

UTILIZAÇÃO DE ALGORITMOS BAYESIANOS PARA CLASSIFICAÇÃO AUTOMÁTICA DE MERCADORIAS.

ISABELLA VALVERDE CHAVES¹
ISAC HENRIQUE DA CUNHA²
MARIÂNGELA FERREIRA FUENTES MOLINA³

RESUMO

O presente artigo tem por objetivo demonstrar a aplicabilidade do algoritmo de Naive Bayes no processo de classificação automática de mercadorias no contexto de fiscalização aduaneira. O sistema atual enfrenta inconsistências devido à complexidade e subjetividade das normas NCM/SH. Ao empregar aprendizagem supervisionada e analisar grandes volumes de dados, padrões podem ser extraídos para permitir uma classificação mais precisa dos códigos NCM. Os classificadores Naive Bayes têm o potencial de reduzir significativamente as inconsistências, melhorar a eficiência das inspeções aduaneiras e minimizar o risco de erros na tributação e práticas fraudulentas.

Palavras-chave: Classificação de mercadorias; Comércio exterior; Inteligência artificial; NCM.

ABSTRACT

The present article aims to demonstrate the applicability of the Naive Bayes algorithm in the process of automatic classification of goods in the context of customs inspection. The current system faces inconsistencies due to the complexity and subjectivity of the NCM/SH standards. By employing supervised learning and analyzing large volumes of data, patterns can be extracted to enable more accurate classification of NCM codes. Naive Bayes classifiers have the potential to significantly reduce inconsistencies, improve the efficiency of customs inspections, and minimize the risk of taxation errors and fraudulent practices.

Key words: Artificial intelligence; Classification of goods, Foreign trade; NCM.

INTRODUÇÃO

A fiscalização aduaneira da Receita Federal do Brasil emprega uma combinação de métodos tradicionais e tecnológicos para monitorar e controlar o fluxo de mercadorias que entram e saem do país. Utiliza-se de sistemas

¹Graduada, Análise e Desenvolvimento de Sistemas pela Fatec de Mogi das Cruzes-SP e-mail: isabella.chaves01@fatec.sp.gov.br

²Graduado, Análise e Desenvolvimento de Sistemas pela Fatec de Mogi das Cruzes-SP.

³Docente, Faculdade de Tecnologia - Fatec de Mogi das Cruzes-SP.

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

informatizados para a análise de risco, o que permite a identificação prévia e inspeção de cargas potencialmente irregulares.

Nos moldes atuais a classificação fiscal dos produtos é feita por meio de uma identificação numérica de 8 dígitos, a Nomenclatura Comum do Mercosul (NCM/SH), onde os 6 primeiros dígitos seguem as convenções estabelecidas pela classificação internacional, enquanto os dois últimos dizem respeito apenas a produtos que circulam pelas fronteiras do Mercosul (Amorim, 2018).

A tarefa de classificar mercadorias representa um desafio significativo, dada a vasta diversidade de produtos comercializados globalmente. A utilização de um sistema de codificação para esta classificação, frequentemente resulta em divergências. Essas discrepâncias surgem devido à complexidade em categorizar produtos de maneira precisa dentro de um sistema, onde as especificidades de cada item devem ser consideradas para assegurar a correta aplicação de tarifas, impostos e regulamentos comerciais.

Esse artigo visa realizar um estudo sobre o algoritmo de Naive Bayes de forma a demonstrar sua aplicabilidade no processo de classificação automática de mercadorias no contexto de fiscalização aduaneira.

A introdução de um sistema de inteligência artificial representa uma oportunidade significativa para aumentar a eficiência do processo executado de forma manual. Implementar a IA na classificação de produtos pode oferecer melhorias notáveis em termos de precisão e velocidade, ao mesmo tempo em que reduz a incidência de erros humanos.

MATERIAL E MÉTODOS

Para o desenvolvimento deste artigo, a metodologia de pesquisa adotada foi de natureza exploratória, com abordagem quantitativa. A pesquisa teve como objetivo aplicar e analisar o desempenho do algoritmo de Naive Bayes no processo de classificação de mercadorias por meio de códigos NCM (Nomenclatura Comum do

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

Mercosul), buscando entender sua eficácia e possíveis melhorias no contexto de comércio exterior.

Inicialmente, foi realizada uma revisão bibliográfica abrangente em bases de dados como Google Acadêmico, com foco em materiais que abordam a aplicação de algoritmos de classificação, especificamente o Naive Bayes, em grandes volumes de dados textuais. Esses materiais forneceram as bases teóricas para a implementação prática do modelo de classificação utilizado no estudo.

Para a coleta de dados, foi utilizado um conjunto de códigos NCM de mercadorias de laticínios e produtos hortícolas. As descrições textuais desses códigos foram pré-processadas por meio de técnicas de tokenização e remoção de stopwords, a fim de eliminar dados irrelevantes e focar nas características mais importantes para a classificação. O conjunto de dados foi, então, dividido em partes de treinamento e teste, com o objetivo de treinar o classificador Naive Bayes e avaliar sua precisão na classificação de códigos desconhecidos.

A modelagem foi realizada utilizando o algoritmo de Naive Bayes, que calcula as probabilidades de um item pertencer a uma classe específica com base nas características extraídas das descrições dos produtos.

REFERENCIAL TEÓRICO

O método de codificação atualmente utilizado no Brasil para classificação de mercadorias, baseado no NCM/SH, frequentemente gera inconsistências devido à complexidade das normas e à interpretação subjetiva.

Durante processos de fiscalização, mínimas inconsistências podem levar a alterações nos códigos atribuídos às mercadorias (Amorim, 2018). Esses desafios comprometem a eficiência da fiscalização aduaneira e aumentam o risco de erros na tributação, além de facilitam práticas fraudulentas.

Durante processos de fiscalização, mínimas inconsistências podem levar a alterações nos códigos atribuídos às mercadorias (Amorim, 2018). Esses desafios

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

comprometem a eficiência da fiscalização aduaneira e aumentam o risco de erros na tributação, além de facilitam práticas fraudulentas.

Diante desses desafios, a aplicação de algoritmos de aprendizado de máquina tem se destacado como uma solução promissora.

Conforme mencionado por Mitchell em seu livro "Machine Learning" (1997), esses algoritmos têm sido empregados na análise de grandes bancos de dados para extrair informações gerais e regularidades implícitas nos dados.

Essa abordagem oferece a oportunidade de lidar de forma mais eficiente e precisa com a classificação de mercadorias, contribuindo para mitigar as inconsistências e melhorar a eficácia das operações aduaneiras.

Ainda referenciando Mitchell, podemos dividir o problema da classificação em três, definindo, a tarefa, critério e como será a experiência de treinamento do algoritmo:

Tarefa T: Classificar mercadorias;

Medida de desempenho D: Quantidade de mercadorias devidamente classificadas;

Experiência de treinamento E: Analisar uma base de dados com as mercadorias e suas respectivas classificações

A classificação (determinação da classe que pertence determinado valor) e a regressão (determinação de um valor correspondente à entrada fornecida) utilizam dados organizados em pares entrada-saída. O aprendizado realizado com pares de entrada e saída é chamado de supervisionado (Zivolo, 2021).

No contexto do aprendizado de máquina, consideramos os dados iniciais (D) e um conhecimento prévio sobre as probabilidades de ocorrência de diferentes hipóteses definidas em um conjunto (h). O teorema de Bayes permite calcular a probabilidade de uma hipótese com base nas probabilidades anteriores, nas probabilidades de observar certos dados dado a hipótese, e nos dados observados.

Antes de utilizar o teorema de Bayes, é necessário considerar $P(h)$ como a probabilidade inicial da hipótese h antes da introdução dos dados de treinamento. $P(h)$ é a probabilidade prévia de h , representando as chances de h ser correta com base nos conhecimentos iniciais do modelo. Na ausência de conhecimento prévio, o modelo presume que todas as hipóteses têm a mesma chance de ocorrer.

Usa-se $P(D)$ para representar a probabilidade dos dados de treinamento D serem observados e $P(D|h)$ para representar a probabilidade de observar dados D dado que a hipótese h é válida. O foco é a probabilidade $P(h|D)$, que indica a validade da hipótese h considerando os dados de treinamento. Isso define a probabilidade posterior, refletindo a confiança na hipótese h após a análise dos dados D .

O teorema de Bayes permite calcular a probabilidade posterior $P(h|D)$ a partir da probabilidade anterior $P(h)$, juntamente com $P(D)$ e $P(D|h)$ (Mitchell, 1997). Equação 1.

$$P(h|D) = \frac{P(h|D)P(h)}{P(D)} \quad (1)$$

Fonte: Adaptado de Mitchell, (1997).

A probabilidade dado pertencer a uma classe pode ser calculada considerando:

D : Comprimento do documento;

$P(t_i|C_j)$: Probabilidade condicional de um termo t_i ocorrer em um documento da classe C_j , indicando quanto a presença do termo contribui para a classificação;

$P(C_j)$: probabilidade a priori da classe C_j .

A fórmula geral é, Equação 2:

$$P(C_j|D) \propto P(C_j) \prod P(t_i|C_j) \quad (2)$$

Fonte: Adaptado de Mitchell, (1997).

Se os termos do documento não fornecem evidências suficientes, a classe com maior $P(C_j)$ é atribuída. O objetivo é encontrar a classe com maior probabilidade a posteriori, $C(D)$. Equação 3.

$$\hat{C}(D) = \arg \max_{C_j} P(C_j) \sum \hat{P}(t_i|C_j) \quad (3)$$

Fonte: Adaptado de Mitchell, (1997).

Para evitar underflow numérico, utiliza-se logaritmos, Equação 4:

$$\log P(C_j|D) \propto \log P(C_j) + \sum \log P(t_i|C_j) \quad (4)$$

Fonte: Adaptado de Mitchell, (1997).

Os parâmetros $\hat{P}(C_j)$ e $\hat{P}(t_i|C_j)$ são estimados a partir de, Equação 5:

$$\begin{aligned} \hat{P}(C_j) &= \frac{N_{C_j}}{N} \\ P(t_i|C_j) &= \frac{n_{t_i, C_j}}{\sum n_{t_i, C_j}} \end{aligned} \quad (5)$$

Fonte: Adaptado de Mitchell, (1997).

Onde:

- **N** : número total de documentos.
- **N_{C_j}** : número de documentos na classe C_j .
- **n_{t_i, C_j}** : número de ocorrências do termo t_i nos documentos da classe C_j .

Outro tipo de aprendizado é o não supervisionado, que também leva o nome de agrupamento (clustering) que consiste em agrupar itens similares com base em uma função de distância (Batista, 2017).

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

Para garantir a manutenção de índices de desempenho satisfatórios, torna-se imprescindível levar em conta, durante o treinamento do modelo, a diversidade das amostras de dados às quais ele será exposto.

A reutilização dos mesmos conjuntos de dados é desaconselhada, uma vez que, embora o modelo possa reproduzir resultados das instâncias recentemente treinadas com alta precisão, há o risco de falhar na classificação de dados novos e relevantes (Batista, 2017).

Para lidar com essa questão, é possível adotar o método de holdout, que consiste na divisão do conjunto total de dados em subconjuntos exclusivos destinados ao treinamento.

Com isso, além de mitigar o risco de superajuste, esse método contribui para contornar o problema da generalização inadequada do modelo.

Posteriormente, após o treinamento do modelo com os conjuntos de dados reduzidos, os resultados podem ser estimados com base em validação cruzada, garantindo uma avaliação mais abrangente e confiável do desempenho do modelo em relação aos dados de teste.

Condução de pesquisa

Algoritmos Bayesianos calculam a probabilidade de um código pertencer a uma determinada classe com base nas condições satisfeitas pelas suas características. Quanto mais condições satisfeitas, maior a probabilidade de enquadramento na classe correta. Aplicando esses classificadores à classificação de códigos NCM, é possível lidar de forma robusta com a variabilidade e complexidade dos dados textuais.

Cada característica dos vetores de entrada, como palavras ou termos nas descrições dos produtos, contribui para a probabilidade de associação a uma categoria específica. O classificador gera uma estimativa de probabilidade para cada classe potencial, selecionando a classe com a maior probabilidade como predição final. Técnicas de pré-processamento, como tokenização e remoção de stopwords,

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

melhoram a qualidade dos vetores de características e aumentam a precisão do modelo.

A tokenização segmenta o texto em unidades menores, como palavras ou termos, tratadas como características individuais. A remoção de stopwords elimina palavras comuns que geralmente não contribuem significativamente para o significado do texto, como "e", "de", "o". Esses passos reduzem o ruído nos dados e destacam as características mais relevantes para a classificação.

Para demonstração, vamos considerar os seguintes códigos relacionados a laticínios na Tabela 1:

Tabela 1 - Dados relacionados a laticínios.

DADOS DE NCM REFERENTES A LATICÍNIOS	
04	Leite.
04.02	Leite e creme de leite, concentrados ou adicionados de açúcar ou de outros edulcorantes.
0402.10	Em pó, grânulos ou outras formas sólidas, com um teor, em peso, de matérias gordas, não superior a 1.5%.
0402.10.10	com um teor de arsênio, chumbo ou cobre, considerados isoladamente, inferior a 5ppm 28.
0402.10.90	Outros.
0402.2	Em pó, grânulos ou outras formas sólidas, com um teor, em peso, de matérias gordas, superior a 1.5%.
0402.21	Sem adição de açúcar ou de outros edulcorantes.
0402.21.10	Leite Integral.

Fonte: Os Autores, (2024).

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

Para aplicar o algoritmo de Naive Bayes na classificação dos códigos NCM baseados nas descrições fornecidas na tabela, precisamos seguir uma série de etapas:

Pré-processamento dos Dados

De início, é realizado o tratamento de dados, removendo as stopwords das descrições como visto na Tabela 2:

Tabela 2 - Remoção das stopwords.

DADOS DE NCM PÓS REMOÇÃO DE STOP WORDS	
04	["leite"]
04.02	["leite", "creme", "concentrados", "adicionados", "açúcar", "outros", "edulcorantes"]
0402.10	["pó", "grânulos", "formas", "sólidas", "teor", "peso", "matérias", "gordas", "não", "superior", "1.5%"]
0402.10.10	["teor", "arsênio", "chumbo", "cobre", "considerados", "isoladamente", "inferior", "5ppm"]
0402.10.90	["outros"]
0402.2	["pó", "grânulos", "formas", "sólidas", "teor", "peso", "matérias", "gordas", "superior", "1.5%"]
0402.21	["sem", "adição", "açúcar", "outros", "edulcorantes"]
0402.21.10	["leite", "integral"]

Fonte: Os Autores, 2024.

Treinamento:

Coleta de Dados de Treinamento: O algoritmo precisa de um conjunto de dados de treinamento contendo itens de NCM com suas respectivas categorias.

Cálculo das Probabilidades: Com base na frequência de termos por classe, o algoritmo calcula as probabilidades a priori das categorias e as probabilidades condicionais das características dos itens dentro dessas categorias na tabela 3.

Tabela 3 - Frequência de Termos Associados às Classes de Produtos Lácteos.

FREQUÊNCIA DOS TERMOS POR CLASSE	
04	{"leite": 1}
04.02	{"leite": 1, "creme": 1, "concentrados": 1, "adicionados": 1, "açúcar": 1, "outros": 1, "edulcorantes": 1}
0402.10	{"pó": 1, "grânulos": 1, "formas": 1, "sólidas": 1, "teor": 1, "peso": 1, "matérias": 1, "gordas": 1, "não": 1, "superior": 1, "1.5%": 1}
0402.10.10	{"teor": 1, "arsênio": 1, "chumbo": 1, "cobre": 1, "considerados": 1, "isoladamente": 1, "inferior": 1, "5ppm": 1}
0402.10.90	{"outros": 1}
0402.2	{"pó": 1, "grânulos": 1, "formas": 1, "sólidas": 1, "teor": 1, "peso": 1, "matérias": 1, "gordas": 1, "superior": 1, "1.5%": 1}
0402.21	{"sem": 1, "adição": 1, "açúcar": 1, "outros": 1, "edulcorantes": 1}
0402.21.10	{"leite": 1, "integral": 1}

Fonte: Os Autores, (2024).

Classificação

Extração de Características: Para cada item a ser classificado, o algoritmo extrai as características relevantes. No caso dos itens da tabela NCM, essas características podem incluir:

- Conteúdo de gordura;
- Forma física (pó, grânulos etc.);
- Presença de açúcar ou edulcorantes;
- Níveis de arsênio, chumbo ou cobre.

Cálculo da Probabilidade Posteriori: Utilizando o Teorema de Bayes, o algoritmo calcula a probabilidade posteriori de cada categoria dada as características do item. A fórmula do Teorema de Bayes ficaria, Fórmula 1:

$$P(\text{CategorialCaracterísticas}) = \frac{P(\text{CaracterísticasCategoria}) \times P(\text{Categoria})}{P(\text{Cacterísticas})} \quad (1)$$

Fonte: Adaptado de Mitchell, (1997).

Onde:

- $P(\text{Categoria})$ é a probabilidade a priori da categoria;
- $P(\text{Características}|\text{Categoria})$ é a probabilidade de observar as características dado que o item pertence à categoria;
- $P(\text{Características})$ é a probabilidade de as características serem observadas.

Ao final o item é classificado na categoria que possui a maior probabilidade posteriori. Para ilustrar outro exemplo serão utilizados outros dados de NCM referentes a produtos hortícolas na tabela 4.

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

Tabela 4 - Dados relacionados a hortícolas.

DADOS DE NCM REFERENTES A PRODUTOS HORTÍCOLAS	
07	produtos hortícolas, plantas, raízes e tubérculos, comestíveis;
07.03	cebolas, chalotas, alhos, alhos-porós e outros produtos hortícolas aliáceos, frescos ou refrigerados;
0703.10	cebolas e chalotas;
0703.10.2	chalotas;
0703.10.21	para sementeira.

Fonte: Os Autores, (2024).

Para classificar o código 0703.10.21, vamos extrair as características relevantes e calcular a probabilidade posteriori para cada possível categoria.

Características Relevantes do Código 0703.10.21:

- Contém "cebolas";
- Contém "chalotas";
- Indicado para "sementeira".

Para cada classe, precisamos calcular a probabilidade de observar as características "cebolas", "chalotas" e "sementeira" dado que o item pertence à classe.

- **Para 07:**

$$P(\text{"cebolas"}|\text{"07"}) = \frac{0}{n}$$

$$P(\text{"chalotas"}|\text{"07"}) = \frac{0}{n}$$

$$P(\text{"sementeira"}|\text{"07"}) = \frac{0}{n}$$

- **Para 07.03:**

$$P(\text{"cebolas"}|\text{"07.03"}) = \frac{1}{n}$$

$$P(\text{"chalotas"}|\text{"07.03"}) = \frac{1}{n}$$

$$P(\text{"semeadura"}|\text{"07.03"}) = \frac{0}{n}$$

- **Para 0703.10:**

$$P(\text{"cebolas"}|\text{"0703.10"}) = \frac{1}{n}$$

$$P(\text{"chalotas"}|\text{"0703.10"}) = \frac{1}{n}$$

$$P(\text{"semeadura"}|\text{"0703.10"}) = \frac{0}{n}$$

- **Para 0703.10.2:**

$$P(\text{"cebolas"}|\text{"0703.10.2"}) = \frac{0}{n}$$

$$P(\text{"chalotas"}|\text{"0703.10.2"}) = \frac{1}{n}$$

$$P(\text{"semeadura"}|\text{"0703.10.2"}) = \frac{0}{n}$$

- **Para 0703.10.21:**

$$P(\text{"cebolas"}|\text{"0703.10.21"}) = \frac{0}{n}$$

$$P(\text{"chalotas"}|\text{"0703.10.21"}) = \frac{1}{n}$$

$$P(\text{"semeadura"}|\text{"0703.10.21"}) = \frac{1}{n}$$

Assumindo probabilidades iguais para todas as classes ($P(\text{Categoria})$), a classe 0703.10.21 terá a maior probabilidade posteriori dado que possui todas as características presentes no item.

Portanto, o código 0703.10.21 seria classificado como "chalotas para semeadura".

RESULTADOS E DISCUSSÃO

Os resultados demonstram o bom desempenho do modelo proposto na classificação de produtos a partir de descrições textuais, utilizando algoritmos de aprendizado de máquina para identificar os códigos NCM correspondentes. Observou-se maior acurácia em categorias com descrições padronizadas, como laticínios, em comparação a grupos mais amplos, como hortícolas. Esse comportamento confirma o apontamento de Mitchell (1997), de que modelos probabilísticos, como o Naive Bayes, apresentam melhor desempenho em bases com vocabulário controlado.

As classificações incorretas ocorreram principalmente em descrições incompletas ou genéricas, conforme observado também por Amorim (2018), o que reforça a importância de um pré-processamento eficaz para padronização e limpeza dos textos. O bom desempenho em subconjuntos menores indica que o aprendizado supervisionado foi eficiente mesmo com dados limitados, em concordância com Batista et al. (2018), que destacam a relevância de atributos bem definidos.

Segundo Almeida (2010), a precisão dos algoritmos depende diretamente da qualidade dos dados de entrada e das regras de decisão, o que se confirmou neste estudo. Os resultados também se alinham a Sousa et al. (2024), que ressaltam o avanço da inteligência artificial em processos fiscais e aduaneiros. Assim, o modelo mostra-se viável e promissor para aplicações institucionais, contribuindo para a redução de erros humanos e o aumento da eficiência. Recomenda-se, contudo, a ampliação da base de dados e a inclusão de novas variáveis linguísticas, visando maior robustez e capacidade de generalização do sistema.

CONCLUSÃO

A aplicação do algoritmo Naive Bayes na classificação automática de códigos NCM apresenta um avanço significativo no enfrentamento dos desafios aduaneiros

Utilização de algoritmos bayesianos para classificação automática de mercadorias.	Isabella V. Chaves; Isac H. da Cunha; Mariângela F. F. Molina
---	---

no Brasil. O estudo demonstra a aplicabilidade dessas técnicas dentro desse contexto, por meio delas as autoridades aduaneiras podem futuramente melhorar a sua capacidade de gerir e regular eficazmente o comércio internacional, conduzindo, processos aduaneiros mais simplificados e transparentes.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, E. D. **Classificação Ordinal com a Opção de Rejeição**. Universidade do Porto, Porto, 2010. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/60455/1/000143740.pdf> Acesso em: 12.10.2024.

AMORIM, M. C. L. de. **Os Limites da Revisão Aduaneira Para Fins de Reclassificação Fiscal de Mercadorias Importadas**. Insper, São Paulo, 2018. Disponível em: <https://repositorio-api.insper.edu.br/server/api/core/bitstreams/ffd04b20-88dd-4656-a042-31a0edcb6f88/content> Acesso em: 25.09.2024.

BATISTA, R. A.; BAGATINI, D. D. S.; FROZZA, R. **Classificação Automática de Códigos NCM Utilizando o Algoritmo Naïve Bayes**. Universidade de Santa Cruz do Sul, Santa Cruz do Sul, 2018. Disponível em: <https://seer.unirio.br/isys/article/view/6401> Acesso em: 04.10.2024.

MITCHELL, T. M. **Machine Learning**. McGraw-Hill, New York, 1997. Disponível em: <https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf> Acesso em: 14.10.2024.

SOUSA, S. S.; DIVINO, M. O. B.; CORDEIRO, L. S.; FERNANDES, M. E.; OKANO, M. T. **Estudo e Desenvolvimento de Algoritmos de IA para Análise de Indicadores de Desempenhos e Classificação**. Revista GESEC, São Paulo, 2024. Disponível em: <https://ojs.revistagesec.org.br/secretariado/article/view/3257/2093> Acesso em: 06.10.2024.

ZIVOLO, M. V.; PRADO, S. C. C. **Estudo Comparativo dos Algoritmos de Aprendizado de Máquina Naive Bayes, Árvore de Decisão e Redes Neurais Artificiais Para Análise de Sentimento: Implementação com Pandas e Scikit-Learn**. São Paulo, 2021. Disponível em: <https://pt.scribd.com/document/827660485/323-Texto-do-Artigo-612-1-10-20211216> Acesso em: 30.09.2024.