Análise e Classificação de E-mails SPAM com Machine Learning

João Vítor Batistella[1], **Andrws Aires Vieira**[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - *Campus* Ibirubá
Rua Nelsi Ribas Fritsch, 1111 – CEP: 98200-000 – Ibirubá – RS – Brasil

joaovitor.batistella.ifrs@gmail.com, andrws.vieira@ibiruba.ifrs.edu.br

Abstract. *This paper proposes the development of a Machine Learning model to classify e-mails as SPAM or HAM (not SPAM). Considering the growing relevance of unwanted e-mails in global data traffic, the research aims to develop a solution that optimizes computational resources and supports digital marketing companies. Using Python, the study applies Logistic Regression to analyze e-mail content and compares its performance with the well-established Naïve Bayes classifier. The study seeks to promote more sustainable digital marketing practices by reducing unwanted communications and preserving sender reputation.*

Resumo. *O trabalho propõe desenvolver um modelo de Machine Learning para classificar e-mails como SPAM ou HAM (não SPAM). Considerando a relevância crescente do problema dos e-mails indesejados no tráfego global de dados, a pesquisa busca criar uma solução que otimize recursos computacionais e auxilie empresas de marketing digital. Utilizando Python, o estudo será fundamentado na aplicação de Regressão Logística para analisar o conteúdo dos e-mails, além de comparar seu desempenho com o consolidado classificador Naïve Bayes. O estudo busca promover práticas mais sustentáveis de marketing digital, reduzindo o impacto das comunicações indesejadas e preservando a reputação dos remetentes.*

1. Introdução

O *e-mail* é uma ferramenta essencial no marketing digital, mas a alta taxa de mensagens classificadas como SPAM tem gerado desafios significativos. Segundo [Statista 2024a], em abril de 2024, usuários nos Estados Unidos enviaram 9,7 bilhões de e-mails em um único dia. Já em dezembro, a média diária de e-mails classificados como SPAM chegou a 7,8 bilhões [Statista 2024b].O envio em massa, de baixo custo, favorece a disseminação de SPAM, que inclui anúncios comerciais, golpes financeiros e mensagens fraudulentas, além de resultar em desperdício de recursos computacionais [Sherwin 2023].

Para mitigar esse problema, propõe-se o desenvolvimento de um modelo de aprendizado de máquina capaz de identificar SPAM antes do envio, auxiliando empresas que realizam disparos em massa. A classificação eficaz dessas mensagens pode otimizar a infraestrutura, melhorar a reputação dos domínios de remetentes e aumentar a taxa de entrega nas caixas de entrada. Diante do avanço dos algoritmos de detecção de SPAM por provedores como Google e Microsoft, manter uma boa reputação de domínio tornou-se crucial para o sucesso das campanhas de marketing digital [Cappy 2024].

A pesquisa focará especificamente no desenvolvimento de um modelo de *Machine Learning* para identificação e classificação de *e-mails* SPAM, com ênfase na análise do corpo da mensagem como principal elemento de predição. O estudo se concentrará na aplicação de técnicas de processamento de linguagem natural e aprendizado de máquina, utilizando Regressão Logística como algoritmo principal, e *Naïve Bayes* com intuito de comparação e avaliação, para desenvolver um modelo capaz de classificar um *e-mail* como SPAM ou HAM, termo em inglês utilizado para nomear *e-mails* que geralmente são desejados, ou seja, o oposto de SPAM.

2. Referencial Teórico

Esta seção busca estabelecer as bases conceituais necessárias para a compreensão e análise do tema em estudo. Por meio de uma revisão bibliográfica, serão apresentados os principais conceitos, teorias e discussões que fundamentam esta pesquisa.

2.1. Filtragem de SPAM

Os filtros de SPAM podem ser implantados no lado do cliente, do servidor e de computadores intermediários [Dada et al. 2019]. Um dos itens considerados por filtros de SPAM é a reputação do domínio do remetente. Segundo [Dossetto 2022], a reputação de domínio é a opinião que os destinatários, incluindo provedores e serviços *anti-spam*, têm sobre um domínio. Os domínios utilizados nos remetentes de *e-mail* sofrem constantes análises quanto à sua qualidade. A reputação é uma maneira de determinar se os *e-mails* são confiáveis. Quem faz isso são os ISPs, analisando o histórico de envios em massa e envio de *phishing* e *spoofing* [Dada et al. 2019].

O *Phishing* é um cibercrime, cujo objetivo é roubar informações pessoais, como senhas ou número de cartões [Kosinski 2024]. O *spoofing* está comumente atrelado ao *phishing*, porém ele faz referência ao uso de técnicas de falsificação, onde os criminosos se passam por familiares ou empresas que o destinatário provavelmente conhece a fim de induzir a vítima a entregar as informações pretendidas [Pronnus 2024]. Um exemplo amplamente conhecido é a solicitação de pagamentos de taxa para o Correios. Sendo que, nesses casos, os autores escrevem o modelo de *e-mail* de forma que o destinatário não perceba que o remetente não são os Correios.

O conteúdo de um *e-mail* é composto por diferentes partes que precisam ser consideradas na mineração de dados. O cabeçalho e o corpo do *e-mail* são elementos fundamentais nesse processo. O corpo pode conter marcação HTML, imagens, arquivos anexos e outros tipos de dados, exigindo um tratamento adequado durante o pré-processamento [Dada et al. 2019]. Já o cabeçalho inclui metadados importantes, como o assunto do e-mail, que podem ser utilizados para auxiliar na classificação e análise das mensagens.

2.2. KDD (Knowledge Discovery in Databases)

O processo que abrange todas as etapas, desde a obtenção de dados brutos até a extração de conhecimento por meio da identificação de padrões, é denominado KDD (Knowledge Discovery in Databases), ou Descoberta de Conhecimento em Bancos de Dados. Esse processo é estruturado em cinco etapas fundamentais, que serão apresentadas a seguir.

2.2.1. Seleção de dados

Responsável pela definição do subconjunto de dados que será analisado no processo de KDD, a seleção de dados é uma etapa essencial que pode ser realizada de duas maneiras: por meio da escolha dos atributos relevantes ou da filtragem dos registros que serão submetidos à análise [Goldschmidt 2015].

2.2.2. Pré-processamento de dados

As bases de dados são compostas por diversas atributos. Após definir os atributos de interesse, a etapa de pré-processamento tem o propósito de preparar os dados para uma análise mais eficiente e precisa. Nessa fase, é essencial detectar e corrigir inconsistências ou ruídos, realizar a limpeza dos dados, preencher valores ausentes com métodos apropriados e eliminar elementos indesejáveis que possam impactar negativamente a extração de conhecimento [Mariano et al. 2021].

2.2.3. Transformação de dados

A etapa de transformação tem como objetivo ajustar os dados para formatos compatíveis com os métodos que serão utilizados na próxima etapa, a mineração de dados. Esse processo pode incluir, por exemplo, a conversão de dados contínuos em discretos quando o modelo exige essa estrutura e a normalização dos dados para garantir uma análise mais eficaz [Goldschmidt 2015].

2.2.4. Mineração de dados

A etapa de Mineração de Dados compreende a busca efetiva por conhecimentos úteis no contexto da aplicação de KDD. Esta etapa envolve a aplicação de algoritmos sobre os dados em busca de conhecimento implícito e útil. Algumas tarefas comumente utilizadas nesta etapa serão descritas a seguir [Goldschmidt 2015].

Clustering é uma tarefa que, segundo [Ali et al. 2020], é usada de forma intercambiável para explicar como os grupos de pesquisa podem coletar dados desagregados. São técnicas de classificação que agrupam padrões em classes relacionadas, dividindo objetos em *clusters* (agrupamentos) semelhantes [Dada et al. 2019].

A classificação, busca organizar em uma categoria dentre diversas pré-definidas. [Tan et al. 2009] definem a classificação como a tarefa de aprender uma função alfa f que mapeie cada conjunto de atributos x para um dos rótulos de classes y pré-determinados. Exemplos incluem classificadores de árvore de decisão, classificadores baseados em regras, redes neurais, máquinas de vetores de suporte (SVM) e classificadores *bayesianos*.

2.3. Trabalhos Correlatos

O trabalho de [Dada et al. 2019] revisa as abordagens de *Machine Learning* para a filtragem de SPAM, destacando que, apesar dos avanços significativos, um dos maiores desafios enfrentados é a capacidade dos modelos de acompanhar a rápida evolução das

técnicas utilizadas por *spammers* (pessoas que utilizam da técnica de enviar *e-mails* indesejados em massa). Para eles, o uso de abordagens híbridas, que combinam diferentes algoritmos de *Machine Learning*, pode ser uma solução promissora para aumentar a precisão na classificação de SPAM. No entanto, eles também apontam uma lacuna importante: a adaptação desses modelos a diferentes idiomas e contextos linguísticos ainda é limitada, o que afeta a eficácia dessas soluções em cenários específicos.

O estudo de [Kuchipudi et al. 2020], em complemento, analisa como a manipulação de palavras-chave em modelos de *e-mails* pode influenciar a classificação de mensagens como SPAM ou HAM. Usando o algoritmo *Naïve Bayes*, os autores investigam três técnicas de evasão: substituição de sinônimos HAM, injeção de palavras HAM e espaçamento de palavras SPAM. Eles mostram que, em 60% dos casos, é possível contornar os filtros ao aplicar uma dessas estratégias, expondo uma vulnerabilidade significativa em modelos tradicionais de classificação.

Adicionalmente, o estudo de [Yaseen et al. 2021] investiga modelos avançados na detecção de *e-mails* de SPAM, destacando a eficácia do modelo *Transformer BERT Base Cased*. Esse modelo se mostrou superior em comparação com abordagens anteriores, como o *BiLSTM*, devido à sua capacidade de considerar o contexto das palavras por meio de camadas de atenção. Os resultados indicam que esses modelos não apenas se ajustam bem a novos dados, mas também apresentam robustez, evitando problemas comuns como o sobreajuste. Além disso, a pesquisa sugere que técnicas semelhantes podem ser aplicadas a outros idiomas, ampliando o potencial da detecção em contextos diversos.

O artigo de [Jayapandian et al. 2023], apresenta um modelo de detecção de SPAM usando Regressão Logística, comparando seu desempenho com o algoritmo *Naïve Bayes*. Os resultados mostraram que a Regressão Logística alcançou uma precisão significativamente maior, variando entre 97.41% e 99.35% em diferentes conjuntos de dados, enquanto o *Naïve Bayes* obteve precisão entre 82.63% e 88.26%. O modelo proposto se destacou pela sua interpretabilidade, permitindo entender quais atributos dos *e-mails* têm maior influência na classificação, além de demonstrar boa adaptabilidade a novas estratégias de burlar filtros de SPAM.

3. Metodologia

O presente trabalho adota uma abordagem quantitativa experimental para desenvolver um classificador de *e-mails* utilizando os algoritmos de Regressão Logística e Naïve Bayes. Seguindo o processo do KDD, o desenvolvimento foi realizado com a linguagem de programação *Python*, amplamente utilizada para tarefas de aprendizado de máquina. Para isso, foram empregadas bibliotecas como *pandas*, *numpy* e *sklearn*, utilizadas respectivamente para manipulação de dados, operações matriciais, divisão do conjunto de dados e implementação dos algoritmos de *Machine Learning*.

A Figura 1, representa a metodologia aplicada no presente trabalho. Nota-se a semelhança com o KDD, processo apresentado previamente, que visa a descoberta de conhecimento a partir de um conjunto de dados. Por compartilhar o objetivo do trabalho, a metodologia abrange em sua grande maioria, as mesmas etapas. Porém, a documentação dos resultados será parte essencial do projeto, consolidando as descobertas e percepções gerados durante o processo de análise. Esse material servirá como insumo para a etapa final de verificação da satisfação dos resultados obtidos.

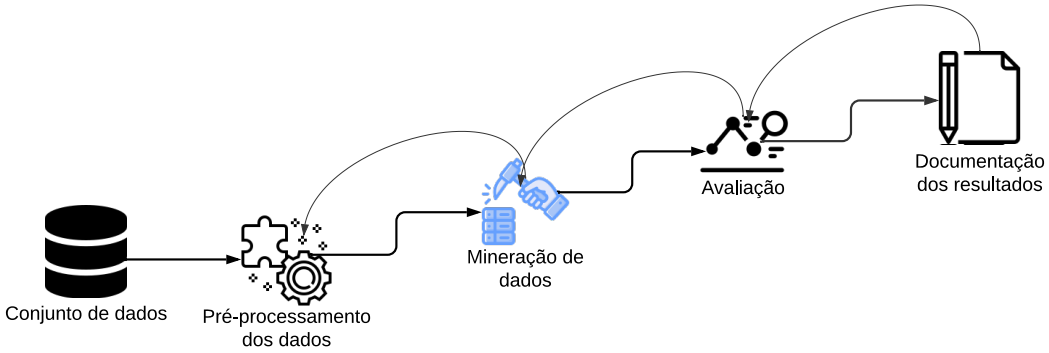

Figura 1. Metodologia. Fonte: Autor.

3.1. Seleção e Preparação dos Dados

Os conjuntos de dados utilizados são públicos, disponíveis na plataforma *Kaggle* e são de propriedade de [Almeida and Hidalgo 2012] e [Garnepudi 2019]. A partir da obtenção dos dados, é feita uma análise para remoção de *Stopwords*, que são palavras que aparecem com frequência em textos, mas carregam pouca informação. A remoção de *stopwords* pode aumentar a relação sinal-ruído em textos não estruturados e, assim, aumentar a significância estatística de termos que podem ser importantes para uma tarefa específica [Sarica and Luo 2021] .

Para a etapa de transformação de dados, é utilizado o processo de tokenização, que envolve a separação do texto em unidades significativas, conhecidas como *tokens*, servindo de entrada para o algoritmo de mineração de dados. Essa substituição de informações por símbolos identificadores, os *tokens*, permite a extração de atributos e termos do *e-mail*, sem considerar seu significado real [Dada et al. 2019] . Um exemplo de técnica de tokenização é o TF-IDF (Term Frequency - Inverse Document Frequency), que extrai informações com base na frequência do termo no documento e na coleção de dados, ajudando a destacar palavras que são importantes em um documento [Martins et al. 2020].

3.2. Desenvolvimento do modelo

O desenvolvimento do modelo de classificação foi conduzido de forma iterativa e exploratória, visando identificar a configuração mais eficaz para a detecção de *e-mails* SPAM. Esta fase da pesquisa foi dividida em duas etapas principais: implementação e otimização dos classificadores.

3.2.1. Implementação do Modelo de Regressão Logística

Regressão Logística é um método estatístico de classificação que, apesar do nome, não é utilizado para regressão, mas sim para problemas de classificação binária. Utiliza a função sigmoide (Equação 2) para transformar uma combinação linear das variáveis de entrada em uma probabilidade entre 0 e 1, onde x representa a soma do intercepto (b_0) com o produto de cada variável ($x_1, x_2, \ldots$) por seus respectivos coeficientes ($b_1, b_2, \ldots$), conforme Equação 1. Esses coeficientes e o intercepto são aprendidos pelo algoritmo durante o treinamento do modelo através de um processo de otimização (geralmente gra-

diente descendente), que busca minimizar o erro entre as previsões e os valores reais dos dados de treinamento [Jurafsky and Martin 2009].

$$x = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \tag{1}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

A regressão logística foi escolhida como algoritmo principal, devido à sua flexibilidade matemática e capacidade de produzir resultados interpretáveis e eficazes em problemas de classificação binária, conforme descrito por [Hastie et al. 2001]. Além disso, o modelo fornece probabilidades diretamente interpretáveis por meio da transformação $logit$, permitindo *feedback* detalhado para o usuário final, o que enriquece a análise preditiva. A implementação foi conduzida de forma progressiva, iniciando com uma configuração básica e adaptando-se com base nos resultados observados, buscando melhorar a precisão do modelo.

3.2.2. Implementação do *Naïve Bayes*

O algoritmo *Naïve Bayes* fundamenta-se no Teorema de Bayes e é amplamente utilizado em tarefas de classificação. Esse método baseia-se no cálculo de probabilidades condicionais e pode ser dividido em duas etapas principais:

1. **Cálculo das probabilidades condicionais** $P(C = c \mid R)$, onde C representa a classe e R o conjunto de atributos observados.
2. **Cálculo das probabilidades** $P(R \mid C = c)$, que representam a verossimilhança dos atributos R serem observados em uma determinada classe C.

O *Naïve Bayes* assume que os atributos são condicionalmente independentes entre si, dado a classe C, o que simplifica significativamente os cálculos e permite a aplicação eficiente do modelo em diversos cenários de classificação [Goldschmidt 2015].

A comparação entre o classificador Bayesiano e o algoritmo de Regressão Logística teve como objetivo identificar qual dos dois apresentaria melhor desempenho na classificação de e-mails, verificando se algum deles se destaca mais na detecção de SPAM e o outro na identificação de e-mails HAM.

3.2.3. Avaliação do Modelo

A avaliação do modelo foi conduzida de forma sistemática, investigando o impacto de diferentes configurações e parâmetros no desempenho do classificador. A validação do modelo foi realizada através de um processo iterativo de experimentação, a validação cruzada, que conforme apresentado por [James et al. 2013], a técnica envolve dividir repetidamente o conjunto de dados em partes distintas: um subconjunto usado para treinar o modelo e outro para avaliar seu desempenho. A principal vantagem dessa abordagem é fornecer uma estimativa mais realista do erro de generalização, evitando o otimismo excessivo que pode ocorrer ao calcular apenas o erro de treinamento.

Foram utilizadas as métricas de precisão (*precision*), que indica a proporção de predições positivas corretas entre todas as predições positivas feitas pelo modelo; revocação (recall), que representa a proporção de casos positivos corretamente identificados; e F1-score, que fornece a média harmônica entre precisão e revocação, oferecendo uma medida balanceada do desempenho do modelo.

Adicionalmente, utilizou-se a matriz de confusão para uma análise detalhada dos tipos de erros cometidos pelos classificadores, permitindo visualizar explicitamente os falsos positivos (mensagens legítimas classificadas incorretamente como SPAM) e falsos negativos (SPAM não detectados). Esta abordagem múltipla de avaliação permitiu uma análise completa do desempenho dos modelos, possibilitando identificar suas características específicas e adequação para diferentes cenários de aplicação.

4. Resultados Experimentais

Esta seção apresenta os resultados experimentais obtidos através da aplicação de dois modelos de classificação - Regressão Logística e *Naïve Bayes* - em dois conjuntos de dados distintos para a classificação de SPAM. Os experimentos foram conduzidos utilizando métricas padrão de avaliação: precisão, *recall*, *f1-score* e acurácia.

4.1. Análise Comparativa dos Modelos

O primeiro conjunto de dados (A) apresentou uma distribuição de 300 mensagens SPAM e 735 HAM, totalizando 1035 amostras. Neste conjunto, ambos os modelos alcançaram uma acurácia global de 0.96, porém com características distintas em sua performance.

A Regressão Logística demonstrou excelente capacidade em identificar mensagens SPAM, alcançando recall perfeito (1.00) para esta classe, indicando que nenhuma mensagem SPAM passou despercebida pelo modelo. Para mensagens HAM, o modelo apresentou precisão perfeita (1.00) e recall de 0.95, resultando em um f1-score de 0.97. A matriz de confusão revelou que o modelo não gerou falsos positivos, embora tenha produzido 40 falsos negativos. A Tabela 1, com as métricas, pode ser encontrada a seguir.

Tabela 1. Desempenho da Regressão Logística com o conjunto de dados A

Classe	Precision	Recall	F1-Score	Support
HAM	1.00	0.95	0.97	735
SPAM	0.88	1.00	0.94	300
Accuracy	-	-	0.96	1035
Macro Avg	0.94	0.97	0.95	1035
Weighted Avg	0.97	0.96	0.96	1035

O *Naïve Bayes*, por sua vez, apresentou um comportamento de minimização de falsos positivos. Conforme a Tabela 2, o modelo alcançou recall perfeito (1.00) para mensagens HAM, com precisão de 0.95. Para mensagens SPAM, obteve alta precisão (0.99) mas recall menor (0.87). A matriz de confusão mostrou apenas 3 falsos positivos, porém 38 falsos negativos, evidenciando uma tendência do modelo em minimizar a classificação incorreta de mensagens legítimas como SPAM.

Tabela 2. Desempenho do *Naïve Bayes* com o conjunto de dados A

Classe	Precision	Recall	F1-Score	Support
HAM	0.95	1.00	0.97	735
SPAM	0.99	0.87	0.93	300
Accuracy	-	-	0.96	1035
Macro Avg	0.97	0.93	0.95	1035
Weighted Avg	0.96	0.96	0.96	1035

O segundo conjunto de dados (B), composto por 149 mensagens SPAM e 966 HAM (total de 1115 amostras), permitiu avaliar a robustez dos modelos em um cenário mais desbalanceado.

A Regressão Logística manteve performance consistente, com acurácia global de 0.98. De acordo com a Tabela 3, o modelo apresentou métricas equilibradas para ambas as classes, com precisão e recall próximos a 0.99 para mensagens HAM e 0.94 e 0.91 respectivamente para SPAM. A matriz de confusão registrou 8 falsos positivos e 13 falsos negativos.

Tabela 3. Desempenho da Regressão Logística com o conjunto de dados B

Classe	Precision	Recall	F1-Score	Support
HAM	0.99	0.99	0.99	966
SPAM	0.94	0.91	0.93	149
Accuracy	-	-	0.98	1115
Macro Avg	0.97	0.95	0.96	1115
Weighted Avg	0.98	0.98	0.98	1115

Na Tabela 4, é possível perceber que o *Naïve Bayes* alcançou acurácia global de 0.97, mantendo sua preferência por minimização de falsos positivos. O modelo obteve recall perfeito (1.00) para HAM e precisão de 0.97, enquanto para SPAM apresentou alta precisão (0.99) mas recall menor (0.81). A matriz de confusão revelou apenas 1 falso positivo, contrastando com 28 falsos negativos.

Tabela 4. Desempenho do *Naïve Bayes* com o conjunto de dados B

Classe	Precision	Recall	F1-Score	Support
HAM	0.97	1.00	0.99	966
SPAM	0.99	0.81	0.89	149
Accuracy	-	-	0.97	1115
Macro Avg	0.98	0.91	0.94	1115
Weighted Avg	0.97	0.97	0.97	1115

4.2. Discussão dos Resultados

Os resultados indicam que ambos os modelos são muito eficazes na classificação de SPAM, com acurácias acima de 0.96 em ambos os conjuntos de dados. Porém, suas diferenças os tornam mais adequados para cenários distintos.

A Regressão Logística apresenta um equilíbrio maior entre falsos positivos e falsos negativos, sendo mais adequada para contextos onde é aceitável um balanço entre os tipos de erro. O *Naïve Bayes*, por sua vez, demonstra maior precisão na redução de falsos positivos, embora isso resulte em um maior número de falsos negativos, tornando-o mais apropriado para cenários onde o custo de classificar incorretamente uma mensagem legítima como SPAM é significativamente alto. Esta diferença de comportamento se manteve consistente nos dois conjuntos de dados, sugerindo que estas, até o presente momento avaliadas, são características intrínsecas dos modelos e não artefatos específicos de um conjunto de dados particular.

Com base nos resultados obtidos em nossos experimentos, observamos algumas diferenças significativas em relação aos trabalhos correlatos. Diferentemente dos resultados apresentados por [Jayapandian et al. 2023], onde a Regressão Logística demonstrou superioridade significativa sobre o Naive Bayes (99.35% versus 88.26%), nossos experimentos mostraram um desempenho mais equilibrado entre os dois modelos. Em nossos testes, ambos os classificadores alcançaram acurácia superior a 96% nos dois conjuntos de dados avaliados. Esta discrepância, pode ser atribuída às diferentes características dos conjuntos de dados utilizados e às técnicas de pré-processamento empregadas.

5. Conclusões e Trabalhos Futuros

Este estudo apresentou o desenvolvimento e a avaliação de modelos de *Machine Learning* para a classificação de *e-mails* como SPAM ou HAM. A partir da aplicação de Regressão Logística e Naïve Bayes, observou-se que ambos os modelos apresentaram alto desempenho, com acurácias superiores a 96% nos conjuntos de dados analisados. A Regressão Logística demonstrou maior equilíbrio entre falsos positivos e falsos negativos, enquanto o Naive Bayes priorizou a minimização de falsos positivos, resultando em uma taxa mais elevada de falsos negativos.

Os resultados indicam que o modelo depende do contexto de aplicação. Para campanhas de marketing, onde a minimização de falsos positivos é essencial para evitar bloqueios de provedores, o Naïve Bayes pode ser mais apropriado. Já em cenários onde se busca maior recall para SPAMs, a Regressão Logística se mostra uma alternativa viável.

Em trabalhos futuros, pretende-se realizar testes com datasets mais amplos e diversificados, de forma a avaliar o desempenho dos modelos em diferentes contextos. Além disso, ajustes nos hiperparâmetros e a utilização de outra técnica de *tokenização*.

Referências

Ali, A., Bin Faheem, Z., Waseem, M., Draz, U., Safdar, Z., Hussain, S., and Yaseen, S. (2020). Systematic review: A state of art ml based clustering algorithms for data mining. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–6.

Almeida, T. and Hidalgo, J. (2012). Sms spam collection dataset. Último acesso em: 10/02/2025.

Cappy, P. (2024). Email sending reputation: How does domain reputation work? Último acesso em: 26/10/2024.

Dada, E. G., Bassi, J. S., Chiroma, H., Adetunmbi, A. O., Ajibuwa, O. E., et al. (2019). Machine learning for email spam filtering: review, approaches and open research problems. *Heliyon*, 5(6).

Dossetto, F. (2022). Domain reputation, explained. Último acesso em: 03/01/2025.

Garnepudi, V. (2019). Spam mails dataset. Último acesso em: 10/02/2025.

Goldschmidt, R. (2015). *Data Mining*. GEN LTC, Rio de Janeiro, RJ, BRA, 2ª edition.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Jayapandian, N. et al. (2023). Machine learning based spam e-mail detection using logistic regression algorithm. In *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)*, pages 1–6. IEEE.

Jurafsky, D. and Martin, J. H. (2009). *Speech and Language Processing (2nd Edition)*. Prentice-Hall, Inc., USA.

Kosinski, M. (2024). What is phishing? Último acesso em: 27/10/2024.

Kuchipudi, B., Nannapaneni, R. T., and Liao, Q. (2020). Adversarial machine learning for spam filters. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*, pages 1–6.

Mariano, D. C. B., Marques, L. T., and Silva, M. S. (2021). *Data Mining*. SAGAH, Porto Alegre, RS, BRA.

Martins, J. S., Lenz, M. L., Silva, M. B. F. d., Oliveira, R. A. d., Pichetti, R. F., Mariano, D. C. B., Martins, J. V., Rodrigues, S. M. A. F., and Bezerra, W. R. (2020). *Processamentos de linguagem natural.* SAGAH, Porto Alegre.

Pronnus (2024). Segurança digital: Você sabe a diferença entre phishing e spoofing? Último acesso em: 27/10/2024.

Sarica, S. and Luo, J. (2021). Stopwords in technical language processing. *PLOS ONE*, 16(8):1–13.

Sherwin, R. (2023). Report spam, misclassified, viral email messages. Último acesso em: 27/10/2024.

Statista (2024a). Daily number of emails sent worldwide as of april 2024 by country. Último acesso em: 26/10/2024.

Statista (2024b). Daily number of spam emails sent worldwide as of august 2024, by country. Último acesso em: 26/10/2024.

Tan, P.-N., Steinbach, M., and Kumar, V. (2009). *Introdução ao Data Mining - Mineração de Dados*. Editora Ciência Moderna Ltda., Rio de Janeiro, RJ, 1ª edition.

Yaseen, Q. et al. (2021). Spam email detection using deep learning techniques. *Procedia Computer Science*, 184:853–858.