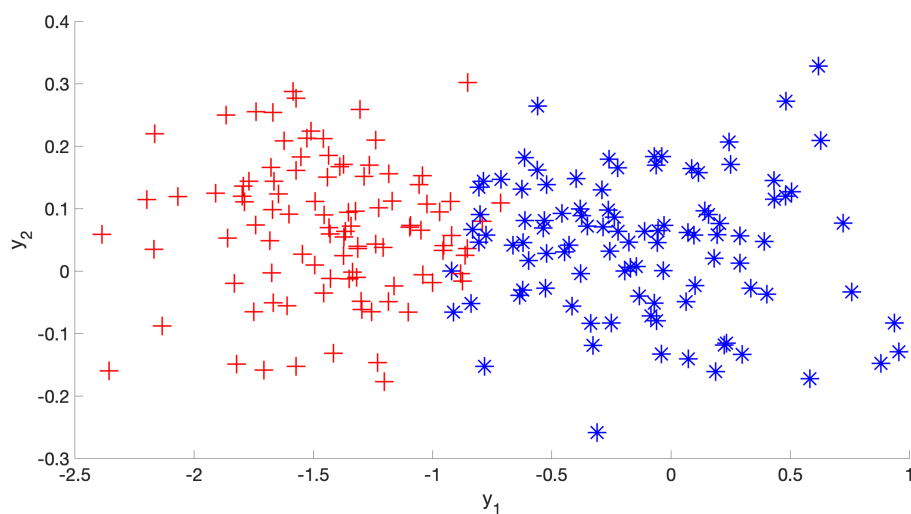
O Coeficiente de *Silhouette* para o conjunto de dados analisado neste capítulo é originalmente de 0,01243. Após a aplicação da redução de dimensionalidade, esse valor sobe para 0,57379. Finalmente, após a aplicação da transformação realizada pela otimização das diferenças de distribuição, o valor encontrado é de 0,74651. A melhora progressiva ratifica a eficácia do método proposto, conforme ilustrado na [Figura 4.22](#) e também na [Figura 4.23](#).

Figura 4.22 – Coeficiente de *Silhouette* para o `crab_dataset`.



Fonte: Elaborado pelo autor.

Figura 4.23 – Classificação do subespaço obtido pelo TSDR.



Fonte: Elaborado pelo autor.

4.2 Métodos de Classificação de Referência

A validação experimental do método NL-TSDR proposto requer a seleção de algoritmos de classificação que sirvam como referência para comparação de desempenho. Para esta finalidade, foram escolhidos três métodos que representam diferentes paradigmas de aprendizado de máquina, todos discutidos na [seção 2.4](#): k -NN, LDA e MLP. Esta seleção estratégica permite uma avaliação abrangente que aborda desde métodos não paramétricos baseados em proximidade até técnicas de modelagem linear e não linear permitindo uma avaliação ampla e compreensiva da técnica proposta.

O algoritmo k -NN ([subseção 2.4.2](#)) foi selecionado devido à sua natureza determinística e simplicidade conceitual. Este método não paramétrico classifica instâncias baseando-se na proximidade no espaço de características, atribuindo a classe majoritária entre os k vizinhos mais próximos.

Por não possuir um modelo interno complexo, o k -NN serve como um indicador direto da qualidade das transformações aplicadas aos dados: melhorias na separabilidade das classes após a aplicação do NL-TSDR devem refletir-se diretamente no desempenho do classificador. Para os experimentos, utilizou-se $k = \sqrt{n}$, onde n representa o número de instâncias de treinamento.

A Análise Discriminante Linear ([subseção 2.4.3](#)) complementa a avaliação representando métodos de classificação com fundamentos estatísticos sólidos que assumem separabilidade linear dos dados. A LDA busca projeções que maximizem a separação entre classes enquanto minimizam a variação intraclasse, sendo fundamentada na suposição de distribuições gaussianas com matrizes de covariância iguais.

A inclusão deste método é particularmente relevante porque permite examinar especificamente como transformações lineares impactam classificadores que dependem de separabilidade linear. Como a LDA possui limitações inerentes a fronteiras de decisão lineares, espera-se que se beneficie significativamente de pré-processamentos que melhorem a organização linear dos dados, tornando-se um excelente indicador da eficácia do primeiro estágio do NL-TSDR.

O MLP ([subseção 2.4.4](#)) foi incluído como representante dos métodos de classificação não linear, proporcionando uma perspectiva essencial sobre a eficácia das transformações propostas. O MLP, com sua capacidade de aproximação universal através de múltiplas camadas de neurônios interconectados, pode teoricamente aprender qualquer mapeamento não linear entre características e classes.

A arquitetura utilizada emprega uma única camada oculta contendo $2n/3 + c$ neurônios, onde n representa o número de características e c o número de classes, seguindo práticas estabelecidas que equilibram capacidade de representação e risco de *overfitting*.

A natureza não determinística do MLP, derivada da inicialização aleatória de pesos e algoritmos de otimização estocásticos, requer considerações especiais na avaliação. Cada experimento é repetido múltiplas vezes com diferentes sementes aleatórias para obter estimativas estatisticamente robustas. No contexto desta pesquisa, o MLP representa um teste: se o NL-TSDR proporcionar melhorias significativas mesmo para um classificador com capacidade de modelagem não linear sofisticada, isso indicaria que as transformações facilitam o processo de otimização, possivelmente evitando mínimos locais subótimos ou acelerando a convergência para soluções de melhor qualidade.

A combinação destes três métodos garante uma avaliação experimental multifacetada que captura diferentes aspectos da eficácia do NL-TSDR. O k -NN oferece medidas diretas da qualidade das transformações; a LDA avalia especificamente benefícios para separabilidade linear; e o MLP testa se as transformações oferecem vantagens mesmo quando o classificador possui flexibilidade inerente para descobrir relações complexas nos dados originais. Esta diversidade metodológica fortalece a validade das conclusões experimentais e reforça a robustez das contribuições propostas em diferentes contextos.

4.3 Conjuntos de Dados de Referência

A avaliação experimental engloba cinco conjuntos de dados cuidadosamente selecionados para representar desafios diversos de reconhecimento de padrões. Os conjuntos de dados de referência, resumidos em [Tabela 4.1](#), variam de problemas balanceados de baixa dimensionalidade a cenários desbalanceados de alta dimensionalidade, permitindo avaliação sistemática da robustez do NL-TSDR através de condições variadas.

O `crab_dataset` fornece um cenário balanceado de baixa dimensionalidade ideal para avaliação de desempenho de linha de base e validação de método, apresentando distribuições de classes iguais que eliminam efeitos confusos do desbalanceamento de classes. O `cancer_dataset` introduz desbalanceamento moderado (65,5%/34,5%) enquanto mantém dimensionalidade gerenciável, testando a robustez do método a cenários realistas de distribuição de classes. Em contraste, o `glass_dataset` apresenta desbalanceamento significativo de classes (23,8%/76,2%), desafiando todos os métodos de classificação com disparidade substancial de classes que potencialmente leva a previsões enviesadas em direção à classe majoritária.

O `ovarian_dataset` representa cenários de alta dimensionalidade com 100 características, avaliando diretamente o desempenho sob condições de maldição da dimensionalidade onde métodos tradicionais potencialmente falham ([HE; GARCIA, 2009](#)). Ela representa um cenário que combina alta dimensionalidade com tamanho moderado de amostra (216 instâncias), testando o desempenho sob condições de maldição da dimensionalidade.

Tabela 4.1 – Características dos conjuntos de dados de referência

Conjunto	Características	Classes	Amostras	Distribuição de Classes
crab_dataset	6	2	200	{100, 100}
cancer_dataset	9	2	699	{458, 241}
glass_dataset	9	2	214	{51, 163}
ovarian_dataset	100	2	216	{121, 95}

Fonte: Elaborado pelo autor com base na *Deep Learning Toolbox do MATLAB*.

Esta estrutura métrica abrangente permite uma avaliação através de diferentes aspectos do desempenho do método enquanto considera os desafios específicos apresentados por cada base de dados. A combinação das métricas de avaliação exploradas em [seção 2.5](#) fornece uma avaliação que aborda tanto melhorias locais de nível de característica quanto aprimoramento global de estrutura de classe, permitindo avaliação da eficácia do NL-TSDR através de cenários diversos de classificação.

4.4 Conjunto de Dados do Instagram

Para demonstrar aplicabilidade em cenários do mundo real, avalia-se o NL-TSDR na predição de engajamento do Instagram usando dados coletados em 2019 ([SOBRINHO, 2019](#)). Esta base de dados compreende 6.522 registros de mídia de 606 usuários únicos, representando tipos de conteúdo diversos e padrões de engajamento típicos de aplicações de análises de redes sociais. Cada registro inclui 11 características, variando de atributos de conteúdo a informações demográficas.

Os dados foram coletados da rede social Instagram a partir de 25 diferentes usuários. Para cada um deles, foram analisadas 2.522 interações, por meio da análise das informações de cada uma das (até) últimas 12 publicações (imagens ou vídeos), repetindo esse mesmo processo para todos os usuários que nelas interagiram. Foi construído um conjunto com 6.522 registros de mídias, pertencentes a 606 usuários distintos, enriquecidos com informações a respeito da idade, gênero, cor do cabelo e demais dados cognitivos.

A estruturação é um trabalho realizado em dois atos. O primeiro deles é manipular os dados brutos, ainda em Node.js, para criar um arquivo JSON com todas as *features* e *labels* devidamente tratado. O segundo é importar esses dados para o MATLAB, rotulando corretamente cada um dos registros e criando duas matrizes (x e t), com os respectivos dados de entrada e saída. Foram considerados os seguintes atributos na matriz x .

- Largura da mídia em *pixels*.
- Altura da mídia em *pixels*.

- Quantidade de *hashtags* utilizadas.
- Comprimento (em número de caracteres) da legenda.
- Quantidade de pessoas do sexo masculino identificada na mídia.
- Média de idade das pessoas do sexo masculino identificadas na mídia.
- Quantidade de pessoas do sexo feminino identificada na mídia.
- Média de idade das pessoas do sexo feminino identificadas na mídia.
- Sexo do usuário do perfil proprietário da mídia.
- Idade do usuário do perfil proprietário da mídia.
- Cor do cabelo do usuário do perfil proprietário da mídia.

O conjunto de dados compreende 6.522 registros de mídia classificados binariamente quanto ao engajamento: 3.835 amostras (58,8%) como alto engajamento e 2.687 (41,2%) como baixo engajamento. Esta classificação segue padrões estabelecidos da indústria (TRUNFIO; ROSSI, 2021), onde o engajamento é mensurado pela razão entre interações totais (curtidas e comentários) e número de seguidores. O limiar de 6% para alto engajamento alinha-se com marcos reconhecidos do marketing digital e oferece interpretação prática para aplicações comerciais.

Esta base de dados Instagram constitui um ambiente de teste robusto para algoritmos de mascaramento de classes devido às suas características intrínsecas. O conjunto apresenta heterogeneidade de atributos com escalas e distribuições variadas, relações não lineares complexas entre características de usuário e métricas de engajamento, além de ruído natural proveniente da variabilidade comportamental dos usuários. Particularmente relevante é o potencial de mascaramento através de padrões de engajamento sobrepostos entre diferentes segmentos demográficos e categorias de conteúdo, fornecendo assim uma avaliação realística da capacidade do NL-TSDR em abordar desafios práticos de aprendizado de máquina.

4.5 Processo de Avaliação e Parametrização dos Métodos

O protocolo experimental avalia sistematicamente múltiplas configurações para fornecer avaliação abrangente de desempenho. Aplicam-se métodos originais a bases de dados não modificadas para estabelecer desempenho de linha de base, então são avaliados os mesmos métodos tanto em espaços de características lineares reduzidos quanto em espaços de características não lineares reduzidos pelo NL-TSDR para avaliar a contribuição de cada estágio.

A avaliação emprega a acurácia como métrica primária. No entanto, reconhecendo que a acurácia sozinha pode ser enganosa para bases de dados desbalanceadas, complementa-se isso com o *F1 Score*, calculado como a média harmônica de *Precision* e *Recall*, conforme detalhado na [subseção 2.5.3.3](#). O *F1 Score* fornece avaliação equilibrada, onde a precisão representa a razão de observações positivas corretamente preditas para o total de observações positivas preditas, e a revocação mede a razão de observações positivas corretamente preditas para todas as observações positivas reais.

Para avaliar de forma abrangente a eficácia da abordagem de redução de dimensionalidade de duas etapas, são empregadas medidas quantitativas adicionais que fornecem detalhes sobre relevância de características e características de separabilidade de classes ao longo do processo de transformação. Para avaliar a relevância de características dentro dos espaços de características originais e transformados, utiliza-se o Coeficiente de Correlação de Pearson (ρ). Este coeficiente quantifica a força da relação linear entre variáveis de características e rótulos de classe, uma medida sobre quão efetivamente o método preserva e aprimora informações discriminativas.

Conforme descrito na [subseção 2.5.1](#), o coeficiente de correlação varia de -1 a 1, onde valores aproximando-se de 1 indicam relações lineares positivas fortes, valores aproximando-se de -1 sugerem correlações negativas fortes, e valores próximos de zero indicam relações lineares fracas ou ausentes. No contexto da avaliação de redução de dimensionalidade, valores absolutos de correlação mais altos entre características transformadas e rótulos de classe indicam poder discriminativo aprimorado, apoiando a hipótese de que o NL-TSDR efetivamente aborda o mascaramento de classes melhorando relações característica-classe nos diferentes conjuntos de dados.

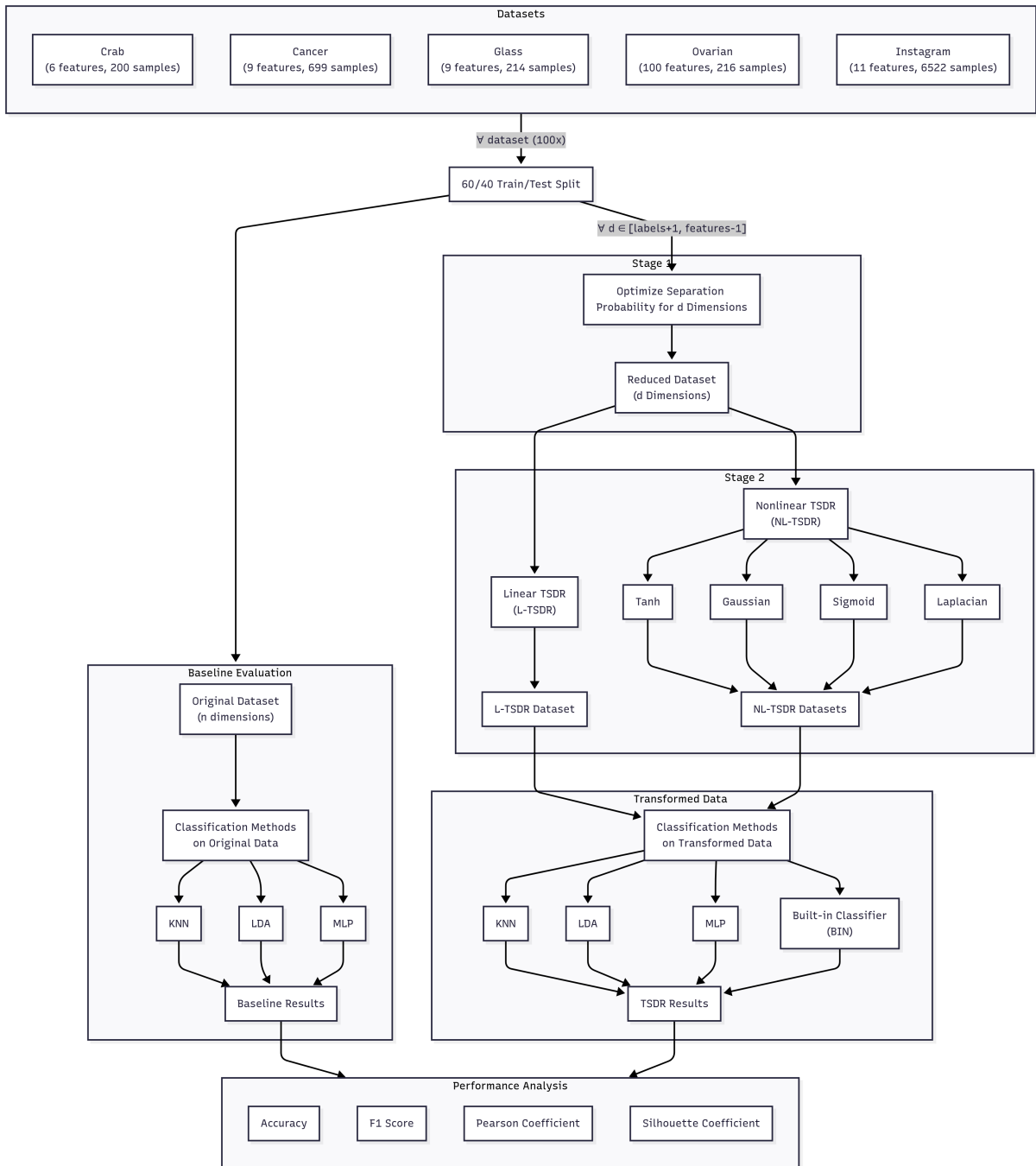
Para avaliar melhorias de separabilidade de classes alcançadas através da abordagem de duas etapas, utiliza-se o Coeficiente de *Silhouette* (s). Esta métrica quantifica a qualidade de separação entre diferentes classes no espaço de características transformado, conforme detalhado em [subseção 2.5.2](#). Para um ponto de dados i , seja a_i representando a distância média intra-classe (distância média para todos os outros pontos dentro da mesma classe) e b_i denotando a distância média inter-classe mais próxima (distância média para todos os pontos na classe diferente mais próxima).

O Coeficiente de *Silhouette* geral para uma base de dados é computado como a média dos valores individuais através de todos os pontos de dados. Valores aproximando-se de 1 indicam classes bem separadas com alta coesão intra-classe e separação inter-classe, valores próximos de 0 sugerem classes sobrepostas ou mal separadas, e valores negativos indicam potencial classificação incorreta ou estrutura de classe pobre. Na estrutura experimental, espera-se observar aumentos sistemáticos nos valores do Coeficiente de *Silhouette* conforme os dados progridem através do Estágio 1 e Estágio 2, demonstrando a melhoria progressiva na separabilidade de classes que o método alcança.

Estas métricas são computadas em três estágios críticos da avaliação: em espaços de características originais para estabelecer características de separabilidade de linha de base, após transformação do Estágio 1 para avaliar melhorias de separabilidade linear, e após transformação completa NL-TSDR para quantificar a eficácia do método completo. Na [Figura 4.24](#), ilustra-se a estrutura experimental abrangente projetada para avaliar a eficácia do método NL-TSDR através de cenários de classificação diversos.

Para permitir conclusões confiáveis sobre eficácia do método, são conduzidos 100 ensaios independentes para cada combinação entre método de classificação e conjunto de dados, seguindo práticas estabelecidas em experimentação de aprendizado de máquina. Cada ensaio emprega amostragem aleatória estratificada para manter consistência de distribuição de classes através de partições de treinamento e teste, com bases de dados divididas aleatoriamente em conjuntos de treinamento (60%) e teste (40%). Esta estratégia de particionamento fornece dados de treinamento suficientes para aprendizado robusto do modelo enquanto mantém tamanho adequado do conjunto de teste para estimativa confiável de desempenho.

Figura 4.24 – Diagrama metodología experimental



5 Resultados

Os resultados experimentais obtidos demonstram que a abordagem proposta em duas etapas (NL-TSDR) alcança desempenho superior de classificação quando comparada a outros métodos da literatura em diferentes conjuntos de dados, proporcionando simultaneamente eficiência computacional através de redução significativa da dimensionalidade.

A metodologia experimental foi conduzida utilizando cinco conjuntos de dados de referência distintos, cada um apresentando características específicas que permitem avaliar diferentes aspectos do método NL-TSDR, incluindo tanto problemas clássicos de classificação quanto aplicações do mundo real.

Os resultados de acurácia apresentados na [Tabela 5.1](#) revelam melhorias positivas consistentes em todos os conjuntos de dados avaliados, com melhorias especificamente em cenários de alta dimensionalidade. Esta observação contribui para a validação da hipótese central desta pesquisa, que propõe que a redução de dimensionalidade em duas etapas pode não apenas manter, mas efetivamente melhorar o poder discriminativo dos dados, corroborada pela melhoria na acurácia dos métodos de classificação.

Para o `ovarian_dataset`, que contém 100 características, o método NL-TSDR com transformação tangente hiperbólica alcança 96,51% de acurácia utilizando MLP, representando uma melhoria substancial de 8,25 pontos percentuais em relação à linha de base original de 88,26%. Esta melhoria é particularmente significativa considerando o fator de redução de dimensionalidade de $50\times$, demonstrando que o método não apenas mantém, mas aprimora o poder discriminativo enquanto reduz drasticamente a complexidade computacional.

A magnitude desta melhoria sugere que o `ovarian_dataset` continha informações redundantes e ruído dimensional que mascaravam os padrões discriminativos essenciais, e a capacidade do método NL-TSDR de identificar e preservar as direções mais informativas enquanto elimina dimensões ruidosas representa uma contribuição significativa para o processamento de dados biomédicos de alta dimensionalidade.

Mesmo para conjuntos de dados onde o desempenho base já é elevado, como no caso do `cancer_dataset` com 96,22% de acurácia original, o NL-TSDR fornece melhorias mensuráveis enquanto alcança redução de dimensionalidade de $4,5\times$. O método k -NN com transformação sigmoide alcança 97,00% de acurácia, representando uma melhoria de 0,78 pontos percentuais que, embora possa parecer modesta em termos absolutos, é estatisticamente significativa e demonstra a capacidade do método de otimizar espaços de características já bem estruturados.

Tabela 5.1 – Acurácias de classificação (%) em conjuntos de dados de referência

Dataset		Original	L-TSDR	NL-TSDR				Δ
				Tanh	Gaussiana	Sigmoide	Laplaciana	
Crab	KNN	70,84	93,84	92,14	94,53	94,24	94,18	3×
	LDA	94,50	95,06	93,45	93,88	94,20	93,25	
	MLP	94,29	94,75	94,71	95,70	95,93	93,28	
	BIN	–	95,19	93,28	93,73	94,81	93,59	
Cancer	KNN	96,22	96,97	96,33	96,27	97,00	96,68	4,5×
	LDA	95,90	95,67	95,51	95,54	95,84	95,66	
	MLP	96,01	96,77	96,19	96,47	96,60	96,49	
	BIN	–	95,70	95,70	95,38	96,19	95,73	
Glass	KNN	89,27	90,52	90,32	90,32	90,33	92,96	4,5×
	LDA	91,58	91,69	90,39	90,36	90,39	92,48	
	MLP	91,39	91,36	90,89	90,58	90,55	92,48	
	BIN	–	91,82	92,89	92,91	92,91	91,94	
Ovarian	KNN	92,44	96,05	95,35	94,19	95,47	96,05	50×
	LDA	73,60	95,81	96,40	94,65	94,88	95,47	
	MLP	88,26	94,42	96,51	93,02	93,95	96,05	
	BIN	–	95,00	95,81	95,35	94,53	95,47	
Instagram	KNN	78,48	80,00	77,83	75,07	77,90	77,97	5,5×
	LDA	78,26	78,41	80,07	80,72	80,43	81,30	
	MLP	75,06	74,07	72,36	74,15	75,69	73,17	
	BIN	–	61,81	55,29	59,06	53,26	60,36	

Fonte: Elaborado pelo autor.

A coluna original mostra a linha de base sem redução de dimensionalidade; A coluna L-TSDR mostra a acurácia após a aplicação do TSDR com a variação linear conforme conforme [Equação 3.7](#); As demais colunas mostram a acurácia após a aplicação do TSDR com cada transformação, conforme [Equação 3.12](#) à [Equação 3.15](#); Δ é o fator de redução de dimensionalidade; BIN refere-se à função discriminante [Equação 3.11](#); Valores em negrito indicam o melhor desempenho.

O conjunto de dados de engajamento do Instagram revela o valor prático da abordagem para análise de mídias sociais do mundo real, onde o método LDA com transformação laplaciana alcança 81,30% de acurácia, representando uma melhoria de 3,04 pontos percentuais em relação à linha de base original de 78,26%. Esta melhoria representa progresso significativo em um domínio desafiador, onde padrões de engajamento exibem relacionamentos não lineares, complexos e com sobreposição significativa entre classes.

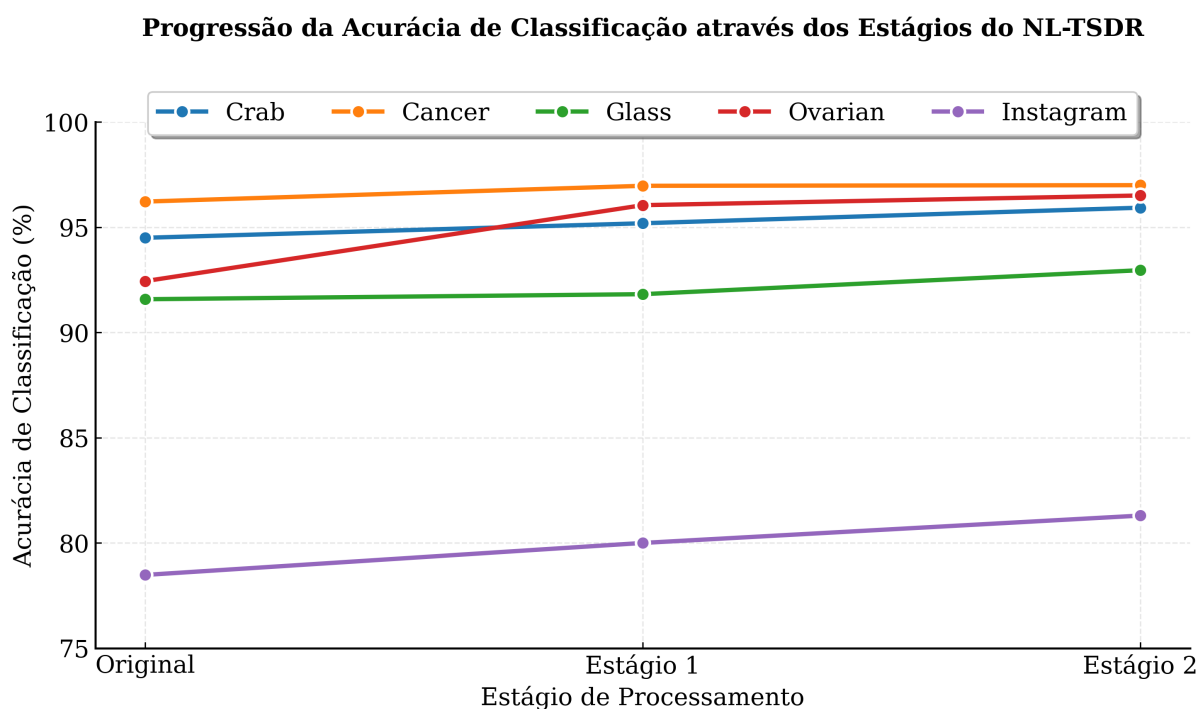
A análise do conjunto de dados de engajamento é particularmente relevante para aplicações práticas, pois os dados de mídias sociais são caracterizados por alta variabilidade, ruído inerente e dependências temporais complexas, e o fato de o método NL-TSDR conseguir melhorias consistentes neste cenário desafiador valida sua aplicabilidade em domínios do mundo real.

O `crab_dataset` e o `glass_dataset` apresentam características intermediárias que permitem avaliar o comportamento do método em cenários de dimensionalidade moderada. Para o `crab_dataset`, a transformação sigmoide com MLP alcança 95,93% de acurácia,

enquanto o `glass_dataset` obtém melhor desempenho com transformação laplaciana, atingindo 92,96% com k -NN. Estes resultados demonstram a importância da seleção adequada da função de transformação baseada nas características específicas dos dados, uma observação que fundamenta a abordagem adaptativa e exploratória proposta nesta pesquisa.

A melhoria progressiva alcançada através de cada estágio do processamento NL-TSDR é apresentada na Figura 5.1, revelando que a maioria dos conjuntos de dados se beneficia de ambos os estágios de processamento, com o Estágio 1 fornecendo melhorias substanciais na linha de base que o Estágio 2 aprimora ainda mais. Esta progressão demonstra que cada estágio contribui de forma para a otimização do espaço de características. O Estágio 1, focado na separação probabilística, identifica as direções mais discriminativas, enquanto o Estágio 2 refina esta representação através de transformações não lineares adaptadas às características específicas dos dados.

Figura 5.1 – Progressão da acurácia da classificação através das etapas



Fonte: Elaborado pelo autor.

A escolha da função de transformação não linear demonstra influenciar o desempenho ótimo, com diferentes funções se destacando para diferentes características dos dados. As transformações laplacianas demonstram força particular para conjuntos de dados desafiadores, com desempenho superior no Instagram alcançando 81,30% de acurácia LDA e no `glass_dataset` atingindo 92,96% de acurácia k -NN.

Estes resultados sugerem que o mecanismo de decaimento exponencial da função laplaciana é efetivo no tratamento de anomalias de engajamento e preservação de classes minoritárias em cenários desbalanceados. A natureza exponencial da transformação laplaciana permite um tratamento suave de *outliers* enquanto preserva a estrutura local dos dados, características particularmente valiosas em dados de mídias sociais onde a presença de comportamentos atípicos é comum.

Por outro lado, a transformação tangente hiperbólica alcança resultados ótimos no `ovarian_dataset` com 96,51% de acurácia MLP, onde a compressão de valores extremos enquanto preserva ordenação relativa aparenta ser particularmente benéfica quando fatores de redução de dimensionalidade são grandes e o tratamento de *outliers* é crítico. Sua capacidade de mapear o domínio para um intervalo limitado $[-1, 1]$ fornece estabilidade numérica enquanto preserva as relações ordinais entre os pontos de dados, características essenciais para *dataset* biomédicos onde a variabilidade inter-paciente pode ser significativa.

A transformação sigmoide, por sua vez, entrega consistentemente resultados competitivos em todas as condições experimentais, alcançando os melhores resultados para `crab_dataset` e `cancer_dataset`. A função sigmoide oferece um compromisso equilibrado entre preservação de informação e regularização, tornando-se uma escolha robusta para cenários onde as características dos dados não são completamente conhecidas.

Os resultados de F1 *Score* apresentados na [Tabela 5.2](#) indicam que as melhorias de acurácia não são alcançadas às custas do equilíbrio entre classes, com F1 *Scores* acompanhando de perto as tendências de acurácia em todos os conjuntos de dados. Esta consistência é particularmente relevante para *datasets* desbalanceados, onde o desequilíbrio de classes poderia potencialmente enviesar classificadores em direção à classe majoritária.

Para conjuntos de dados severamente desbalanceados como o `glass_dataset`, com distribuição 23,8%/76,2%, o método mantém desempenho sólido de F1 *Score*, com transformação laplaciana alcançando 95,37% usando k -NN, uma melhoria de 2,06 pontos percentuais que indica que o método proposto trata efetivamente o mascaramento de classes minoritárias que tipicamente ocorre em cenários desbalanceados de alta dimensionalidade.

Esta capacidade de preservar informações de classes minoritárias é potencialmente relevante para aplicações práticas, onde frequentemente o interesse reside na detecção precisa de casos raros ou atípicos. A manutenção de alto F1 *Score* demonstra que o método NL-TSDR não sacrifica sensibilidade pela especificidade global, uma característica relevante para sua aplicabilidade em diferente domínios.

A [Tabela 5.3](#) quantifica a melhoria progressiva da qualidade do espaço de características através de ambos os estágios do método, com os aumentos sistemáticos tanto no Coeficiente de *Silhouette* quanto nos valores de Correlação de Pearson demonstrando que cada estágio contribui significativamente para a separabilidade aprimorada entre classes.

Tabela 5.2 – F1 Scores de classificação (%) em conjuntos de dados de referência

Dataset		Original	L-TSDR	NL-TSDR				Δ
				Tanh	Gaussiana	Sigmoide	Laplaciana	
Crab	KNN	66,79	93,75	91,60	94,23	93,96	94,03	3×
	LDA	94,44	95,05	93,04	93,45	93,78	92,93	
	MLP	94,17	94,69	94,58	95,65	95,88	93,16	
	BIN	–	95,17	92,73	93,21	94,45	93,24	
Cancer	KNN	94,51	95,69	94,76	94,66	95,73	95,27	4,5×
	LDA	93,99	93,66	93,34	93,46	93,89	93,64	
	MLP	94,28	95,46	94,55	94,98	95,18	95,01	
	BIN	–	93,69	93,63	93,22	94,42	93,72	
Glass	KNN	93,31	93,89	93,69	93,69	93,70	95,37	4,5×
	LDA	94,56	94,57	93,74	93,73	93,74	95,13	
	MLP	94,42	94,35	93,97	93,83	93,81	95,08	
	BIN	–	94,59	95,24	95,25	95,25	94,66	
Ovarian	KNN	91,85	95,66	94,93	93,46	95,18	95,66	50×
	LDA	71,40	95,42	96,14	94,11	94,50	95,10	
	MLP	86,98	93,79	96,20	92,15	93,51	95,68	
	BIN	–	94,64	95,51	94,90	94,23	95,15	
Instagram	KNN	77,58	79,23	76,95	74,12	77,01	77,15	5,5×
	LDA	77,41	77,52	79,18	79,84	79,55	80,42	
	MLP	74,16	73,19	71,48	73,27	74,81	72,29	
	BIN	–	60,93	54,41	58,18	52,38	59,48	

Fonte: Elaborado pelo autor.

A coluna original mostra a linha de base sem redução de dimensionalidade; A coluna L-TSDR mostra o F1 Score após a aplicação do TSDR com a variação linear conforme conforme Equação 3.7; As demais colunas mostram a F1 Score após a aplicação do TSDR com cada transformação, conforme Equação 3.12 à Equação 3.15; Δ é o fator de redução de dimensionalidade; BIN refere-se à função discriminante Equação 3.11; Valores em negrito indicam o melhor desempenho.

Os resultados do Estágio 1 mostram melhorias na separação probabilística em todos os conjuntos de dados, com o `ovarian_dataset` exibindo a transformação mais positiva, onde o Coeficiente de *Silhouette* aumenta de 0,46 para 0,58 e a Correlação de Pearson melhora de 0,17 para 0,58 após reduzir a dimensionalidade de 100 para 29 características.

Esta melhoria demonstra que a otimização de separação probabilística identifica com sucesso e preserva as direções de características mais discriminativas enquanto elimina dimensões de ruído que contribuem para o mascaramento entre classes. O fato de que esta melhoria ocorre simultaneamente com uma redução significativa de dimensionalidade valida a eficácia da abordagem de otimização baseada em informação mútua.

A otimização não linear do Estágio 2 aprimora ainda mais estas métricas em todos os conjuntos de dados, com o `dataset` Instagram, apesar de suas características desafiadoras, mostrando melhoria consistente onde o Coeficiente de *Silhouette* progride de 0,08 através de 0,11 para 0,14, enquanto a Correlação de Pearson aumenta de 0,08 através de 0,19 para 0,29.

Embora estes valores absolutos permaneçam relativamente modestos, eles representam melhorias significativas em um domínio onde padrões de engajamento são inerentemente ruidosos e exibem dependências complexas. A progressão consistente através dos estágios valida a abordagem hierárquica proposta e demonstra que mesmo em cenários altamente desafiadores, o método consegue extrair estruturas discriminativas latentes.

Em termos de correlação de características, a [Figura 5.2](#) revela o aprimoramento progressivo das relações característica-classe, com melhorias particularmente dramáticas para o *dataset* Ovarian onde a correlação progride de 0,17 para 0,75, demonstrando como o método transforma características de alta dimensionalidade fracamente correlacionadas em representações reduzidas fortemente preditivas.

Tabela 5.3 – Avaliação da redução de dimensionalidade e métricas de qualidade

<i>Dataset</i>	Dataset Original			Após Etapa 1 ¹			Após Etapa 2 ²		
	p^3	s_i	ρ	d^4	s_i	ρ	c^5	s_i	ρ
Crab	6	0,01	0,60	4	0,43	0,62	2	0,72	0,66
Cancer	9	0,71	0,65	5	0,75	0,71	2	0,85	0,87
Glass	9	0,55	0,03	3	0,63	0,25	2	0,77	0,55
Ovarian	100	0,46	0,17	29	0,58	0,58	2	0,70	0,75
Instagram	11	0,08	0,08	8	0,11	0,19	2	0,14	0,29

Fonte: Elaborado pelo autor.

¹Considerando o melhor valor de d para cada conjunto de dados.

²Considerando a melhor transformação não linear para cada conjunto de dados.

³Quantidade de dimensões no conjunto de dados original.

⁴Quantidade de dimensões ao fim da primeira etapa.

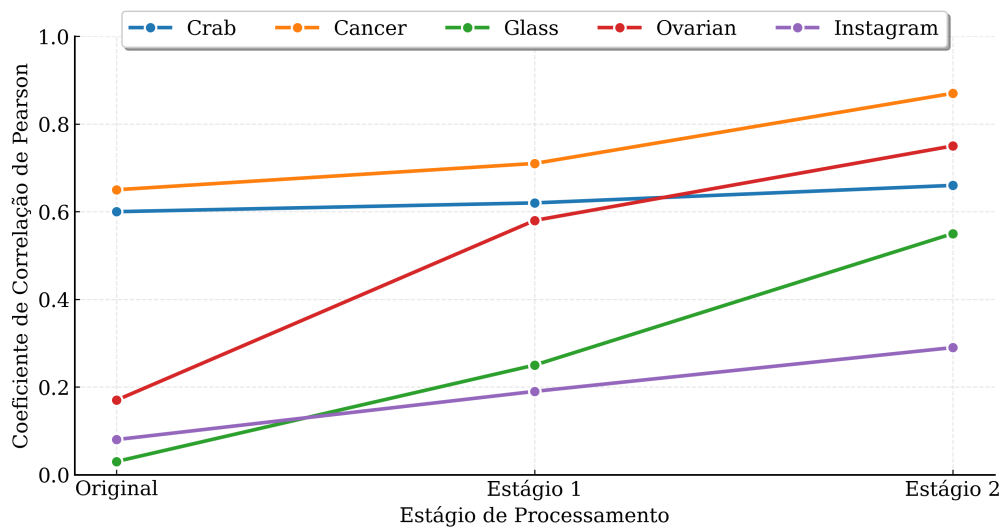
⁵Quantidade de dimensões ao fim da segunda etapa.

O `glass_dataset` mostra padrões similares, progredindo de correlação fraca em 0,03 para correlação moderada em 0,55, explicando as melhorias substanciais de classificação observadas. Esta transformação de relacionamentos fracos em associações fortes sugere que o método NL-TSDR é particularmente efetivo em revelar padrões latentes que não são evidentes no espaço de características original.

Além da análise das métricas de desempenho de classificação, avalia-se também a eficiência computacional do método NL-TSDR. A [Tabela 5.4](#) apresenta os tempos médios de execução em segundos para uma iteração de cada combinação de conjunto de dados, método de classificação e transformação. Como o método proposto envolve a solução de dois problemas de otimização, espera-se tempos de treinamento superiores quando aplicado. Contudo, os tempos médios de inferência permanecem similares para os classificadores, sugerindo que a relação custo-benefício pode ser favorável quando se considera que o maior investimento computacional no treinamento pode ser compensado pelos ganhos observados nas métricas de desempenho.

Figura 5.2 – Progressão do coeficiente de correlação de Pearson

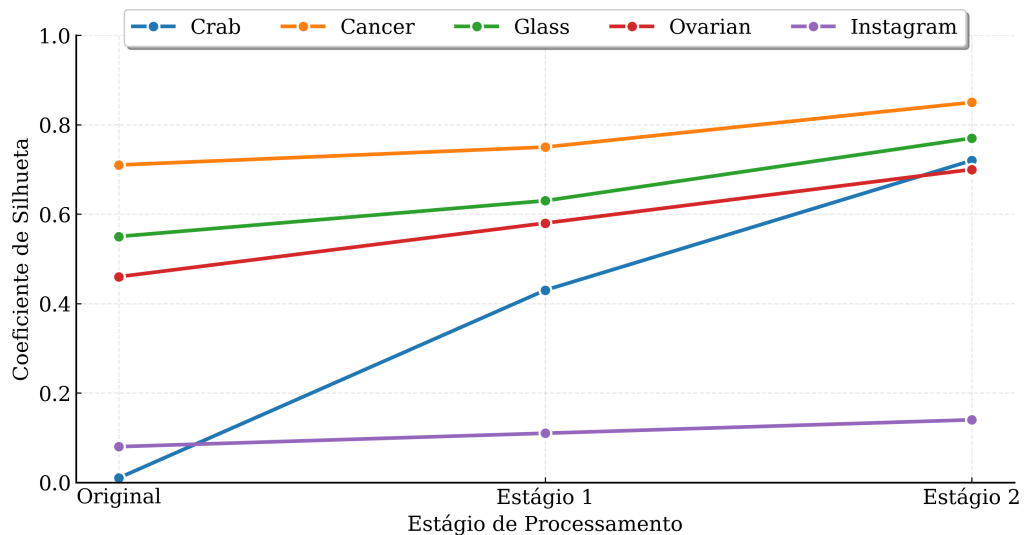
Progressão do Coeficiente de Correlação de Pearson através dos Estágios do NL-TSDR



Fonte: Elaborado pelo autor.

Figura 5.3 – Progressão do coeficiente de *Silhouette*

Progressão do Coeficiente de Silhueta através dos Estágios do NL-TSDR



Fonte: Elaborado pelo autor.

Tabela 5.4 – Tempo de execução (segundos) em conjuntos de dados de referência

Dataset		Original	L-TSDR	NL-TSDR				Δ
				Tanh	Gaussiana	Sigmoide	Laplaciana	
Crab	KNN	0,01	0,21	0,20	0,22	0,21	0,19	3×
	LDA	0,01	0,21	0,20	0,22	0,21	0,19	
	MLP	0,10	0,29	0,28	0,30	0,28	0,27	
	BIN ¹	–	0,21	0,19	0,21	0,20	0,19	
Cancer	KNN	0,01	0,25	0,32	0,44	0,33	0,37	4,5×
	LDA	0,01	0,25	0,32	0,44	0,33	0,37	
	MLP	0,09	0,32	0,39	0,52	0,41	0,45	
	BIN ¹	–	0,24	0,31	0,44	0,33	0,37	
Glass	KNN	0,01	0,62	0,68	0,70	0,68	0,61	4,5×
	LDA	0,01	0,62	0,68	0,70	0,67	0,61	
	MLP	0,09	0,70	0,76	0,78	0,75	0,69	
	BIN ¹	–	0,62	0,67	0,69	0,67	0,61	
Ovarian	KNN	0,03	28,74	28,82	29,15	28,78	28,78	50×
	LDA	0,01	28,74	28,82	29,15	28,78	28,78	
	MLP	10,28	28,82	28,90	29,23	28,86	28,87	
	BIN ¹	–	28,73	28,82	29,14	28,77	28,78	
Instagram	KNN	0,01	1,89	2,21	2,55	2,15	2,08	5,5×
	LDA	0,01	1,89	2,21	2,55	2,15	2,08	
	MLP	0,12	2,00	2,32	2,66	2,25	2,18	
	BIN ¹	–	1,88	2,21	2,54	2,15	2,07	

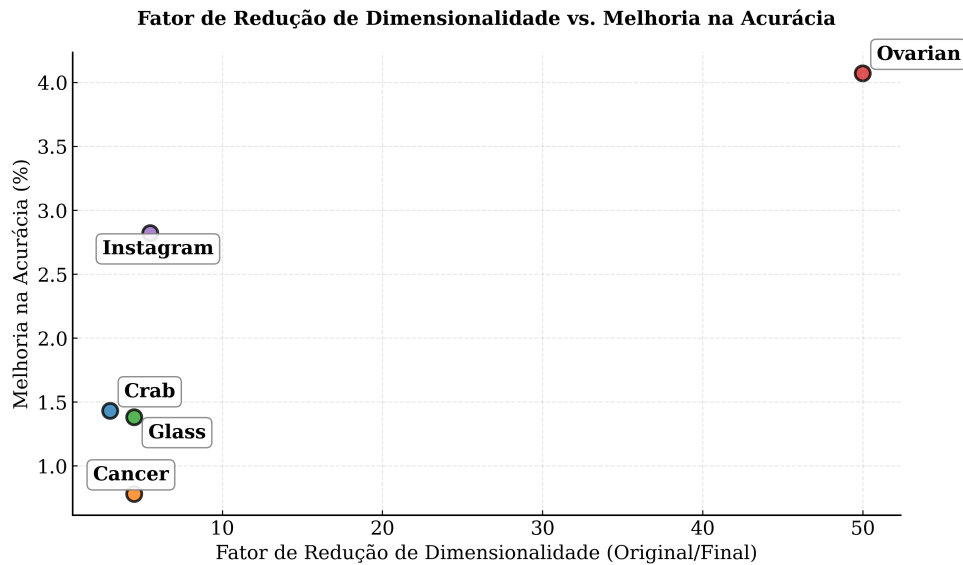
Fonte: Elaborado pelo autor.

A coluna original mostra a linha de base sem redução de dimensionalidade; A coluna L-TSDR mostra o tempo com a aplicação do TSDR com a variação linear conforme conforme [Equação 3.7](#); As demais colunas mostram o tempo com a aplicação do TSDR com cada transformação, conforme [Equação 3.12](#) à [Equação 3.15](#); Δ é o fator de redução de dimensionalidade; BIN refere-se à função discriminante [Equação 3.11](#).

Na [Figura 5.3](#) ilustra-se a evolução da estrutura de *clusters* ao longo do processamento, indicando que o método separa progressivamente classes sobrepostas, alcançando melhores valores no Estágio 2. A melhoria consistente nos Coeficientes de *Silhouette* demonstra que o método não apenas reduz dimensionalidade, mas efetivamente melhora a estrutura geométrica dos dados, criando representações onde classes naturalmente se agrupam em regiões mais compactas e bem separadas do espaço de características.

A relação entre fator de redução de dimensionalidade e melhoria de acurácia é ilustrada na [Figura 5.4](#). Conjuntos de dados de menor dimensionalidade alcançam fatores de redução menores mas ainda demonstram melhorias significativas, indicando que o método fornece valor mesmo quando restrições de dimensionalidade são menos severas. A correlação positiva consistente entre fator de redução e magnitude de melhoria sugere que os benefícios do NL-TSDR aumentam com a dimensionalidade do conjunto de dados, uma observação particularmente relevante para aplicações modernas de aprendizado de máquina, onde conjuntos de dados de alta dimensionalidade são cada vez mais comuns.

Figura 5.4 – Relação entre redução de dimensionalidade e melhoria de acurácia



Fonte: Elaborado pelo autor.

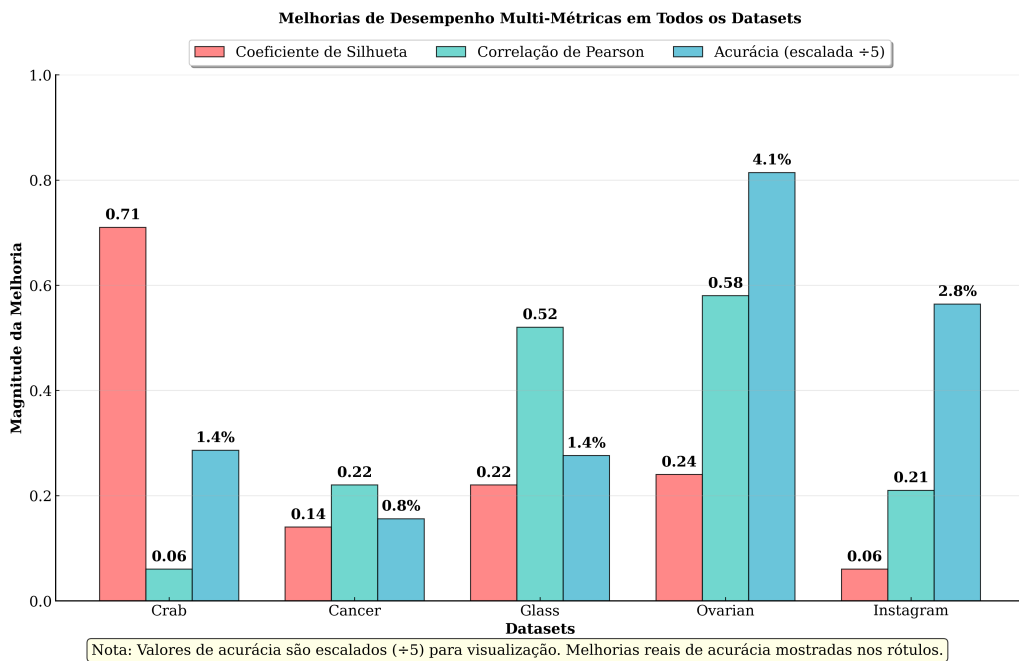
A comparação com a função discriminante incorporada definida na equação [Equação 3.11](#) revela que classificadores tradicionais geralmente superam o discriminante personalizado quando aplicados a espaços transformados por NL-TSDR. Este resultado é esperado e benéfico, pois demonstra que a transformação aprimora a qualidade do espaço de características suficientemente para que algoritmos de classificação padrão possam alcançar desempenho superior.

O desempenho consistentemente superior de classificadores tradicionais em dados transformados por NL-TSDR comparado aos espaços de características originais valida a hipótese central de que o método aprimora características intrínsecas dos dados em vez de favorecer paradigmas de classificação específicos.

Finalmente, a comparação multi-métrica na [Figura 5.5](#) demonstra que a abordagem proposta entrega melhorias simultâneas em três dimensões críticas de sistemas inteligentes de informação: acurácia de classificação, separabilidade entre classes e correlação característica-classe. A análise específica por transformação na [Figura 5.6](#), [Figura 5.7](#) e na [Figura 5.8](#) revela padrões de desempenho distintos que fornecem orientação prática para implementação do método.

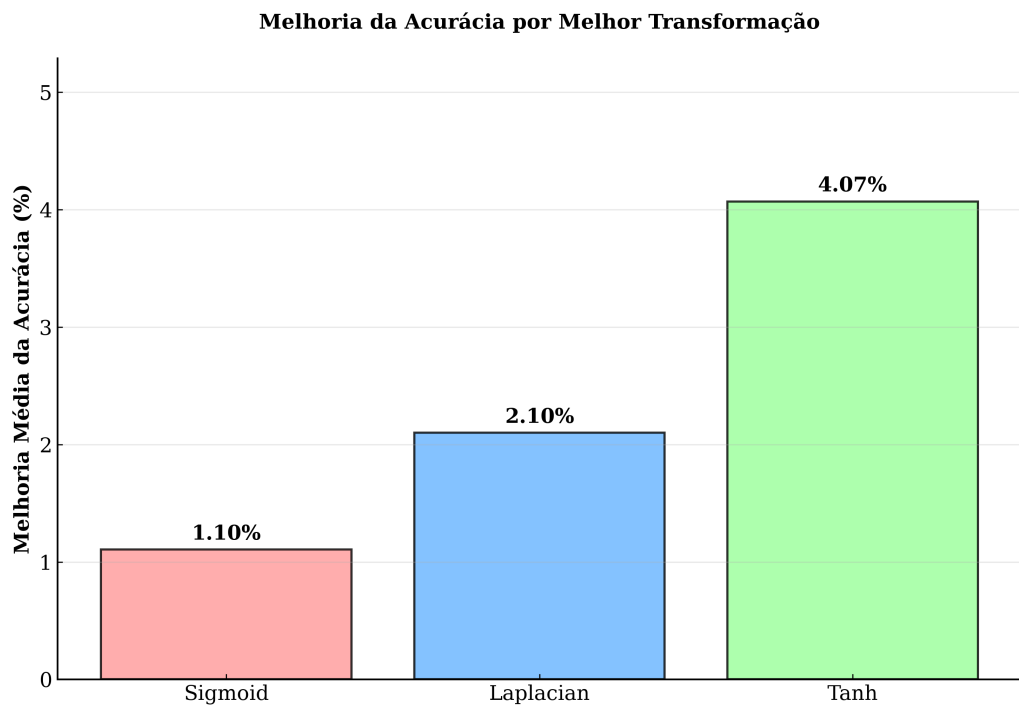
Estas visualizações coletivamente estabelecem que diferentes transformações não lineares exibem forças complementares, apoiando a estrutura adaptativa do método e demonstrando sua capacidade de abordar diversos desafios de classificação através de seleção direcionada de transformações baseada em características do conjunto de dados e objetivos de otimização.

Figura 5.5 – Melhorias no desempenho demonstrando melhoria simultânea

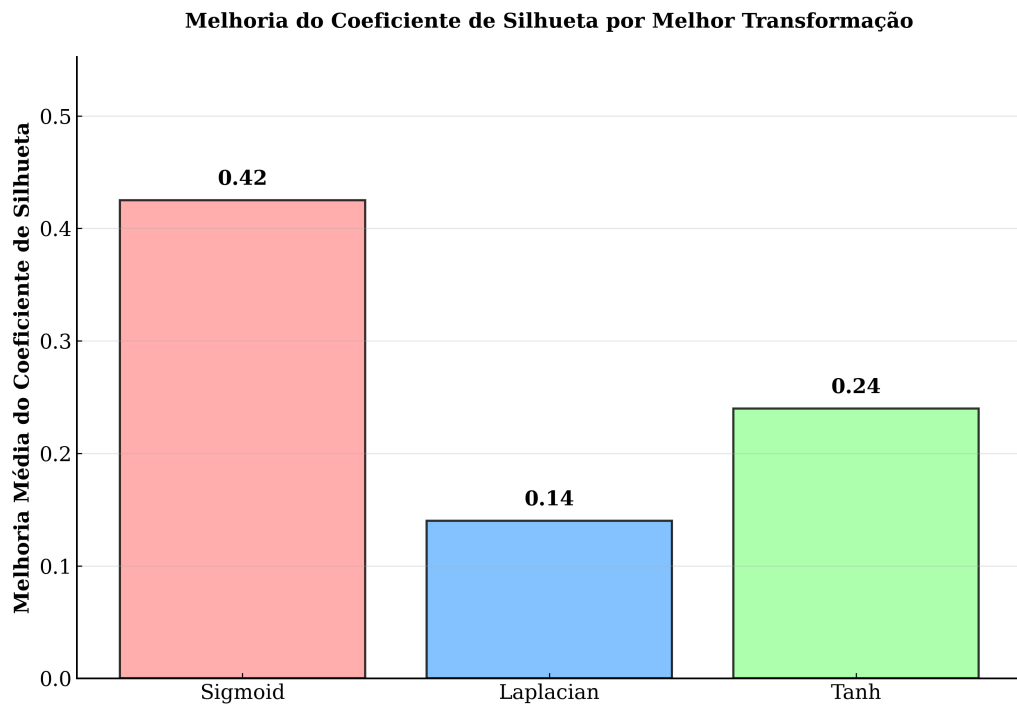


Fonte: Elaborado pelo autor.

Figura 5.6 – Comparação da acurácia da classificação entre transformações

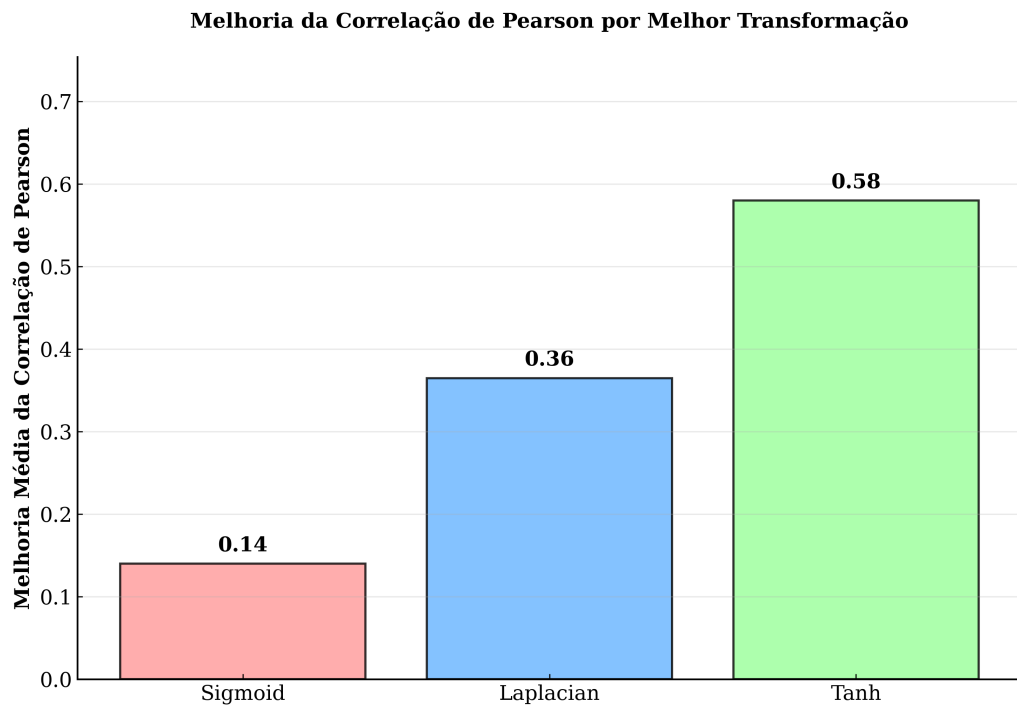


Fonte: Elaborado pelo autor.

Figura 5.7 – Melhorias no Coeficiente de *Silhouette* por transformação

Fonte: Elaborado pelo autor.

Figura 5.8 – Melhorias no Coeficiente de Correlação de Pearson por transformação



Fonte: Elaborado pelo autor.

Para avaliar a eficácia do método NL-TSDR, conduziu-se análise estatística comparando a acurácia de classificação dos conjuntos de dados originais (sem redução de dimensionalidade) com a acurácia de classificação dos conjuntos de dados resultantes do método proposto. A análise utilizou nível de significância $\alpha = 0,05$, que representa a probabilidade máxima aceitável de cometer erro ao concluir que existe diferença quando na verdade não existe. Este é o critério padrão adotado na literatura científica (WALPOLE et al., 2016).

Dentre as análises estatísticas empregadas nesta tese, utilizou-se a análise de variância, ou *analysis of variance* (ANOVA), técnica que permite comparar diferentes condições experimentais aplicadas aos mesmos sujeitos. O ANOVA pode oferecer uma abordagem estatisticamente mais robusta do que testes independentes ao considerar a estrutura de dependência dos dados. Diferentemente de uma comparação simples entre grupos independentes, o ANOVA modela explicitamente que cada combinação experimental (conjunto de dados avaliado com um classificador específico) é testada sob duas condições: utilizando os dados originais sem redução e utilizando os dados após aplicação das diferentes variações do método NL-TSDR. Esta característica permite controlar a variabilidade entre os sujeitos experimentais, aumentando o poder estatístico do teste ao focar na variação dentro dos sujeitos. Como teste paramétrico, ele exige que os resíduos sigam uma distribuição normal e que a condição de esfericidade seja satisfeita, mas compensa essas exigências com maior sensibilidade para detectar diferenças reais quando seus pressupostos são atendidos (FIELD, 2013).

O ANOVA verifica se a diferença média entre as acurácias de classificação das duas condições avaliadas (conjuntos de dados originais sem redução de dimensionalidade e conjuntos de dados resultantes do NL-TSDR) é estatisticamente significativa, onde a estatística F mede a magnitude desta diferença: valores elevados de F combinados com valores baixos de p indicam que os métodos potencialmente produzem resultados diferentes. Complementarmente, aplicaram-se dois testes adicionais para validar os resultados de forma mais específica e robusta. O teste t pareado é uma ferramenta paramétrica que compara diretamente as médias de dois grupos dependentes, exigindo normalidade das diferenças pareadas. Este teste produz uma estatística t que quantifica a distância entre a diferença observada e a ausência de diferença (zero), oferecendo uma análise mais simples e direta da comparação entre as duas condições. Já o teste de Wilcoxon é sua versão não paramétrica, que compara distribuições em vez de médias e é sugerido quando os dados violam a normalidade, mantendo a robustez da análise mesmo diante de *outliers* ou distribuições assimétricas. Este teste produz uma estatística z com interpretação similar. Em todos os testes, o valor p representa a probabilidade de observar uma diferença igual ou maior que a encontrada se não houvesse diferença real entre os métodos (MONTGOMERY; RUNGER, 2017).

A regra de decisão estatística é direta: se $p \leq 0,05$, conclui-se que existe diferença significativa entre os métodos; se $p > 0,05$, não há evidência estatística suficiente para afirmar diferença. Ao optar pelo ANOVA complementado pelos testes t pareado e Wilcoxon, obtém-se uma análise triangulada que combina o controle de variabilidade entre sujeitos do ANOVA com a simplicidade interpretativa do teste t e a robustez não-paramétrica do Wilcoxon, fornecendo evidências convergentes sobre o desempenho do método proposto. Os resultados apresentados na [Tabela 5.5](#) demonstram convergência entre os três testes aplicados: o ANOVA apresentou $F = 9,07$ com $p = 0,0034$, indicando diferença estatística significativa; o teste t pareado resultou em $t = 2,37$ com $p = 0,0199$; e o teste de Wilcoxon produziu $z = -2,43$ com $p = 0,0153$. A concordância entre estes testes sugere que o método NL-TSDR produz melhorias estatisticamente significativas na acurácia de classificação quando comparado aos conjuntos de dados originais sem redução de dimensionalidade.

Tabela 5.5 – Análise estatística do método NL-TSDR

Teste Estatístico	Estatística	Resultado
ANOVA	F -statistic	9,0711
	p -value	0,0034
Teste t Pareado	t -statistic	2,3668
	p -value	0,0199
Teste Wilcoxon	z -score	-2,4257
	p -value	0,0153

Fonte: Elaborado pelo autor.

Os resultados experimentais apresentados têm várias implicações para a área de redução de dimensionalidade e classificação de padrões. A progressão consistente de melhorias através dos estágios valida a estrutura hierárquica do método, sugerindo que a separação probabilística e otimização não linear abordam aspectos complementares da otimização de espaço de características. Os testes estatísticos por sua vez sugerem que as melhorias são estatisticamente significativas.

O sucesso em conjuntos de dados que variam de problemas clássicos de classificação a aplicações de mídias sociais do mundo real demonstra a generalidade e robustez da abordagem, enquanto a capacidade de melhorar acurácia enquanto reduz drasticamente a dimensionalidade indica que o método identifica efetivamente e preserva informações essenciais para discriminação entre classes. Particularmente notável é o desempenho em conjuntos de dados desbalanceados e de alta dimensionalidade, indicando que o método aborda limitações específicas de técnicas tradicionais de redução de dimensionalidade.

Estes resultados posicionam o método NL-TSDR como uma contribuição positiva para a literatura, oferecendo uma solução que combina eficiência computacional com melhoria de desempenho de classificação, estabelecendo uma potencial abordagem para o tratamento do problema da maldição da dimensionalidade em problemas de classificação.

6 Discussão

6.1 Discussão dos Resultados Experimentais e Limitações

Os resultados experimentais obtidos sugerem que o método NL-TSDR proposto aborda os desafios do mascaramento de classes e da maldição da dimensionalidade em aplicações de aprendizado de máquina como planejado. As melhorias observadas através de conjuntos de dados diversos e heterogêneos fornecem indicações sobre a efetividade do método e sua aplicabilidade em diferentes domínios. A consistência das melhorias de acurácia alcançadas em todos os conjuntos de dados avaliados corrobora a hipótese de que a combinação de otimização de separação probabilística com transformações não lineares pode criar espaços de características superiores para classificação.

Uma análise dos resultados sugere que a magnitude das melhorias obtidas correlaciona-se com as características iniciais dos conjuntos de dados, particularmente a dimensionalidade e o nível da sobreposição entre classes. Esta correlação sugere que o método NL-TSDR possui uma capacidade adaptativa inerente, ajustando-se automaticamente às complexidades específicas de cada conjunto de dados. O `ovarian_dataset` exemplifica este comportamento, apresentando uma melhoria de 8,25 pontos percentuais em acurácia acompanhada de uma redução de dimensionalidade de $50\times$, demonstrando como conjuntos de dados de alta dimensionalidade podem se beneficiar da abordagem proposta nesta tese.

Este padrão de melhoria é possivelmente relevante para aplicações práticas, onde a diversidade de domínios de aplicação requer métodos capazes de se adaptar a diferentes complexidades estruturais. A potencial capacidade do NL-TSDR de identificar e explorar as características discriminativas latentes em dados de alta dimensionalidade pode fornecer uma vantagem significativa sobre outros métodos, que podem falhar quando confrontados com a escassez relativa de amostras em relação ao número de características.

O aprimoramento progressivo observado através de ambos os estágios de processamento, conforme demonstrado na [Figura 5.1](#), revela a natureza complementar da arquitetura de dois estágios proposta. Esta sinergia entre os estágios não é meramente aditiva, mas potencialmente multiplicativa em seus efeitos. O Estágio 1, através de sua otimização probabilística, estabelece fundações discriminativas ao maximizar as probabilidades de separação pareada entre classes. O Estágio 2, por sua vez, adapta essas fundações através de transformações não lineares que capturam padrões complexos que poderiam permanecer ocultos.

Esta arquitetura hierárquica pode permitir que o método capture tanto relações lineares quanto não lineares nos dados, oferecendo uma flexibilidade que outros métodos de

redução de dimensionalidade tradicionais, tipicamente limitados a uma única perspectiva de transformação, podem não proporcionar. A progressão gradual da qualidade de separação entre classes através dos estágios também sugere a estabilidade numérica do método, que parece não apresentar oscilações e instabilidades potencialmente observadas em métodos que tentam otimizar múltiplos objetivos simultaneamente.

A progressão dos Coeficientes de *Silhouette* apresentada na [Tabela 5.3](#) fornece evidência quantitativa de que o método pode mitigar o problema de mascaramento de classes. A transformação dos dados do Instagram demonstra a resolução de fronteiras de classe sobrepostas que podem confundir abordagens de classificação tradicionais. Esta melhoria é particularmente relevante considerando a complexidade inerente da classificação de engajamento em mídias sociais.

O Coeficiente de *Silhouette*, como métrica de qualidade de *clustering*, oferece uma perspectiva complementar à acurácia de classificação, permitindo avaliar a estrutura geométrica dos dados transformados. A melhoria consistente nesta métrica através de diferentes conjuntos de dados sugere que o NL-TSDR não apenas melhora o desempenho de classificação, mas também aprimora a separabilidade estrutural dos dados. Esta característica é potencialmente relevante para aplicações onde a interpretabilidade e a compreensão da estrutura dos dados são tão importantes quanto o desempenho de classificação.

O desempenho das funções de transformação através dos conjuntos de dados revela percepções para aplicações práticas do método proposto. As transformações laplacianas parecem exceder em cenários desafiadores envolvendo desbalanceamento de classes e presença de *outliers*, alcançando desempenho ótimo tanto no conjunto de dados Instagram (81,30% de acurácia LDA) quanto no `glass_dataset` (92,96% de acurácia k -NN).

Esta potencial superioridade das transformações laplacianas em cenários com *outliers* pode ser atribuída à sua capacidade de suavizar gradualmente as diferenças extremas nos dados, criando transições mais suaves entre regiões de diferentes densidades de dados. O *kernel* laplaciano, com sua característica de decaimento exponencial, oferece uma robustez natural contra pontos extremos que poderiam distorcer significativamente transformações baseadas em distâncias Euclidianas ou outras métricas mais sensíveis a *outliers*.

Por outro lado, as transformações tangente hiperbólica demonstram bom desempenho para conjuntos de dados de alta dimensionalidade, alcançando 96,51% de acurácia MLP no conjunto de dados Ovarian. A compressão de valores enquanto preserva a ordenação relativa prova-se benéfica quando métodos lineares tradicionais podem falhar. A função tangente hiperbólica oferece uma compressão não linear que mapeia o domínio infinito dos números reais para o intervalo limitado $[-1, 1]$, proporcionando uma normalização natural que é potencialmente valiosa em dados de alta dimensionalidade onde diferentes características podem apresentar escalas diferentes.

Embora a transformação gaussiana não tenha apresentado o melhor valor de acurácia em nenhum dos conjuntos de dados, seu desempenho consistente sugere sua utilidade como uma escolha padrão quando características específicas dos dados são desconhecidas. A distribuição Gaussiana, com sua propriedade de máxima entropia, oferece uma transformação que minimiza suposições sobre a estrutura subjacente dos dados, tornando-a uma opção conservadora mas confiável. No entanto, o desempenho superior de transformações mais especializadas indica que o conhecimento de domínio sobre características de distribuição dos dados pode guiar a seleção de transformações para resultados ótimos, fornecendo aos praticantes diretrizes fundamentadas para customização do método.

Esta flexibilidade na seleção de transformações representa uma vantagem do NL-TSDR sobre métodos de transformação fixa, permitindo que diferentes domínios incorporem conhecimento específico da aplicação na configuração do método. A capacidade de adaptar a função de transformação às características dos dados oferece um caminho para otimização orientada por domínio que pode resultar em melhorias de desempenho em aplicações específicas.

Os efeitos da de redução de dimensionalidade alcançados (variando de $3\times$ a $50\times$) enquanto mantêm ou melhoram a acurácia de classificação abordam requisitos críticos de eficiência computacional para sistemas de aprendizado de máquina de grande escala. A arquitetura de dois estágios fornece características favoráveis de complexidade computacional: o Estágio 1 reduz espaços de características para tamanhos gerenciáveis que permitem otimização eficiente do Estágio 2, criando uma ferramenta escalável não apenas para redes sociais mas outras áreas do conhecimento.

Esta redução substancial de dimensionalidade tem implicações que se podem se estender além da eficiência computacional. A redução no número de características também diminui significativamente os requisitos de armazenamento, acelera operações de I/O, e reduz a complexidade de etapas de processamento subsequentes. Em ambientes de produção, onde um grande volume de amostras é processado diariamente, essas economias podem se acumular em benefícios operacionais substanciais.

Além disso, a redução de dimensionalidade pode facilitar a visualização e interpretação dos dados, aspectos desejados na validação e compreensão de modelos em aplicações práticas. Espaços de características de menor dimensionalidade permitem análises exploratórias mais efetivas e podem revelar detalhes sobre a estrutura dos dados que permaneceriam ocultos em espaços de alta dimensionalidade.

A relação entre o fator de redução de dimensionalidade e a melhoria de acurácia, ilustrada na [Figura 5.4](#), demonstra que os benefícios do NL-TSDR podem aumentar com a dimensionalidade do conjunto de dados. Esta característica torna o método potencialmente valioso para aplicações onde a dimensionalidade de características continua a crescer mais rapidamente que a disponibilidade de amostras.

Esta relação crescente entre dimensionalidade e benefício sugere que o NL-TSDR pode ser adequado para aplicações de *big data*, onde a proliferação de sensores, dispositivos IoT (*Internet of Things*) e plataformas digitais resulta em conjuntos de dados com milhares ou dezenas de milhares de características. O método pode oferecer uma solução escalável para o problema de extrair informação útil de dados de alta dimensionalidade sem sucumbir aos potenciais problemas da maldição da dimensionalidade.

O desempenho superior de classificadores tradicionais (k -NN, LDA, MLP) em espaços transformados pelo NL-TSDR comparada aos espaços de características originais sugere que a abordagem proposta aprimora características intrínsecas dos dados ao invés de favorecer paradigmas de classificação específicos. Esta característica distingue o NL-TSDR de métodos que podem alcançar melhorias apenas através de complexidade algorítmica ao invés de aprimoramento do espaço de características e melhores fronteiras de decisão.

Esta universalidade de melhoria entre diferentes paradigmas de classificação é potencialmente importante para aplicações práticas, onde a escolha do classificador pode ser determinada por fatores como interpretabilidade, tempo de treinamento ou requisitos específicos de desempenho. As melhorias nos F1 *Scores* acompanhando os ganhos de acurácia sugere ainda que o desempenho aprimorado não é alcançado às custas do balanceamento de classes. O F1 *Score*, como média harmônica de *Precision* e *Recall*, oferece uma perspectiva mais holística do desempenho do modelo, especialmente em cenários onde diferentes classes têm importâncias variadas ou custos de erro assimétricos.

A manutenção ou melhoria dos F1 *Scores* sugere que o NL-TSDR não introduz vieses que favoreceriam classes majoritárias às custas de classes minoritárias, uma preocupação em métodos de transformação de dados. Esta característica é potencialmente valiosa em aplicações como detecção de fraude, diagnóstico médico, ou análise de sentimentos, onde classes minoritárias frequentemente representam os casos de maior interesse prático.

Embora o NL-TSDR possa demonstrar desempenho de classificação superior, o método introduz sobrecarga computacional através de sua arquitetura de dois estágios. Esta sobrecarga computacional representa uma relação custo-benefício que deve ser cuidadosamente avaliada no contexto de cada aplicação específica. A complexidade das operações matriciais, pode se tornar proibitiva para conjuntos de dados extremamente grandes ou aplicações com restrições rigorosas de tempo. No entanto, é importante considerar que este custo computacional é tipicamente incorrido apenas durante a fase de treinamento ou preparação dos dados.

O Estágio 2 apresenta demandas computacionais adicionais através de sua otimização multiobjetivo, particularmente ao avaliar múltiplas funções de transformação para identificar configurações ótimas. A abordagem de busca em grade (*grid search*) para otimização de parâmetros pode tornar-se cara para espaços de parâmetros de alta dimensionalidade, potencialmente requerendo muitas avaliações para alcançar desempenho

satisfatório. Esta intensidade computacional cria uma relação custo-benefício: acurácia de classificação aprimorada e capacidades de redução de dimensionalidade, resultando em inferência mais rápida, *versus* tempos aumentados de pré-processamento e treinamento que podem limitar certas aplicações.

Contudo, esta relação custo benefício deve ser avaliada dentro do contexto de eficiência geral do sistema. A redução substancial de dimensionalidade alcançada (compressão de $3\times$ a $50\times$) reduz significativamente os requisitos computacionais para operações de classificação subsequentes, demandas de armazenamento, etc. Para cenários de processamento em lote, comuns em análises de dados em grande escala, onde previsões são computadas periodicamente ao invés de em tempo real, o custo inicial de otimização é potencialmente amortizado através de operações de classificação no espaço de características reduzido.

Adicionalmente, as transformações otimizadas podem ser pré-computadas e reutilizadas através de múltiplos conjuntos de dados com características similares, reduzindo o fardo computacional prático. Esta reutilização é particularmente valiosa em ambientes de produção onde múltiplos modelos podem beneficiar-se das mesmas transformações otimizadas, distribuindo o custo computacional inicial por várias aplicações.

Apesar da efetividade demonstrada, várias limitações merecem reconhecimento e sugerem direções de pesquisa futura. A abordagem atual de otimização de parâmetros depende da busca em grade dentro de faixas teoricamente motivadas, que pode não capturar combinações ótimas de parâmetros para todas as características possíveis dos dados. Mecanismos adaptativos de seleção de parâmetros que automaticamente ajustem parâmetros de transformação baseados em propriedades do conjunto de dados poderiam aprimorar a generalizabilidade do método enquanto reduzem a sobrecarga computacional.

O desenvolvimento de estratégias de otimização de parâmetros mais sofisticadas representa uma oportunidade significativa para pesquisa futura. Técnicas como algoritmos evolutivos poderiam potencialmente identificar configurações de parâmetros superiores com menor custo computacional comparado à busca em grade exaustiva. Estas abordagens poderiam também incorporar conhecimento prévio sobre as características dos dados ou desempenho histórico em conjuntos de dados similares.

A integração com arquiteturas de aprendizado profundo apresenta oportunidades para aprimorar as capacidades do método. Tal integração poderia potencialmente abordar as limitações de sobrecarga computacional enquanto mantém as vantagens discriminativas do método. Outras direções de pesquisa promissoras incluem o desenvolvimento de métricas de qualidade de dados que possam prever a adequação do NL-TSDR para conjuntos de dados específicos e a exploração de funções de transformação adaptativas que se ajustem dinamicamente durante o treinamento. A aplicação do método a domínios emergentes como processamento de linguagem natural, visão computacional e análise de grafos também oferece oportunidades para validação adicional e refinamento da abordagem.

6.2 Direções Futuras

Ao se abordar técnicas de redução de dimensionalidade e classificação, observa-se uma correspondência frequente com o uso de sistemas de recomendação (PARK et al., 2012). No contexto de redes sociais, essa associação se manifesta de forma recorrente. Utilizam-se tais sistemas em redes sociais, plataformas de *streaming* e demais aplicações que dependem da análise de comportamento para realizar sugestões aos usuários.

Sistemas de recomendação combinam técnicas computacionais para selecionar itens personalizados com base nos interesses dos usuários e no contexto em que se inserem. Esses itens podem assumir diferentes formas, como livros, filmes, notícias, músicas, vídeos, anúncios, links patrocinados, páginas de internet e produtos de lojas virtuais. Empresas como Amazon, Netflix e Google destacam-se pelo uso intensivo de sistemas de recomendação.

A Netflix, por exemplo, constitui objeto de estudo tanto no contexto empresarial quanto acadêmico. Originalmente, a empresa permitia que os usuários criassem listas de filmes para assistir posteriormente. Quando dispusessem de tempo, a Netflix enviaria lembretes sobre esses filmes. Entretanto, identificou-se um padrão: os usuários preenchiam suas filas com numerosos filmes, mas raramente os acessavam quando notificados.

Verificou-se que, ao serem questionados sobre suas intenções futuras de visualização, os usuários selecionavam filmes como documentários sobre guerras ou produções estrangeiras. Dias depois, contudo, tendiam a consumir comédias ou filmes de romance. Observou-se, assim, uma discrepância entre as intenções declaradas e o comportamento.

Diante dessa disparidade, a Netflix deixou de solicitar declarações sobre preferências futuras e passou a construir modelos baseados em cliques e visualizações de usuários com perfis semelhantes. A empresa começou a apresentar listas de filmes baseadas não nas preferências declaradas, mas nos padrões indicados pelos dados. Como resultado, registrou-se aumento na frequência de visitas e no volume de filmes assistidos. Nas palavras de Xavier Amatriain (STEPHENS-DAVIDOWITZ, 2017), ex-cientista de dados da empresa, “os algoritmos conhecem você melhor do que você mesmo”.

Em 2000, introduziram-se recomendações personalizadas de filmes na Netflix e, em 2006, lançou-se o Netflix Prize, uma competição de aprendizado de máquina e mineração de dados. Na época, a empresa utilizava seu sistema de recomendação proprietário, que apresentava uma raiz do erro quadrático médio (RMSE ou *root mean squared error*) de 0,9525, e desafiou a comunidade a superar esse resultado de referência em 10%. A equipe que atingisse a meta ou dela se aproximasse após um ano receberia o prêmio em dinheiro.

O vencedor do prêmio em 2007 alcançou um RMSE de 0,88. A Netflix implementou esses algoritmos em produção após adaptações. Algumas equipes alcançaram um RMSE de 0,8567 em 2009, porém a empresa não implementou esses algoritmos devido ao esforço de engenharia necessário para obter o aumento marginal de acurácia. Isso ilustra um aspecto

relevante nos sistemas de recomendação: requer-se uma relação positiva entre as melhorias do modelo e os esforços para implementá-lo (AMATRIAIN; BASILICO, 2015).

No contexto das redes sociais, observam-se dinâmicas semelhantes. O Instagram introduziu em 2019 um recurso denominado “explorar”, no qual se apresentam recomendações de conteúdo baseadas no comportamento dos usuários. Segundo dados da própria empresa, mais da metade da comunidade visita mensalmente o Instagram Explore para descobrir fotos, vídeos e histórias relacionadas aos seus interesses. Recomendar o conteúdo mais relevante entre bilhões de opções em tempo real apresenta desafios de aprendizado de máquina que exigem soluções de engenharia.

Para enfrentar esses desafios, a empresa desenvolveu linguagens de consulta personalizadas, técnicas de modelagem e ferramentas que permitem experimentação em alta velocidade. Essas soluções representam um sistema de IA baseado em um funil de classificação de três partes que extrai 65 bilhões de recursos e realiza 90 milhões de previsões de modelos a cada segundo (Meta, 2023).

O modelo utiliza sistemas de recomendação combinados a redes neurais artificiais e ao k -NN. Essas informações corroboram a validade da hipótese do uso combinado do NL-TSDR com sistemas de recomendação. Assim como o NL-TSDR foi utilizado para classificar se uma publicação apresenta engajamento satisfatório, é provável que esse método apresente resultados quando atue como pré-processador para um sistema de recomendação que, seja por meio do conteúdo ou da filtragem colaborativa, indique a receptividade das publicações pela audiência.

Sistemas de recomendação representam, para alguns, a distinção algorítmica mais significativa entre empresas nascidas digitais na era contemporânea e empresas anteriores. Seu uso é extenso e contribui para ciclos de negócios: quanto mais usuários absorvem dos sistemas de recomendação de uma empresa, mais valiosas se tornam suas bases de dados e, quanto mais valiosas se tornam, mais usuários atraem.

Convém considerar os sistemas de recomendação não apenas como ferramenta de vendas online, mas como instrumento de melhoria contínua para empresas e clientes. No caso da Netflix, 80% do que os usuários assistem na plataforma tem origem em sistemas de recomendação (GOMEZ-URIBE; HUNT, 2016). Na Amazon, 35% das compras dos clientes originam-se de recomendações de produtos. No Airbnb, 99% das reservas têm origem no ranking de acomodações e nas sugestões de acomodações semelhantes.

Dessa maneira, a capacidade potencial do NL-TSDR de aprimorar o desempenho de múltiplos paradigmas de classificação, aliada à redução substancial de dimensionalidade, sugere que o método pode atuar como pré-processador em sistemas de recomendação. Considerando-se que empresas como Netflix, Instagram e Amazon dependem de processamento eficiente de dados de alta dimensionalidade para gerar recomendações em tempo

real, a aplicação do NL-TSDR neste contexto pode contribuir tanto para a acurácia das predições quanto para a eficiência computacional dos sistemas. Esta convergência entre as capacidades do método proposto e as demandas dos sistemas de recomendação estabelece uma base para investigações futuras sobre sua integração em plataformas de larga escala.

Conclusão

Nesta pesquisa, desenvolveu-se um método de Redução de Dimensionalidade Não Linear de Dois Estágios (NL-TSDR) para abordar os desafios de mascaramento de classes e maldição da dimensionalidade em aplicações de aprendizado de máquina. Integrou-se otimização de separação probabilística e transformações não lineares adaptativas em uma arquitetura que busca aprimorar a qualidade do espaço de características.

A arquitetura de dois estágios proposta situa-se entre abordagens lineares tradicionais e métodos não lineares de redução de dimensionalidade. Buscou-se estabelecer um compromisso entre complexidade computacional e eficiência, mantendo-se a tratabilidade necessária para sistemas reais.

Na avaliação experimental através de conjuntos de dados diversos, observaram-se melhorias na acurácia de classificação, com ganhos em cenários de alta dimensionalidade. A consistência dos resultados em diferentes tipos de dados sugere que o método captura aspectos da estrutura de dados que se manifestam em diversos domínios.

A análise dos resultados indica que as melhorias ocorrem com maior intensidade em conjuntos de dados de alta dimensionalidade com classes sobrepostas e estruturas complexas. Esta característica sugere que o método identifica e explora oportunidades de transformação onde as características dos dados assim o permitem.

Alcançaram-se fatores de redução de dimensionalidade entre $3\times$ e $50\times$ com manutenção ou melhoria do desempenho de classificação. Esta capacidade de redução relaciona-se a requisitos de eficiência para sistemas que processam grandes conjuntos de dados, com possíveis implicações para armazenamento, transmissão de dados e análises em tempo real.

A redução de dimensionalidade obtida pelo NL-TSDR pode facilitar a aplicação de técnicas analíticas que apresentam limitações em espaços de alta dimensionalidade. Métodos de visualização, análise de *clusters* e técnicas de interpretabilidade tornam-se aplicáveis quando se reduz a dimensionalidade para intervalos gerenciáveis. Esta facilitação de análises subsequentes amplia as possibilidades de aplicação do método.

Adicionalmente, a redução de dimensionalidade pode contribuir para a mitigação de problemas de *overfitting* em modelos subsequentes, aspecto relevante em cenários onde o número de características se aproxima ou excede o número de amostras disponíveis.

A transformação de fronteiras de classe sobrepostas, quantificada através de melhorias nos Coeficientes de *Silhouette* e análise de correlação, indica a reorganização do espaço de características. Esta reorganização pode expor padrões discriminativos que se apresentavam ocultos no espaço original.

A evolução dos Coeficientes de *Silhouette* em diversos conjuntos de dados sugere que o método atua em situações onde classes apresentavam baixa separabilidade no espaço original. Esta capacidade de reorganização de estruturas complexas apresenta implicações para aplicações onde a separabilidade natural das classes é limitada. Observou-se melhoria simultânea em múltiplas métricas de qualidade (acurácia, *F1 Score*, entre outras), indicando que os ganhos não se concentram em aspectos específicos do desempenho. Esta característica sugere que o NL-TSDR atua sobre aspectos estruturais dos dados.

O aprimoramento progressivo através de ambos os estágios de processamento corrobora as fundações teóricas e demonstra aplicabilidade em diversos domínios. A natureza complementar dos dois estágios cria um efeito onde o resultado combinado pode exceder os benefícios que seriam alcançados por cada estágio individualmente.

A validação através de diferentes paradigmas de classificação, desde métodos como *k*-NN e LDA até abordagens como MLP, indica que as melhorias se manifestam independentemente do algoritmo específico. Esta independência pode ser relevante para aplicações práticas, onde a escolha do classificador é determinada por considerações que incluem interpretabilidade, tempo de treinamento e requisitos de produção.

Na aplicação à classificação de engajamento em mídias sociais, a transformação laplaciana alcançou uma melhoria de 3,04% sobre o método de referência. Estes resultados indicam o potencial de transformações não lineares em sistemas especialistas, particularmente para análise de mídias sociais.

A efetividade da transformação laplaciana em dados de mídias sociais pode relacionar-se à sua capacidade de lidar com a natureza ruidosa e heterogênea destes domínios. As características de decaimento exponencial do *kernel* laplaciano proporcionam robustez contra *outliers* e variações extremas comuns em dados de engajamento.

Esta aplicação em mídias sociais demonstra a aplicabilidade do método em contextos contemporâneos. O volume e velocidade de geração de dados em plataformas digitais criam desafios que métodos tradicionais de redução de dimensionalidade podem não abordar adequadamente. A capacidade do NL-TSDR de processar estes dados enquanto extrai informações discriminativas representa uma possível contribuição para o domínio.

As direções de pesquisa futura incluem a investigação de mecanismos de seleção adaptativa de parâmetros, exploração de integração com arquiteturas de aprendizado profundo e sistemas de recomendação. As fundações probabilísticas estabelecidas neste trabalho fornecem base para estas extensões. O desenvolvimento de mecanismos adaptativos de seleção de parâmetros representa uma oportunidade para pesquisa futura. A implementação atual requer conhecimento de domínio para seleção de parâmetros de transformação. Sistemas adaptativos que identifiquem automaticamente configurações baseadas em características estatísticas dos dados podem simplificar o uso do método.

A aplicação do NL-TSDR em conjuntos de dados de referência e em aplicações reais de mídias sociais estabelece seu potencial como ferramenta para abordar desafios de redução de dimensionalidade em pesquisa de aprendizado de máquina. Esta validação através de múltiplos domínios e tipos de dados corrobora a capacidade de generalização do método.

À medida que a dimensionalidade de dados continua a crescer em diversos domínios, métodos que podem extrair informação discriminativa enquanto mantêm tratabilidade computacional podem se tornar relevantes para sistemas de informação. Esta tendência se observa em áreas como análise de texto de alta dimensionalidade, visão computacional de alta resolução e análise de redes sociais.

A crescente demanda por interpretabilidade e explicabilidade em sistemas de IA relaciona-se à redução de dimensionalidade. Modelos operando em espaços de características reduzidos apresentam maior interpretabilidade, e as transformações aplicadas pelo NL-TSDR podem revelar estruturas de dados que contribuem para a compreensão dos domínios subjacentes.

A eficiência computacional do método posiciona-o em contextos de *edge computing* e processamento distribuído, onde recursos computacionais podem ser limitados. A capacidade de reduzir dimensionalidade enquanto preserva informação discriminativa permite o uso de métodos de aprendizado de máquina em ambientes com restrições de recursos, expandindo aplicações para dispositivos móveis, sensores IoT e sistemas embarcados.

Em conclusão, o método NL-TSDR apresentado nesta pesquisa constitui uma contribuição para o campo de redução de dimensionalidade e classificação. Suas fundações teóricas, desempenho experimental e aplicabilidade prática estabelecem-no como potencial ferramenta para pesquisadores que enfrentam os desafios de análise de dados de alta dimensionalidade em sistemas contemporâneos. A combinação de fundamentação científica, desenvolvimento metodológico e aplicabilidade prática posiciona este trabalho como contribuição para o desenvolvimento de técnicas de análise de dados e suas aplicações em diversos domínios.

Referências

ABDALLA, H. I.; AMER, A. A. Towards highly-efficient k-nearest neighbor algorithm for big data classification. In: *2022 5th International Conference on Networking, Information Systems and Security: Envisage Intelligent Systems in 5g//6G-based Interconnected Digital Worlds (NISS)*. [S.l.: s.n.], 2022. p. 1–5. Citado na página 52.

ABOU-MOUSTAFA, K. T.; De La Torre, F.; FERRIE, F. P. Pareto models for discriminative multiclass linear dimensionality reduction. *Pattern Recognition*, v. 48, n. 5, p. 1863–1877, 2015. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320314004580>>. Citado na página 15.

ADAMIC, L.; BUYUKKOKTEN, O.; ADAR, E. A social network caught in the web. *First Monday*, v. 8, n. 6, Jun. 2003. Disponível em: <<https://firstmonday.org/ojs/index.php/fm/article/view/1057>>. Citado na página 25.

Agência Senado. *Senado aprova regulamentação da inteligência artificial; texto vai à Câmara*. 2024. Disponível em: <<https://www12.senado.leg.br/noticias/materias/2024/12/10/senado-aprova-regulamentacao-da-inteligencia-artificial-texto-vai-a-camara>>. Citado 3 vezes nas páginas 31, 32 e 34.

Akhter, J. et al. Common pitfalls to avoid while using multiobjective optimization in machine learning. *arXiv e-prints*, p. arXiv:2405.01480, maio 2024. Citado na página 46.

ALJUNID, M. F. et al. A collaborative filtering recommender systems: Survey. *Neurocomputing*, v. 617, p. 128718, 2025. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231224014899>>. Citado na página 22.

ALPAYDIN, E. *Machine Learning*. The MIT Press, 2021. ISBN 9780262365369. Disponível em: <<https://doi.org/10.7551/mitpress/13811.001.0001>>. Citado 2 vezes nas páginas 50 e 51.

AMATRIAIN, X.; BASILICO, J. Recommender systems in industry: A netflix case study. In: _____. *Recommender Systems Handbook*. Boston, MA: Springer US, 2015. p. 385–419. Disponível em: <https://doi.org/10.1007/978-1-4899-7637-6_11>. Citado na página 129.

ANPD. *ANPD determina suspensão cautelar do tratamento de dados pessoais para treinamento da IA da Meta*. 2024. Disponível em: <<https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta>>. Citado na página 31.

ANPD. *Meta cumpre exigências da ANPD e poderá retomar, com restrições, o uso de dados pessoais para treinamento de inteligência artificial*. 2024. Disponível em: <<https://www.gov.br/anpd/pt-br/assuntos/noticias/meta-cumprer-exigencias-da-anpd-e-podera-retomar-com-restricoes-o-uso-de-dados-pessoais-para-treinamento-de-inteligencia-artificial>>. Citado na página 31.

- ARAMBURU, M. J.; BERLANGA, R.; LANZA-CRUZ, I. A data quality multidimensional model for social media analysis. *Business & Information Systems Engineering*, v. 66, n. 6, p. 667–689, dec 2024. ISSN 1867-0202. Disponível em: <<https://doi.org/10.1007/s12599-023-00840-9>>. Citado na página 16.
- AWASTHI, P.; HAGHTALAB, N.; ZHAO, E. Open problem: The sample complexity of multi-distribution learning for vc classes. In: NEU, G.; ROSASCO, L. (Ed.). *Proceedings of Thirty Sixth Conference on Learning Theory*. PMLR, 2023. (Proceedings of Machine Learning Research, v. 195), p. 5943–5949. Disponível em: <<https://proceedings.mlr.press/v195/awasthi23a.html>>. Citado na página 23.
- AZEVEDO, B. F.; ROCHA, A. M. A. C.; PEREIRA, A. I. Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Machine Learning*, v. 113, n. 7, p. 4055–4097, 7 2024. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1007/s10994-023-06467-x>>. Citado 2 vezes nas páginas 48 e 49.
- BAGIROV, A. M.; ALIGULIYEV, R. M.; SULTANOVA, N. Finding compact and well-separated clusters: Clustering using silhouette coefficients. *Pattern Recognition*, v. 135, p. 109144, 2023. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320322006239>>. Citado na página 36.
- BARABÁSI, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature*, v. 435, n. 7039, p. 207–211, 2005. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/nature03459>>. Citado na página 25.
- BARBIERI, M. C.; GRISCI, B. I.; DORN, M. Analysis and comparison of feature selection methods towards performance and stability. *Expert Systems with Applications*, v. 249, p. 123667, 2024. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417424005335>>. Citado na página 19.
- BENESTY, J. et al. Pearson correlation coefficient. In: _____. *Noise Reduction in Speech Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 1–4. ISBN 978-3-642-00296-0. Disponível em: <https://doi.org/10.1007/978-3-642-00296-0_5>. Citado na página 62.
- BERISHA, V. et al. Digital medicine and the curse of dimensionality. *npj Digital Medicine*, v. 4, n. 1, p. 153, 10 2021. ISSN 2398-6352. Disponível em: <<https://doi.org/10.1038/s41746-021-00521-5>>. Citado na página 15.
- BEYER, K. et al. When is “nearest neighbor” meaningful? In: BEERI, C.; BUNEMAN, P. (Ed.). *Database Theory — ICDT’99*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999. p. 217–235. ISBN 978-3-540-49257-3. Citado 4 vezes nas páginas 23, 40, 53 e 68.
- BISHOP, C. M.; BISHOP, H. *Deep Learning: Foundations and Concepts*. Cham, Switzerland: Springer, 2024. ISBN 978-3-031-45468-4. Citado na página 37.
- BLAGUS, R.; LUSA, L. Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics*, v. 11, n. 1, p. 523, 2010. ISSN 1471-2105. Disponível em: <<https://doi.org/10.1186/1471-2105-11-523>>. Citado na página 41.

BLOCKEEL, H. et al. Decision trees: from efficient prediction to responsible ai. *Frontiers in Artificial Intelligence*, Volume 6 - 2023, 2023. ISSN 2624-8212. Disponível em: <<https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1124553>>. Citado na página 19.

BONNIN, R. *Machine Learning for Developers: Uplift your regular applications with the power of statistics, analytics, and machine learning*. [S.l.]: Packt Publishing, 2017. ISBN 9781786466969. Citado 2 vezes nas páginas 64 e 65.

BONYADI, M. R.; TIENG, Q. M.; REUTENS, D. C. Optimization of distributions differences for classification. *IEEE Transactions on Neural Networks and Learning Systems*, v. 30, n. 2, p. 511–523, 2019. Citado 3 vezes nas páginas 79, 80 e 81.

BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, NY, USA: Association for Computing Machinery, 1992. (COLT '92), p. 144–152. ISBN 089791497X. Disponível em: <<https://doi.org/10.1145/130385.130401>>. Citado na página 44.

BOUSQUET, O.; ELISSEEFF, A. Stability and generalization. *J. Mach. Learn. Res.*, JMLR.org, v. 2, p. 499–526, mar. 2002. ISSN 1532-4435. Disponível em: <<https://doi.org/10.1162/153244302760200704>>. Citado na página 23.

BOUTABA, R. et al. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications*, v. 9, n. 1, p. 16, jun 2018. ISSN 1869-0238. Disponível em: <<https://doi.org/10.1186/s13174-018-0087-2>>. Citado na página 22.

BRANDAO, J. G. et al. Optimization of machine learning models for sentiment analysis in social media. *Information Sciences*, v. 694, p. 121704, 2025. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0020025524016189>>. Citado na página 22.

Brasil. Lei Federal, *Lei n° 13.709 – Lei Geral de Proteção de Dados*. 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm>. Citado 2 vezes nas páginas 31 e 33.

BYRD, R. H.; HRIBAR, M. E.; NOCEDAL, J. An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, v. 9, n. 4, p. 877–900, 1999. Disponível em: <<https://doi.org/10.1137/S1052623497325107>>. Citado 2 vezes nas páginas 47 e 49.

CADWALLADR, C. *'I Made Steve Bannon's Psychological Warfare Tool': Meet the Data War Whistleblower*. 2018. Disponível em: <<https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-facebook-nix-bannon-trump>>. Citado na página 27.

California State Legislature. *California Consumer Privacy Act of 2018*. 2018. California Civil Code, Division 3, Part 4, Title 1.81.5, §§ 1798.100-1798.199.100. Disponível em: <https://leginfo.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5>. Citado 2 vezes nas páginas 32 e 33.

- CARUANA, R.; NICULESCU-MIZIL, A. An empirical comparison of supervised learning algorithms. In: . New York, NY, USA: Association for Computing Machinery, 2006. (ICML '06), p. 161–168. ISBN 1595933832. Disponível em: <<https://doi.org/10.1145/1143844.1143865>>. Citado na página 36.
- CHAVES, M. *Os impactos da regulação da inteligência artificial para o ambiente de negócios*. 2024. Exame. Disponível em: <<https://exame.com/esferabrasil/os-impactos-da-regulacao-da-inteligencia-artificial-para-o-ambiente-de-negocios/>>. Citado na página 33.
- CHEN, W. et al. A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, v. 57, n. 6, p. 137, may 2024. ISSN 1573-7462. Disponível em: <<https://doi.org/10.1007/s10462-024-10759-6>>. Citado na página 15.
- CHICCO, D.; JURMAN, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, v. 21, n. 1, p. 6, 2020. ISSN 1471-2164. Disponível em: <<https://doi.org/10.1186/s12864-019-6413-7>>. Citado na página 69.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, 1995. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1007/BF00994018>>. Citado na página 44.
- DEB, K.; JAIN, H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, v. 18, n. 4, p. 577–601, 2014. Citado na página 46.
- DESHPANDE, P. On a generalization of zaslavsky’s theorem for hyperplane arrangements. *Annals of Combinatorics*, v. 18, n. 1, p. 35–55, mar 2014. ISSN 0219-3094. Disponível em: <<https://doi.org/10.1007/s00026-013-0210-9>>. Citado na página 24.
- DOMINGOS, P.; PAZZANI, M. On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning*, v. 29, p. 103–130, nov 1997. Disponível em: <<https://doi.org/10.1023/A:1007413511361>>. Citado na página 52.
- DONOHO, D. L. et al. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, v. 1, n. 2000, p. 32, 2000. Citado 2 vezes nas páginas 22 e 23.
- DOURADO, T. M. S. G. *Fake News na eleição presidencial de 2018 no Brasil*. 364 p. Tese (Doutorado) — Universidade Federal da Bahia, Salvador, BA, 2020. Programa de Pós-Graduação em Comunicação e Culturas Contemporâneas. Disponível em: <<https://repositorio.ufba.br/handle/ri/31967>>. Citado na página 27.
- DUDA, R.; HART, P.; STORK, D. *Pattern Classification*. [S.l.]: Wiley, 2012. ISBN 9781118586006. Citado 2 vezes nas páginas 37 e 39.
- DUKES, K. A. Gram–schmidt process. In: _____. *Wiley StatsRef: Statistics Reference Online*. John Wiley Sons, Ltd, 2014. ISBN 9781118445112. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat05633>>. Citado na página 77.

European Parliament; Council of the European Union. *General Data Protection Regulation*. 2016. Regulation (EU) 2016/679. Official Journal of the European Union, L 119, 4.5.2016, p. 1-88. Disponível em: <<https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>>. Citado 2 vezes nas páginas 31 e 33.

European Union. *High-Level Summary of the EU AI Act*. 2024. Disponível em: <<https://artificialintelligenceact.eu/high-level-summary/>>. Citado 2 vezes nas páginas 31 e 34.

Facebook. *Facebook to Acquire Instagram*. 2012. Disponível em: <<https://about.fb.com/news/2012/04/facebook-to-acquire-instagram/>>. Citado na página 26.

FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016786550500303X>>. Citado na página 70.

FIELD, A. *Discovering Statistics using IBM SPSS Statistics*. 4th. ed. [S.l.]: Sage Publications Ltd., 2013. ISBN 1446249182. Citado na página 121.

FLACH, P. A.; KULL, M. Precision-recall-gain curves: Pr analysis done right. In: *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1*. Cambridge, MA, USA: MIT Press, 2015. (NIPS'15), p. 838–846. Citado na página 68.

GAO, L. et al. Learning in high-dimensional multimedia data: the state of the art. *Multimedia Systems*, Springer, v. 23, n. 3, p. 303–313, 2017. ISSN 1432-1882. Disponível em: <<https://doi.org/10.1007/s00530-015-0494-1>>. Citado na página 16.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: TEH, Y. W.; TITTERINGTON, M. (Ed.). *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Chia Laguna Resort, Sardinia, Italy: PMLR, 2010. (Proceedings of Machine Learning Research, v. 9), p. 249–256. Disponível em: <<https://proceedings.mlr.press/v9/glorot10a.html>>. Citado na página 82.

GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st. ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1989. ISBN 0201157675. Citado 2 vezes nas páginas 48 e 49.

GOLDBERGER, J. et al. Neighbourhood components analysis. In: *Proceedings of the 18th International Conference on Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2004. (NIPS'04), p. 513–520. Citado na página 70.

GOMEZ-URIBE, C. A.; HUNT, N. The netflix recommender system: Algorithms, business value, and innovation. Association for Computing Machinery, New York, NY, USA, v. 6, n. 4, dez. 2016. ISSN 2158-656X. Disponível em: <<https://doi.org/10.1145/2843948>>. Citado na página 129.

GUNDERSEN, O. E.; KJENSMO, S. State of the art: Reproducibility in artificial intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 32, n. 1, Apr. 2018. Disponível em: <<https://ojs.aaai.org/index.php/AAAI/article/view/11503>>. Citado na página 37.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, JMLR.org, v. 3, n. null, p. 1157–1182, mar. 2003. ISSN 1532-4435. Citado na página 22.

HAIXIANG, G. et al. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, v. 73, p. 220–239, 2017. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417416307175>>. Citado na página 26.

HAMILTON, W. L.; YING, R.; LESKOVEC, J. Inductive representation learning on large graphs. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 1025–1035. ISBN 9781510860964. Citado na página 74.

HAN, H. Challenges of reproducible ai in biomedical data science. *BMC Medical Genomics*, v. 18, n. 1, p. 8, 2025. ISSN 1755-8794. Disponível em: <<https://doi.org/10.1186/s12920-024-02072-6>>. Citado 2 vezes nas páginas 38 e 41.

HASTIE ROBERT TIBSHIRANI, J. F. T. *Linear Discriminant Analysis*. New York, NY: Springer, 2009. ISBN 978-0-387-84858-7. Disponível em: <<https://doi.org/10.1007/978-0-387-84858-7>>. Citado na página 55.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: data mining, inference, and prediction*. 2. ed. [S.l.]: Springer, 2009. Citado na página 23.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation (3rd Edition)*. USA: Prentice-Hall, Inc., 2007. ISBN 0131471392. Citado 3 vezes nas páginas 56, 57 e 59.

HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 9, p. 1263–1284, 2009. Citado 3 vezes nas páginas 24, 25 e 104.

HENDERSON, P. et al. Deep reinforcement learning that matters. In: *Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18)*. [S.l.]: AAAI Press, 2018. (AAAI'18/IAAI'18/EAAI'18). ISBN 978-1-57735-800-8. Citado na página 38.

HEUSINGER, M.; RAAB, C.; SCHLEIF, F.-M. Dimensionality reduction in the context of dynamic social media data streams. *Evolving Systems*, Springer, v. 13, n. 3, p. 387–401, 2022. ISSN 1868-6486. Disponível em: <<https://doi.org/10.1007/s12530-021-09396-z>>. Citado na página 16.

Hootsuite. *Brazil Social Media Statistics 2024*. 2024. Disponível em: <<https://blog.hootsuite.com/instagram-statistics/>>. Citado na página 27.

HOUSTON, J.; GLAVIN, F. G.; MADDEN, M. G. Robust classification of high-dimensional spectroscopy data using deep learning and data synthesis. *Journal of Chemical Information and Modeling*, American Chemical Society, v. 60, n. 4, p. 1936–1954, 2020. ISSN 1549-9596. Disponível em: <<https://doi.org/10.1021/acs.jcim.9b01037>>. Citado na página 45.

- INDARJO, P. *Statistika Deskriptif untuk Data Scientist*. 2024. Disponível em: <<https://medium.com/@pararawendy19/statistika-deskriptif-untuk-data-scientist-845115435ecb>>. Citado na página 63.
- IZENMAN, A. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer New York, 2009. (Springer Texts in Statistics). ISBN 9780387781891. Disponível em: <<https://books.google.com/books?id=1CuznRORa3EC>>. Citado na página 78.
- JAIN, A.; DUIN, R.; MAO, J. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 1, p. 4–37, 2000. Citado na página 24.
- JAIN, A. K. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, v. 31, n. 8, p. 651–666, 2010. ISSN 0167-8655. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865509002323>>. Citado 2 vezes nas páginas 37 e 39.
- JAPKOWICZ, N.; STEPHEN, S. The class imbalance problem: A systematic study1. *Intelligent Data Analysis*, v. 6, n. 5, p. 429–449, 2002. Disponível em: <<https://journals.sagepub.com/doi/abs/10.3233/IDA-2002-6504>>. Citado na página 26.
- JIA, W. et al. Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, v. 8, n. 3, p. 2663–2693, jun 2022. ISSN 2198-6053. Disponível em: <<https://doi.org/10.1007/s40747-021-00637-x>>. Citado na página 15.
- JIAO, R. et al. A survey on evolutionary multiobjective feature selection in classification: Approaches, applications, and challenges. *IEEE Transactions on Evolutionary Computation*, v. 28, n. 4, p. 1156–1176, 2024. Citado na página 46.
- JOHNSON, J. M.; KHOSHGOFTAAR, T. M. Survey on deep learning with class imbalance. *Journal of Big Data*, v. 6, n. 1, p. 27, 2019. ISSN 2196-1115. Disponível em: <<https://doi.org/10.1186/s40537-019-0192-5>>. Citado na página 26.
- JOHNSTONE, I. M. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 29, n. 2, p. 295–327, 2001. ISSN 00905364, 21688966. Disponível em: <<http://www.jstor.org/stable/2674106>>. Citado na página 24.
- JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, The Royal Society, v. 374, 2016. Disponível em: <<https://doi.org/10.1098/rsta.2015.0202>>. Citado 2 vezes nas páginas 16 e 37.
- KATOCH, S.; CHAUHAN, S. S.; KUMAR, V. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, v. 80, n. 5, p. 8091–8126, 2 2021. ISSN 1573-7721. Disponível em: <<https://doi.org/10.1007/s11042-020-10139-6>>. Citado 2 vezes nas páginas 48 e 49.
- KAZEMI, S. M. et al. Representation learning for dynamic graphs: a survey. *J. Mach. Learn. Res.*, JMLR.org, v. 21, n. 1, jan. 2020. ISSN 1532-4435. Citado na página 74.

KEMP, S. *Digital 2024: Brazil*. 2024. Disponível em: <<https://datareportal.com/reports/digital-2024-brazil>>. Citado na página 27.

KEMP, S. *Digital 2024: Global Overview Report*. 2024. Disponível em: <<https://datareportal.com/reports/digital-2024-global-overview-report>>. Citado na página 72.

KHOSLA, A.; SARMA, A. D.; HAMID, R. What makes an image popular? In: *Proceedings of the 23rd International Conference on World Wide Web*. New York, NY, USA: Association for Computing Machinery, 2014. (WWW '14), p. 867–876. ISBN 9781450327442. Disponível em: <<https://doi.org/10.1145/2566486.2567996>>. Citado na página 73.

KINGMA, D. P.; BA, J. *Adam: A Method for Stochastic Optimization*. 2017. Disponível em: <<https://arxiv.org/abs/1412.6980>>. Citado na página 82.

KOBAK, D.; BERENS, P. The art of using t-sne for single-cell transcriptomics. *Nature Communications*, v. 10, n. 1, p. 5416, 2019. ISSN 2041-1723. Disponível em: <<https://doi.org/10.1038/s41467-019-13056-x>>. Citado na página 72.

KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, v. 26, n. 3, p. 159–190, nov 2006. ISSN 1573-7462. Disponível em: <<https://doi.org/10.1007/s10462-007-9052-3>>. Citado na página 19.

KOTZ, S.; KOZUBOWSKI, T.; PODGORSKI, K. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser Boston, 2012. (SpringerLink : Bücher). ISBN 9781461201731. Disponível em: <<https://books.google.com/books?id=qhPSBwAAQBAJ>>. Citado na página 85.

KUMAR, A. *Linear vs Non-linear Data: How to Know*. 2020. Disponível em: <<https://vitalflux.com/how-know-data-linear-non-linear/>>. Citado na página 42.

LAGARIAS, J. C. et al. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, v. 9, n. 1, p. 112–147, 1998. Disponível em: <<https://doi.org/10.1137/S1052623496303470>>. Citado 3 vezes nas páginas 47, 49 e 81.

LANCKRIET, G. R. et al. A robust minimax approach to classification. *J. Mach. Learn. Res.*, JMLR.org, v. 3, n. null, p. 555–582, mar. 2003. ISSN 1532-4435. Disponível em: <<https://doi.org/10.1162/153244303321897726>>. Citado na página 76.

LAWRENCE, S.; GILES, C. Overfitting and neural networks: conjugate gradient and backpropagation. In: *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*. [S.l.: s.n.], 2000. v. 1, p. 114–119 vol.1. Citado na página 59.

LEVATIĆ, J. et al. Semi-supervised predictive clustering trees for (hierarchical) multi-label classification. *International Journal of Intelligent Systems*, v. 2024, n. 1, p. 5610291, 2024. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1155/2024/5610291>>. Citado na página 19.

- LI, M. et al. Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, Volume 11 - 2023, 2023. ISSN 2296-2565. Disponível em: <<https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1273253>>. Citado na página 22.
- LIN, W.-J.; CHEN, J. J. Class-imbalanced classifiers for high-dimensional data. *Briefings in Bioinformatics*, v. 14, n. 1, p. 13–26, 02 2012. ISSN 1467-5463. Disponível em: <<https://doi.org/10.1093/bib/bbs006>>. Citado na página 15.
- LINARDATOS, P.; PAPASTEFANOPOULOS, V.; KOTSIANTIS, S. Explainable ai: A review of machine learning interpretability methods. *Entropy*, v. 23, n. 1, 2021. ISSN 1099-4300. Disponível em: <<https://www.mdpi.com/1099-4300/23/1/18>>. Citado 2 vezes nas páginas 41 e 43.
- LIU, X.; ZHENG, X. The persuasive power of social media influencers in brand credibility and purchase intention. *Humanities and Social Sciences Communications*, v. 11, n. 1, p. 15, 2024. ISSN 2662-9992. Disponível em: <<https://doi.org/10.1057/s41599-023-02512-1>>. Citado na página 73.
- LIU, Y.; WU, Y.-F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 32, n. 1, Apr. 2018. Disponível em: <<https://ojs.aaai.org/index.php/AAAI/article/view/11268>>. Citado na página 72.
- LNCC. *Plano Brasileiro de Inteligência Artificial (PBIA) 2024-2028*. 2024. Disponível em: <<https://www.gov.br/lncc/pt-br/assuntos/noticias/ultimas-noticias-1/plano-brasileiro-de-inteligencia-artificial-pbia-2024-2028>>. Citado na página 31.
- MAATEN, L. V. D. et al. Dimensionality reduction: A comparative review. *Journal of machine learning research*, v. 10, n. 66-71, p. 13, 2009. Citado na página 22.
- MAATEN, L. van der. Accelerating t-sne using tree-based algorithms. *Journal of Machine Learning Research*, v. 15, n. 93, p. 3221–3245, 2014. Disponível em: <<http://jmlr.org/papers/v15/vandermaaten14a.html>>. Citado na página 40.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, v. 5, n. 4, p. 115–133, 1943. ISSN 1522-9602. Disponível em: <<https://doi.org/10.1007/BF02478259>>. Citado na página 56.
- MEHTA, S.; KUMAR, R. Federated deep learning for multiclass sentiment analysis: Privacy-preserving cnns with optimized aggregation. In: *2024 2nd International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMAACC)*. [S.l.: s.n.], 2024. p. 386–390. Citado na página 73.
- Meta. *The AI behind unconnected content recommendations on Facebook and Instagram*. 2023. Disponível em: <<https://ai.meta.com/blog/ai-unconnected-content-recommendations-facebook-instagram>>. Citado na página 129.
- MONTGOMERY, D. C.; RUNGER, G. C. *Applied Statistics and Probability for Engineers*. 7th. ed. [S.l.]: John Wiley & Sons, 2017. Citado na página 121.

- MOON, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, v. 37, n. 12, p. 1482–1492, 2019. ISSN 1546-1696. Disponível em: <<https://doi.org/10.1038/s41587-019-0336-3>>. Citado na página 72.
- MOULINES, C. U. Metatheoretical structuralism: a general program for analyzing science. *Axiomathes*, v. 20, n. 2, p. 255–268, September 2010. ISSN 1572-8390. Disponível em: <<https://doi.org/10.1007/s10516-010-9106-z>>. Citado na página 39.
- MURPHY, K. P. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. Disponível em: <<http://probml.github.io/book1>>. Citado na página 84.
- MUTAMBIK, I. An entropy-based clustering algorithm for real-time high-dimensional iot data streams. *Sensors*, v. 24, n. 22, 2024. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/24/22/7412>>. Citado na página 16.
- NARAYANAN, U. et al. A survey on various supervised classification algorithms. In: *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. [S.l.: s.n.], 2017. p. 2118–2124. Citado na página 15.
- NARKHEDE, M. V.; BARTAKKE, P. P.; SUTAONE, M. S. A review on weight initialization strategies for neural networks. *Artificial Intelligence Review*, v. 55, n. 1, p. 291–322, 2022. ISSN 1573-7462. Disponível em: <<https://doi.org/10.1007/s10462-021-10033-z>>. Citado 2 vezes nas páginas 40 e 41.
- NETZER, O.; LEMAIRE, A.; HERZENSTEIN, M. When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, v. 56, n. 6, p. 960–980, 2019. Disponível em: <<https://doi.org/10.1177/0022243719852959>>. Citado 2 vezes nas páginas 19 e 21.
- NEWMAN, M. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, Taylor & Francis, v. 46, n. 5, p. 323–351, 2005. Disponível em: <<https://doi.org/10.1080/00107510500052444>>. Citado na página 25.
- NGO, N.-T. et al. An ensemble machine learning model for enhancing the prediction accuracy of energy consumption in buildings. *Arabian Journal for Science and Engineering*, v. 47, n. 4, p. 4105–4117, 4 2022. ISSN 2191-4281. Disponível em: <<https://doi.org/10.1007/s13369-021-05927-7>>. Citado na página 58.
- NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. 2nd. ed. New York: Springer, 2006. Comprehensive treatment of optimization methods for machine learning applications. ISBN 978-0387303031. Citado na página 47.
- OLIVEIRA, M. Martins de; SILVEIRA, D. Barile da; OLIVEIRA, M. d. G. Macena Dias de. Análise comparada da normas de proteção de dados do brasil, da união europeia e do estado da califórnia - eua: Lgpd x gdpr x ccpa: Lgpg x gdpr x ccpa. *Revista de Direito, Governança e Novas Tecnologias*, v. 10, n. 2, mar. 2025. Disponível em: <<https://www.indexlaw.org/index.php/revistadgnt/article/view/10923>>. Citado 2 vezes nas páginas 32 e 34.
- OPITZ, J.; BURST, S. *Macro F1 and Macro F1*. 2021. Disponível em: <<https://arxiv.org/abs/1911.03347>>. Citado na página 70.

- OSISANWO, F. Y. et al. Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, v. 48, n. 2, p. 128–138, 2017. Disponível em: <<https://doi.org/10.14445/22312803/IJCTT-V48P126>>. Citado na página 15.
- PAJILA, P. et al. A comprehensive survey on naive bayes algorithm: Advantages, limitations and applications. In: *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*. [S.l.: s.n.], 2023. p. 1228–1234. Citado na página 21.
- PANCHAL, F. S.; PANCHAL, M. Review on methods of selecting number of hidden nodes in artificial neural network. *International Journal of Computer Science and Mobile Computing*, Zain Publications, v. 3, n. 11, p. 455–464, 2014. Citado na página 60.
- PARK, D. H. et al. A literature review and classification of recommender systems research. *Expert Systems with Applications*, v. 39, n. 11, p. 10059–10072, 2012. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417412002825>>. Citado na página 128.
- PENG, D.; GUI, Z.; WU, H. *Interpreting the Curse of Dimensionality from Distance Concentration and Manifold Effect*. 2025. Disponível em: <<https://arxiv.org/abs/2401.00422>>. Citado na página 22.
- PEREG, D. Information theoretic perspective on sample complexity. *Neural Networks*, v. 167, p. 445–449, 2023. ISSN 0893-6080. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0893608023004537>>. Citado na página 23.
- POWERS, D. M. W. *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. 2020. Disponível em: <<https://arxiv.org/abs/2010.16061>>. Citado 3 vezes nas páginas 67, 68 e 69.
- QIU, Q.; SAPIRO, G. Learning transformations for clustering and classification. *J. Mach. Learn. Res.*, JMLR.org, v. 16, n. 1, p. 187–225, jan. 2015. ISSN 1532-4435. Citado 2 vezes nas páginas 81 e 82.
- QIU, X. et al. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, v. 63, n. 10, p. 1872–1897, 2020. ISSN 1869-1900. Disponível em: <<https://doi.org/10.1007/s11431-020-1647-3>>. Citado na página 72.
- RENNIE, J. D. et al. Tackling the poor assumptions of naive bayes text classifiers. In: *Proceedings of the 20th international conference on machine learning (ICML-03)*. [S.l.: s.n.], 2003. p. 616–623. Citado na página 52.
- RIOS, L. M.; SAHINIDIS, N. V. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, v. 56, n. 3, p. 1247–1293, 2013. ISSN 1573-2916. Disponível em: <<https://doi.org/10.1007/s10898-012-9951-y>>. Citado na página 48.
- Rivery. *Data Statistics (2025) - How much data is there in the world?* 2024. Disponível em: <<https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/>>. Citado na página 71.

- ROGERS, A.; KOVALEVA, O.; RUMSHISKY, A. A primer in bertology: What we know about how bert works. *Transactions of the Association for Computational Linguistics*, v. 8, p. 842–866, 01 2021. ISSN 2307-387X. Disponível em: <https://doi.org/10.1162/tacl_a_00349>. Citado na página 72.
- RUNCIE, D. E.; MUKHERJEE, S. Dissecting high-dimensional phenotypes with bayesian sparse factor analysis of genetic covariance matrices. *Genetics*, v. 194, n. 3, p. 753–767, 07 2013. ISSN 1943-2631. Disponível em: <<https://doi.org/10.1534/genetics.113.151217>>. Citado na página 24.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3rd. ed. USA: Prentice Hall Press, 2009. ISBN 0136042597. Citado na página 36.
- SAINBURG, T.; MCINNES, L.; GENTNER, T. Q. *Parametric UMAP embeddings for representation and semi-supervised learning*. 2021. Disponível em: <<https://arxiv.org/abs/2009.12981>>. Citado na página 72.
- SASSON, J. *Demystifying Artificial Intelligence*. 2023. Disponível em: <<https://medium.com/geekculture/demystifying-artificial-intelligence-bdd9a117d4a6>>. Citado na página 58.
- SCHOLKOPF, B.; SMOLA, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2001. ISBN 9780262256933. Disponível em: <<https://doi.org/10.7551/mitpress/4175.001.0001>>. Citado 2 vezes nas páginas 41 e 45.
- SERANI, A.; DIEZ, M. A survey on design-space dimensionality reduction methods for shape optimization. *Archives of Computational Methods in Engineering*, sep 2025. ISSN 1886-1784. Disponível em: <<https://doi.org/10.1007/s11831-025-10349-x>>. Citado na página 16.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. *Understanding Machine Learning: From Theory to Algorithms*. [S.l.]: Cambridge University Press, 2014. Citado na página 24.
- SHARMA, A.; PALIWAL, K. K. Linear discriminant analysis for the small sample size problem: an overview. *International Journal of Machine Learning and Cybernetics*, v. 6, p. 443–454, jun 2015. Disponível em: <<https://doi.org/10.1007/s13042-013-0226-9>>. Citado na página 55.
- SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, Annual Reviews, v. 19, n. Volume 19, 2017, p. 221–248, 2017. ISSN 1545-4274. Disponível em: <<https://www.annualreviews.org/content/journals/10.1146/annurev-bioeng-071516-044442>>. Citado na página 22.
- SOBRINHO, J. L. V. *Rastreador web não supervisionado para aquisição, enriquecimento e predição de dados de usuários de redes sociais por intermédio de métodos de inteligência computacional*. 93 p. Dissertação (Dissertação (Mestrado em Engenharia Elétrica e da Computação)) — Universidade Federal de Goiás, Goiânia, 2019. Disponível em: <<https://repositorio.bc.ufg.br/tede/items/e28e1544-295c-45ab-9062-c462a649821e>>. Citado 2 vezes nas páginas 73 e 105.

- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information Processing Management*, v. 45, n. 4, p. 427–437, 2009. ISSN 0306-4573. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306457309000259>>. Citado 6 vezes nas páginas 60, 61, 62, 66, 68 e 69.
- SOLTANZADEH, P.; FEIZI-DERAKHSHI, M. R.; HASHEMZADEH, M. Addressing the class-imbalance and class-overlap problems by a metaheuristic-based under-sampling approach. *Pattern Recognition*, v. 143, p. 109721, 2023. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320323004193>>. Citado na página 15.
- SONG, L. et al. Colored maximum variance unfolding. In: PLATT, J. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2007. v. 20. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2007/file/55a7cf9c71f1c9c495413f934dd1a158-Paper.pdf>. Citado na página 71.
- SOUZA, P. V. de C.; LUGHOFER, E. An advanced interpretable fuzzy neural network model based on uni-nullneuron constructed from n-uniforms. *Fuzzy Sets and Systems*, v. 426, p. 1–26, 2022. ISSN 0165-0114. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0165011420304632>>. Citado na página 15.
- Statista. *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023, with forecasts from 2024 to 2028*. 2024. Disponível em: <<https://www.statista.com/statistics/871513/worldwide-data-created/>>. Citado na página 71.
- Statista. *Countries with the Most Instagram Users 2025*. 2025. Disponível em: <<https://www.statista.com/statistics/578364/countries-with-most-instagram-users/>>. Citado na página 30.
- Statista. *Distribution of Instagram Users Worldwide as of April 2024, by Age Group*. 2025. Disponível em: <<https://www.statista.com/statistics/325587/instagram-global-age-group/>>. Citado na página 29.
- Statista. *Instagram: Number of Global Users 2020-2025*. 2025. Disponível em: <<https://www.statista.com/statistics/183585/instagram-number-of-global-users/>>. Citado na página 29.
- Statista. *Most Popular Social Networks Worldwide as of February 2025, by Number of Monthly Active Users (in Millions)*. 2025. Disponível em: <<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>>. Citado na página 28.
- STEINBACH, M.; ERTÖZ, L.; KUMAR, V. The challenges of clustering high dimensional data. In: _____. *New Directions in Statistical Physics: Econophysics, Bioinformatics, and Pattern Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 273–309. ISBN 978-3-662-08968-2. Disponível em: <https://doi.org/10.1007/978-3-662-08968-2_16>. Citado na página 16.
- STEPHENS-DAVIDOWITZ, S. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*. [S.l.]: HarperCollins, 2017. ISBN 9780062390875. Citado na página 128.

SUGIYAMA, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, v. 8, n. 37, p. 1027–1061, 2007. Disponível em: <<http://jmlr.org/papers/v8/sugiyama07b.html>>. Citado na página 71.

SUGIYAMA, M. et al. Semi-supervised local Fisher discriminant analysis for dimensionality reduction. *Machine Learning*, v. 78, n. 1, p. 35–61, 2010. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1007/s10994-009-5125-7>>. Citado na página 71.

SUTTON, R.; BARTO, A. *Reinforcement Learning, second edition: An Introduction*. [S.l.]: MIT Press, 2018. (Adaptive Computation and Machine Learning series). ISBN 9780262352703. Citado 2 vezes nas páginas 37 e 39.

TAN, S. *You are ideating AI solutions wrong*. 2024. Disponível em: <<https://medium.com/design-bootcamp/you-are-ideating-ai-solutions-wrong-heres-how-9b408c3bcfa8>>. Citado na página 38.

THUDUMU, S. et al. A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, v. 7, n. 1, p. 42, jul 2020. ISSN 2196-1115. Disponível em: <<https://doi.org/10.1186/s40537-020-00320-x>>. Citado na página 15.

Tribunal Superior Eleitoral. *Combate à desinformação: TSE derruba mais de uma centena de postagens com narrativas enganosas*. 2022. Disponível em: <<https://www.tse.jus.br/comunicacao/noticias/2022/Outubro/combate-a-desinformacao-tse-derruba-mais-de-uma-centena-de-postagens-com-narrativas-enganosas>>. Citado na página 27.

TRUNFIO, M.; ROSSI, S. Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing*, v. 2021, n. 3, p. 267–292, 9 2021. ISSN 2662-3331. Disponível em: <<https://doi.org/10.1007/s43039-021-00035-8>>. Citado na página 106.

UDDIN, S. et al. Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, v. 19, p. 281, dec 2019. Disponível em: <<https://doi.org/10.1186/s12911-019-1004-8>>. Citado na página 50.

USERA, D.; DURHAM, N. What predicts engagement on linkedin? engagement-boosting strategies for professionals. *Business and Professional Communication Quarterly*, 2025. Disponível em: <<https://doi.org/10.1177/23294906251336710>>. Citado na página 73.

VALLET, D. et al. Characterizing and predicting viral-and-popular video content. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2015. (CIKM '15), p. 1591–1600. ISBN 9781450337946. Disponível em: <<https://doi.org/10.1145/2806416.2806556>>. Citado na página 73.

VAPNIK, V. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, v. 10, n. 5, p. 988–999, 1999. Citado na página 23.

VELLIANGIRI, S.; ALAGUMUTHUKRISHNAN, S.; Thankumar joseph, S. I. A review of dimensionality reduction techniques for efficient computation. *Procedia Computer Science*, v. 165, p. 104–111, 2019. ISSN 1877-0509. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC

- DISRUP - TIV INNOVATION , 2019 November 11-12, 2019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050920300879>>. Citado na página 16.
- VERLEYSSEN, M.; FRANÇOIS, D. The curse of dimensionality in data mining and time series prediction. In: CABESTANY, J.; PRIETO, A.; SANDOVAL, F. (Ed.). *Computational Intelligence and Bioinspired Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. p. 758–770. ISBN 978-3-540-32106-4. Citado na página 24.
- VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. *Science*, v. 359, n. 6380, p. 1146–1151, 2018. Disponível em: <<https://www.science.org/doi/abs/10.1126/science.aap9559>>. Citado na página 72.
- VUTTIPITTAYAMONGKOL, P.; ELYAN, E.; PETROVSKI, A. On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems*, v. 212, p. 106631, 2021. ISSN 0950-7051. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950705120307607>>. Citado na página 15.
- WALPOLE, R. E. et al. *Probability and Statistics for Engineers and Scientists*. 9th. ed. [S.l.]: Pearson, 2016. Citado na página 121.
- WANG, J. et al. Weight masking in image classification networks: class-specific machine unlearning. *Knowledge and Information Systems*, v. 67, n. 4, p. 3245–3265, 2025. ISSN 0219-3116. Disponível em: <<https://doi.org/10.1007/s10115-024-02312-2>>. Citado na página 44.
- WANG, K. et al. *Synthetic Data Generation and Adaption for Object Detection in Smart Vending Machines*. 2019. Disponível em: <<https://arxiv.org/abs/1904.12294>>. Citado na página 74.
- WAY, G. P.; GREENE, C. S. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. In: *Pacific Symposium on Biocomputing*. [S.l.: s.n.], 2018. v. 23, p. 80–91. Citado na página 72.
- WEINBERGER, K. Q.; SAUL, L. K. Distance metric learning for large margin nearest neighbor classification. *Journal of machine learning research*, v. 10, n. 2, 2009. Citado na página 53.
- WEINBERGER, K. Q.; SHA, F.; SAUL, L. K. Learning a kernel matrix for nonlinear dimensionality reduction. In: *Proceedings of the Twenty-First International Conference on Machine Learning*. New York, NY, USA: Association for Computing Machinery, 2004. (ICML '04), p. 106. ISBN 1581138385. Disponível em: <<https://doi.org/10.1145/1015330.1015345>>. Citado na página 70.
- WILIMITIS, D. *The Kernel Trick*. 2018. Disponível em: <<https://medium.com/data-science/the-kernel-trick-c98cdbcaeb3f>>. Citado na página 45.
- XANTHOPOULOS, P.; PARDALOS, P. M.; TRAFALIS, T. B. Linear discriminant analysis. In: _____. *Robust Data Mining*. New York, NY: Springer New York, 2013. p. 27–33. ISBN 978-1-4419-9878-1. Disponível em: <https://doi.org/10.1007/978-1-4419-9878-1_4>. Citado 2 vezes nas páginas 16 e 54.

- XU, D.; TIAN, Y. A comprehensive survey of clustering algorithms. *Annals of Data Science*, v. 2, n. 2, p. 165–193, 6 2015. ISSN 2198-5812. Disponível em: <<https://doi.org/10.1007/s40745-015-0040-1>>. Citado na página 85.
- XUE, J. et al. Word embedding dimension reduction via weakly-supervised feature selection. *APSIPA Transactions on Signal and Information Processing*, Now Publishers, Inc., v. 13, n. 1, 2024. Citado na página 24.
- YADAV, A. et al. Analyzing user behavior in social media through big data analytics. In: *2023 IEEE International Conference on ICT in Business Industry Government (ICTBIG)*. [S.l.: s.n.], 2023. p. 1–5. Citado na página 16.
- YANG, L. et al. Nonparametric dimension reduction via maximizing pairwise separation probability. *IEEE Transactions on Neural Networks and Learning Systems*, v. 30, n. 10, p. 3205–3210, 2019. Citado 3 vezes nas páginas 76, 77 e 78.
- YOKOYAMA, A. M.; FERRO, M.; SCHULZE, B. Multi-objective hyperparameter optimization approach with genetic algorithms towards efficient and environmentally friendly machine learning. *AI Communications*, v. 37, n. 3, p. 429–442, 2024. Disponível em: <<https://journals.sagepub.com/doi/abs/10.3233/AIC-230063>>. Citado na página 48.
- YUAN, G.-X.; HO, C.-H.; LIN, C.-J. Recent advances of large-scale linear classification. *Proceedings of the IEEE*, v. 100, n. 9, p. 2584–2603, 2012. Citado 2 vezes nas páginas 40 e 42.
- ZADEH, A. et al. *Tensor Fusion Network for Multimodal Sentiment Analysis*. 2017. Disponível em: <<https://arxiv.org/abs/1707.07250>>. Citado na página 73.
- ZHANG, J.; YU, J.; TAO, D. Local deep-feature alignment for unsupervised dimension reduction. *IEEE Transactions on Image Processing*, v. 27, n. 5, p. 2420–2432, 2018. Citado na página 71.
- ZHANG, Q.; LI, H. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, v. 11, n. 6, p. 712–731, 2007. Citado na página 46.
- ZHAO, J. et al. Dimension reduction for covariates in network data. *Biometrika*, v. 109, n. 1, p. 85–102, 02 2021. ISSN 1464-3510. Disponível em: <<https://doi.org/10.1093/biomet/asab006>>. Citado na página 16.
- ZHAO, S. et al. Linear discriminant analysis. *Nature Reviews Methods Primers*, v. 4, n. 1, p. 70, 2024. ISSN 2662-8449. Disponível em: <<https://doi.org/10.1038/s43586-024-00346-y>>. Citado na página 42.
- ZHOU, X.; ZAFARANI, R. Network-based fake news detection: A pattern-driven approach. Association for Computing Machinery, New York, NY, USA, v. 21, n. 2, p. 48–60, nov. 2019. ISSN 1931-0145. Disponível em: <<https://doi.org/10.1145/3373464.3373473>>. Citado na página 73.
- ZHOU, Z. *Ensemble Methods: Foundations and Algorithms*. [S.l.]: CRC Press, 2012. (Chapman & Hall/CRC Machine Learning & Pattern Recognition). ISBN 9781439830055. Citado na página 40.

ZHU, Y. et al. Cdf transform-and-shift: An effective way to deal with datasets of inhomogeneous cluster densities. *Pattern Recognition*, v. 117, p. 107977, 2021. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320321001643>>. Citado na página 83.

ZITZLER, E.; THIELE, L. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Transactions on Evolutionary Computation*, v. 3, n. 4, p. 257–271, 1999. Citado na página 46.