

AVALIAÇÃO DE MODELOS BASEADOS EM APRENDIZAGEM DE MÁQUINA SUPERVISIONADO PARA TAREFAS DE RECONHECIMENTO DE EMOÇÕES EM TEXTOS CURTOS.

Douglas Maia dos Santos ; <https://orcid.org/0000-0003-2880-0571>
Universidade Federal do Pará

Fabiola Araújo ; <https://orcid.org/0000-0002-8485-5799>
Universidade Federal do Pará.



EVALUATION OF SUPERVISED MACHINE LEARNING-BASED MODELS FOR EMOTION RECOGNITION TASKS IN SHORT TEXTS.

ABSTRACT

The technological advances of the 4.0 revolution made it possible to create devices and software that aim to improve and/or facilitate the lives of individuals, especially in terms of human-computer interaction, a concept that is immersed in everyday life and has a strong link with the quality of life. Therefore, this work aims to present and evaluate the performance of models based on supervised Machine Learning for tasks of recognition of emotions joy, sadness and surprise in short texts in the Portuguese language collected in comments of empathic games. The implementations were carried out with the aid of the *Scikit-Learn* library, using the *Support Vector Machine* (SVM), *Naive Bayes* (NB), *K-Nearest Neighbors* (KNN), and *Ensemble methods*, which were trained using a predefined database classified. The development of the application is divided into 6 stages of evaluation, where in each stage it was sought to apply a different technique of data pre-processing, thus making it possible to ascertain the impact of each technique on the performance of the models. Results show that the SVM model reached 71% accuracy in scenarios with a large presence of preprocessing functions, while the NB and Voting models act respectively with 77% accuracy in scenarios with little preprocessing robustness and 77% accuracy in both scenarios.

Keywords: Machine Learning, Emotion Mining, Empathic Games, Short-texts.

AVALIAÇÃO DE MODELOS BASEADOS EM APRENDIZAGEM DE MÁQUINA SUPERVISIONADO PARA TAREFAS DE RECONHECIMENTO DE EMOÇÕES EM TEXTOS CURTOS.

RESUMO

Os avanços tecnológicos da revolução 4.0 possibilitaram a criação de dispositivos e softwares que visam melhorar e/ou facilitar a vida dos indivíduos, principalmente no quesito da interação humano-computador, conceito que encontra-se imerso no cotidiano e possui forte vínculo com a qualidade de vida das pessoas. Diante disso, este trabalho tem como objetivo apresentar e avaliar o desempenho de modelos baseados em aprendizagem de máquina supervisionado para tarefas de reconhecimento das emoções alegria, tristeza e surpresa em textos curtos no idioma Português coletados em comentários de jogos empáticos. As implementações foram realizadas com o auxílio da biblioteca *Scikit-Learn*, utilizando os modelos *Support Vector Machine* (SVM), *Naive Bayes* (NB), *K-Nearest Neighbors* (KNN), e métodos *Ensemble* os quais foram treinados utilizando uma base de dados pré-classificada. O desenvolvimento da aplicação divide-se em 6 etapas de avaliação, onde em cada etapa buscou-se aplicar uma técnica distinta de pré-processamento de dados, possibilitando assim, averiguar o impacto de cada técnica no desempenho dos modelos. Resultados mostram que o modelo SVM atingiu 71% de acurácia em cenários com grande presença de funções de pré-processamento (remoção de acentos, stopwords e aplicação de *stemming*), enquanto os modelos NB e *Votting* atuam respectivamente com 77% de precisão em cenários com pouca robustez de pré-processamento e 77% de precisão em ambos os cenários (com muita ou pouca presença de funções de pré-processamento).

Palavras-chave: Aprendizado de Máquina, Mineração de Emoções, Jogos Empáticos, Textos curtos.

1. INTRODUÇÃO

A chegada da revolução tecnológica flexibilizou significativamente a interação humano-computador (IHC), sendo esta não mais concentrada apenas no contato de redes sociais e utilização de softwares comumente acessados via *personal computer* (PC), como também é

acessível via *smartphones* e dispositivos IoT (Internet das Coisas) que estão constantemente em contato com os indivíduos e se encontram cada vez mais presentes no cotidiano. A computação afetiva, por sua vez, tornou-se um ramo cada vez mais presente devido a esta grande variedade tecnológica, pois se concentra em analisar o estado emocional desses usuários perante a grande demanda de recursos tecnológicos por intermédio da mineração de emoções e sentimentos, os quais podem ser realizadas através de reconhecimentos de expressões faciais, por voz e também por textos. De acordo com [5], às emoções refletem diretamente a qualidade de vida das pessoas, portanto, garantir uma boa avaliação emocional dos usuários em ambientes virtuais é também garantir um bem estar social. Diante disso, criar ferramentas dispondo de metodologias modernas e usuais para o cotidiano e que se concentram em avaliar o estado emocional dos usuários na interação humano-computador pode contribuir para uma boa Experiência do Usuário (*User Experience - UX*).

A mineração de emoções consiste em realizar análises de dados (expressões faciais e gestuais, voz e textos) através de algoritmos de Inteligência Artificial (IA) com intuito de encontrar um padrão que auxilie na classificação ou reconhecimento de uma determinada emoção ou sentimento que é expressada com predominância nesses dados. Sendo assim, é comum a aplicação de modelos baseados em IA para realizar essas tarefas, uma vez que, estes dispõem de grande bagagem e recursos matemáticos e computacionais que permitem o processamento de até milhares de dados para garantir o retorno desejado.

O aprendizado de máquina ou comumente conhecido como *Machine Learning* é um subtópico da Inteligência Artificial juntamente com o *Deep Learning*, e também principal recurso aplicado no desenvolvimento de tecnologias da indústria 4.0. O aprendizado de máquina pode ser dividido em três partes: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. O presente artigo busca estabelecer aplicações do primeiro tipo de aprendizagem em tarefas de classificação de emoções em textos curtos, onde os modelos são previamente alimentados por um conjunto de dados pré-classificados e em seguida são testado com dados já conhecidos pelo desenvolvedor, garantindo uma análise detalhada do desempenho conforme ilustrado na Figura 1.

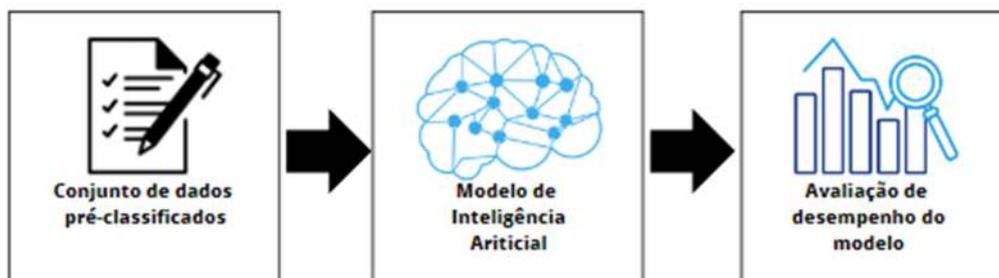


Figura 1: Fluxograma ilustrando o Aprendizado Supervisionado (Fonte: Autor).

Portanto, este trabalho foi organizado em seções bem definidas, na qual a Seção 2 disserta sobre trabalhos correlatos que serviram de inspiração, dispõem de ferramentas e métodos que facilitam de forma significativa o desenvolvimento deste trabalho. Na Seção 3 é abordada a metodologia implementada no trabalho, que visa detalhar sobre as etapas de classificação e o modo em que cada ferramenta foi utilizada. Os resultados obtidos em cada etapa são abordados na Seção 4, que visa mostrar de forma visual através de gráficos, os resultados da acurácia, precisão, *recall*, *F1-score* dos modelos utilizados na etapa de classificação, permitindo uma avaliação geral sobre o desempenho de cada modelo. A Seção 5 foi reservada para as considerações finais sobre o resultado do desempenho dos modelos apresentados na Seção 4.

2. TRABALHOS CORRELATOS

Este trabalho inspirou-se em [1], que buscou realizar a classificação de comentários do jogo empáticos *The Last Day of June* nas polaridades positivo, negativo e neutro, utilizando recursos disponíveis na linguagem de programação R. A base de dados utilizada por [1] foi aproveitada no desenvolvimento do presente trabalho e sofreu adaptações para que os modelos implementados fossem capazes de reconhecer as emoções alegria, tristeza e surpresa. O autor utilizou 3 modelos para a classificação dos comentários em 2 sentimentos e mais o neutro. Os modelos foram *Support Vector Machine*, *Naive Bayes* e *K-Nearest Neighbors* os quais alcançaram 79%, 75% e 45% de acurácia, respectivamente.

Em [2] também há a implementação de aplicações de Processamento de Linguagem Natural, o qual dispõe da biblioteca NLTK (*Natural Language ToolKit*) escrita na linguagem Python 3 para realizar as etapas de pré-processamento dos dados. A mineração de textos foi realizada pela API do *Twitter* e teve como objetivo analisar a satisfação das pessoas acerca de um determinado produto. A análise foi realizada exclusivamente pelo modelo *Naive Bayes*, que foi capaz de classificar comentários nas emoções raiva, expectativa, desgosto, medo, alegria, tristeza, surpresa e confiança, conseguindo classificar até 78% do montante dos dados corretamente.

Em [3] há a disponibilização de uma API nomeada *The Emotion Mining Toolkit* (EMTk), que promove uma série de métodos que permite a classificação de emoções e sentimentos em textos compatíveis com as linguagens de programação *Java*, *Python* e *R*. A grande vantagem da utilização desta API está na velocidade de processamento e classificação dos textos. Em uma comparação realizada pelo autor, a ferramenta *EMTk* foi capaz de realizar a classificação de 4000 textos em polaridade no tempo de 1 minuto e 20 segundos, enquanto a *Senti4SD* realizou o mesmo processo em 56 minutos e 46 segundos. No quesito emoção, a *EMTk* foi comparada com a ferramenta *EmoTxt* na classificação da mesma quantidade de textos, nesse caso a *EMTk* levou 33 minutos e 3 segundos, enquanto a *EmoTxt* demorou 1 hora e 4 minutos para o mesmo procedimento. Entretanto, a API é funcional apenas para textos presentes na língua inglesa, convergindo com o objetivo da aplicação do presente trabalho, servindo apenas de inspiração para criação de métodos semelhantes aos implementados na API.

As métricas adotadas para avaliar os modelos selecionados foram adotadas considerando o trabalho de [4], que implementou os modelos *Support Vector Machine*, *Naive Bayes* e *Logistic Regression* para tarefas de mineração de opinião em textos coletados do *Twitter*, tarefa bastante semelhante à mineração de emoções e que compartilham a mesma lógica na etapa de pré-processamento dos dados textuais. Dessa forma, [4] obteve melhores resultados considerando o modelo *Logistic Regression*, alcançando 82% de acurácia, 78% de precisão e 79% de F1-Score. A Tabela 2 expõe as principais contribuições dos trabalhos supracitados para o desenvolvimento do presente artigo.

Trabalhos	Contribuições
[1] Mineração de texto para análise afetiva da interação dos usuários com jogos empáticos	<ul style="list-style-type: none"> ● Base de dados pré-classificada ● Escolha dos modelos
[2] Análise de dados reconhecendo emoções em textos com objetivo de obter informações úteis para vendas utilizando as ferramentas Naive Bayes e NLTK com Python	<ul style="list-style-type: none"> ● Bibliotecas de tratamento de texto
[3] EMTk - The Emotion Mining Toolkit	<ul style="list-style-type: none"> ● Inspiração de aplicação
[4] Mineração de opinião em textos opinativos utilizando algoritmos de	<ul style="list-style-type: none"> ● Métricas para avaliação do modelo (Acurácia, precisão, recall e f1-

classificação	score)
---------------	--------

Tabela 1: Resumo das contribuições dos trabalhos correlatos.

O presente projeto busca utilizar técnicas semelhantes às implementadas nos trabalhos supracitados, porém contribuir com a criação de uma ferramenta de código aberto que seja capaz de classificar textos em três emoções (alegria, tristeza e surpresa), exclusivamente em Português, uma vez que há escassez de produtos e referências destinados a essa devida finalidade.

3. METODOLOGIA

Nesta seção, busca-se explicar de forma detalhada o desenvolvimento prático deste trabalho, dissertando acerca de conceitos necessários para a compreensão geral do propósito do projeto e também apresentação de ferramentas e API's que foram utilizadas. O presente trabalho busca avaliar o desempenho de modelos de Aprendizado de Máquina Supervisionado para tarefas de classificação de emoções em textos curtos, para isso, optou-se por dividir as etapas de pré-processamento em 6 subetapas, com o objetivo de analisar o desempenho de técnicas comumente aplicadas em mineração de textos na performance dos modelos selecionados. A base de dados utilizada contém 173 comentários, conforme supracitado foi herdada de [1], sendo adicionados mais comentários para diversificar o conjunto, almejando a criação de um modelo mais versátil. Os comentários coletados são dos jogos empáticos *The Last Day of June*, *The Life is Strange 1*, *The Life is Strange 2*, *Detroit: Become Human* e *Brothers- A Tale of Two Sons* os quais foram adquiridos de forma manual das plataformas *Steam* e *Youtube* e podem ser conferidos na base de dados disponibilizada em um arquivo excel¹. A porcentagem da distribuição dos comentários entre as três emoções pode ser consultada na Figura 2. Os jogos empáticos foram selecionados por possuírem a característica de imergir os usuários em diversas emoções com o desenvolvimento de sua trama.

Distribuição das emoções em 173 comentários

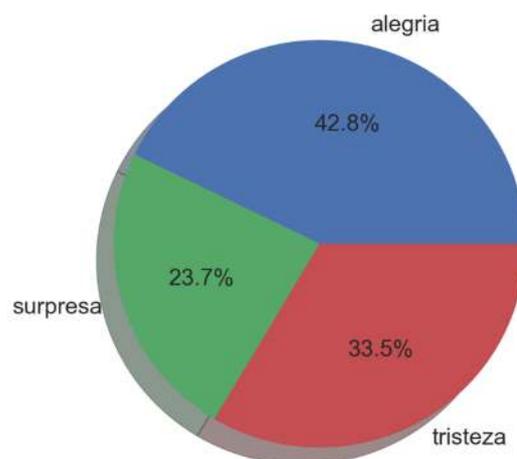


Figura 2: Percentual da distribuição das emoções na base de dados (Fonte: Autor).

O código fonte foi desenvolvido na linguagem de programação *Python 3*, a qual dispõe de grande variedade de bibliotecas com métodos pré-programados e API's que auxiliam no desenvolvimento de códigos de propósito único. Utilizou-se as bibliotecas: *Pandas* para manipulação dos dados e criação de *dataframe* que auxiliam na visualização dos dados, *NLTK* para aplicação de funções de pré-processamento de textos, e a *Scikit-Learn* para criação dos modelos e realização das fases de treino e teste. Em aprendizado supervisionado, a fase de

¹ <https://drive.google.com/drive/folders/19FMAfKALZkDUvwtRD735LkTj5WJV7HKT?usp=sharing>

treino serve para ensinar o modelo a seguir os padrões pré-estabelecidos pelo supervisor, logo, optou-se por reservar 80% da base de dados para realizar o treino do modelo e 20% para avaliar a performance com dados inéditos para o modelo, a proporção de comentários por emoção utilizada está de acordo com a Figura 3.

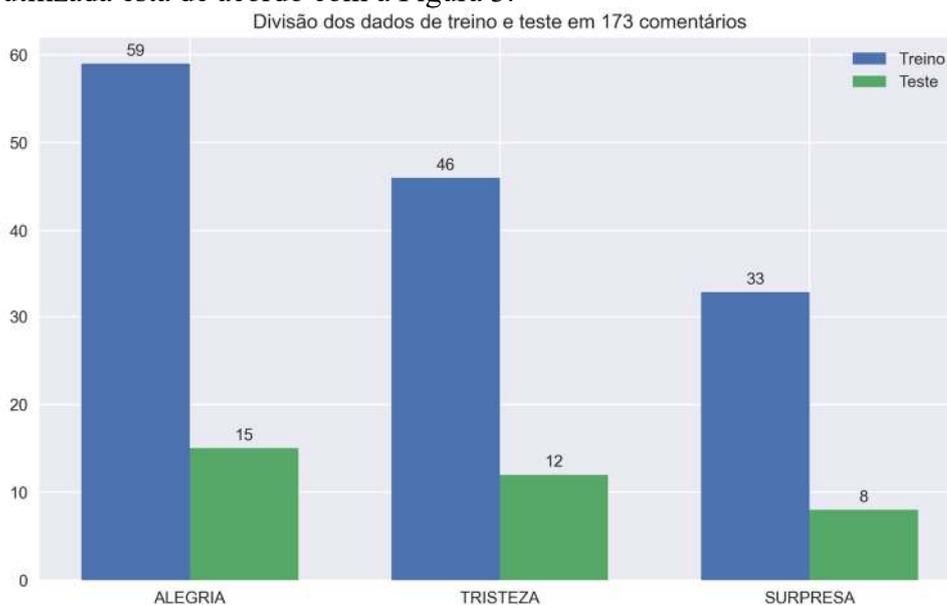


Figura 3: Quantidade de dados para treino e teste (Fonte: Autor).

3.1. Etapas de avaliação

As técnicas de pré-processamento são de grande importância para tarefas de análise de dados e aplicações que envolvem modelos de IA, pois são essas técnicas responsáveis por organizar e ajustar os dados para que os mesmos sejam processados com a menor taxa de ruído pelos modelos. Para analisar o impacto de cada técnica de pré-processamento na performance dos modelos selecionados, optou-se por aplicá-las progressivamente e avaliar os modelos nas métricas propostas em cada etapa. O diagrama da Figura 4 resume as etapas adotadas para a implementação das técnicas, as quais são descritas a seguir.

3.1.1. Primeira etapa

Na primeira etapa foram utilizados os dados sem técnicas de pré-processamento, ou seja, os textos foram repassados diretamente para a função *TFI-DF Vectorizer*, que é responsável por organizar os dados em uma matriz numérica de acordo com as frequências das palavras, permitindo que dados de entrada do tipo *String* fossem devidamente interpretadas pelos modelos que só reconhecem valores numéricos. Os modelos SVM, KNN e NB, assim como nas demais etapas, foram otimizados com a utilização da função *GridsearchCV* disponibilizada pela biblioteca *Scikit-Learn*, que é fundamental para a escolha dos melhores parâmetros na etapa de treino, pois a função é capaz realizar sucessivas comparações entre as combinações de parâmetros dos modelos, garantindo que o modelo seja repassado para a fase de teste de forma otimizada.

3.1.2. Segunda etapa

Na segunda etapa foi aplicada primeira técnica de pré-processamento: remoção dos acentos, que foi criada com o auxílio da biblioteca *unicodedata*, através da função *normalize* que transforma caracteres fora do padrão *Unicode* em seu semelhante.

classificador. A função *RSLPStemmer* também é disponibilizada pela biblioteca NLTK, sendo própria para tratar textos em português.

Original	Stemming
<i>amigos</i>	<i>amg</i>
<i>amigas</i>	<i>amig</i>
<i>amizade</i>	<i>amizad</i>
<i>carreiras</i>	<i>carr</i>

Quadro 1: Exemplo de aplicação da função *stemming*.

3.1.6. Sexta etapa

Na última etapa houve a utilização dos métodos *Ensemble* que recebem os modelos SVM, KNN e NB como parâmetro de entrada e são capazes de entregar na saída o resultado de sucessivas combinações entre os resultados dos modelos anteriores. Os métodos *Votting* e *Stacking* aplicados no presente trabalho foram implementados com a biblioteca *Scikit-Learn*, e foram treinados com a mesma proporção de dados utilizados pelos modelos anteriores.

4. RESULTADOS

Em cada etapa, realizou-se análises gráficas de quatro métricas distintas: *acurácia*, *precisão*, *recall* e *F1-Score*, as quais permitiram visualizar com mais detalhes a performance de cada algoritmo nas fases de teste.

4.1. Primeira etapa:

A Figura 6 expõe o desempenho dos modelos NB, SVM e KNN na primeira etapa, nota-se então, que o modelo NB atua com 77% quando considera-se os dados sem técnicas de pré-processamento. Nas demais métricas não há valores de destaque, entretanto ambos modelos atuam com mais de 50% de performance.

4.2. Segunda etapa:

A Figura 7 retorna às métricas na segunda etapa, onde os modelos atuaram com textos isentos de acentuações, resultando na recaída da precisão do modelo NB de 77% para 65% , e nenhuma melhora significativa nos demais modelos.

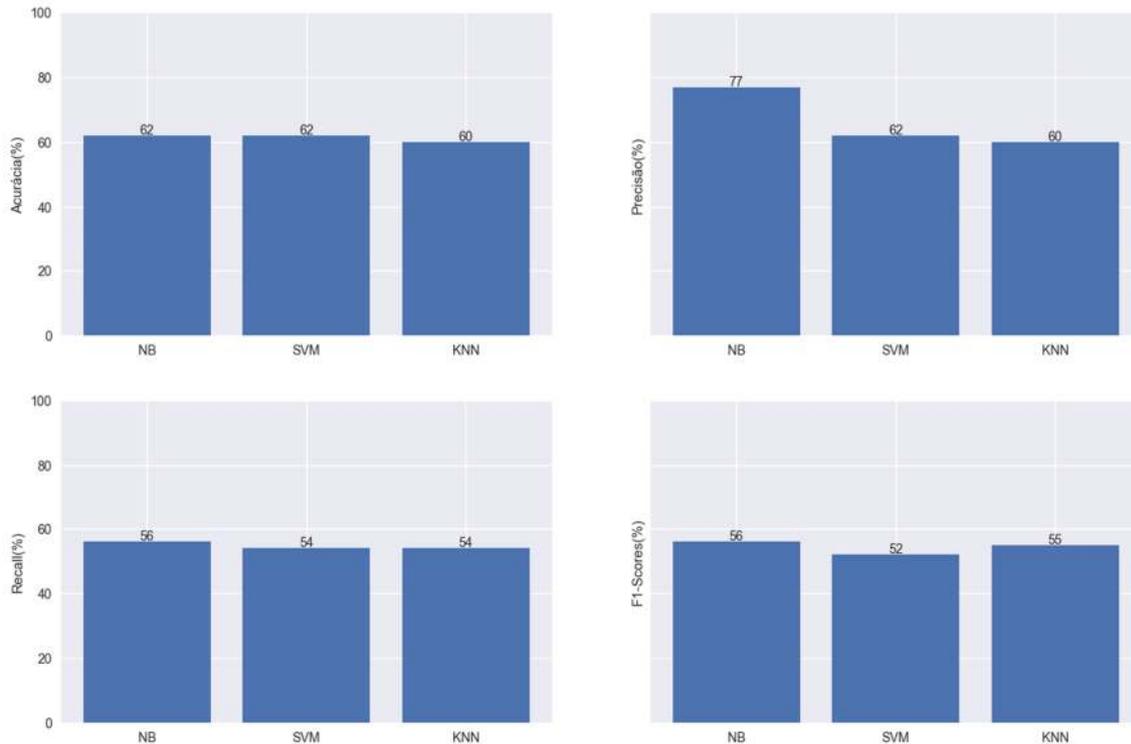


Figura 6: Métricas de avaliação da primeira etapa (Fonte: Autor).

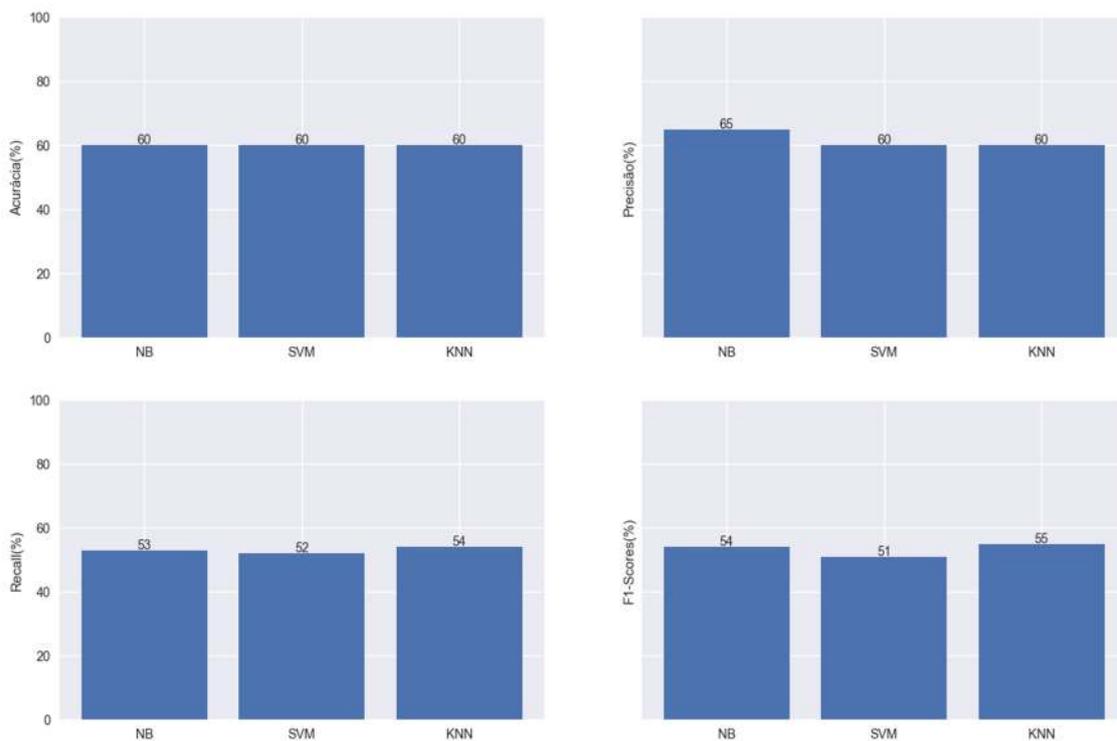


Figura 7: Métricas de avaliação da segunda etapa (Fonte: Autor).

4.3. Terceira etapa:

A partir da terceira etapa, é possível observar ganhos significativos em todas as métricas avaliadas no modelo SVM, entretanto, o modelo Naive Bayes ainda possui boa performance considerando dados sem acentuação e caracteres especiais, garantindo 73% e 65% de acurácia (Figura 8).

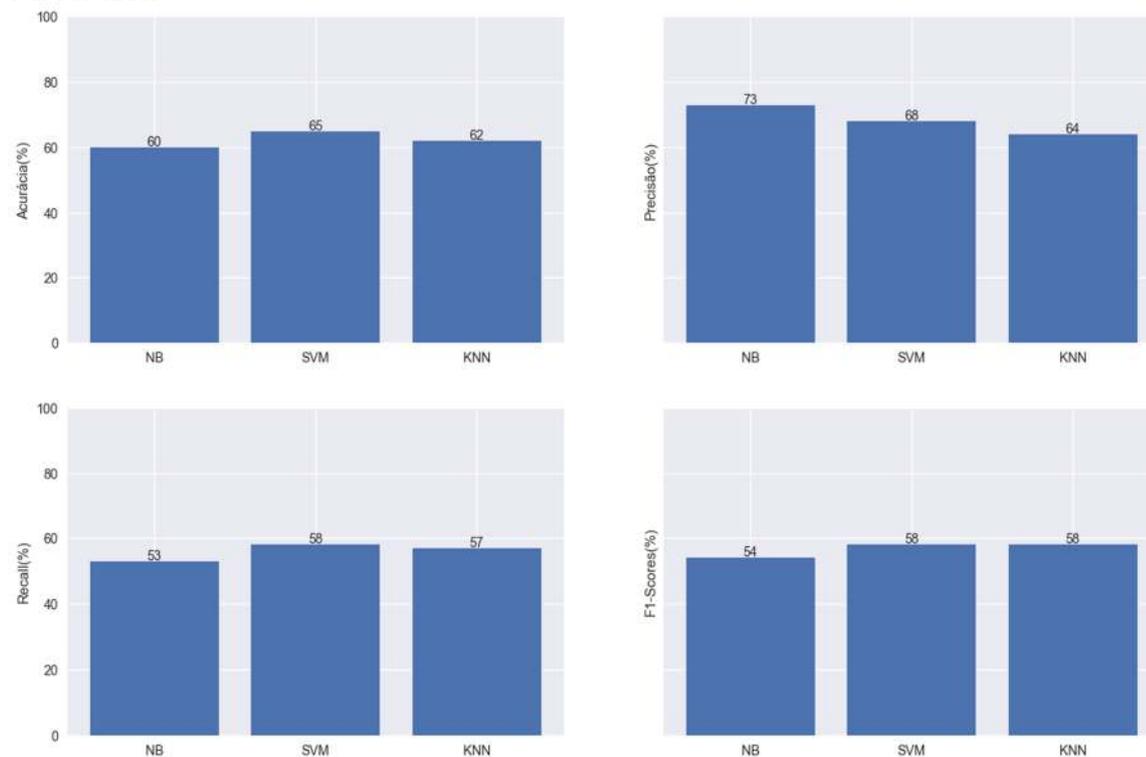


Figura 8: Métricas de avaliação da terceira etapa (Fonte: Autor).

4.4. Quarta etapa:

A quarta etapa de avaliação pode ser considerada como um pequeno empate entre a performance dos modelos SVM e NB, que conforme a Figura 9, alcançaram os mesmos percentuais com a exceção da precisão. O modelo KNN apresenta-se pouco promissor para tarefas de reconhecimento de emoções quando são considerados textos sem *stopwords*.

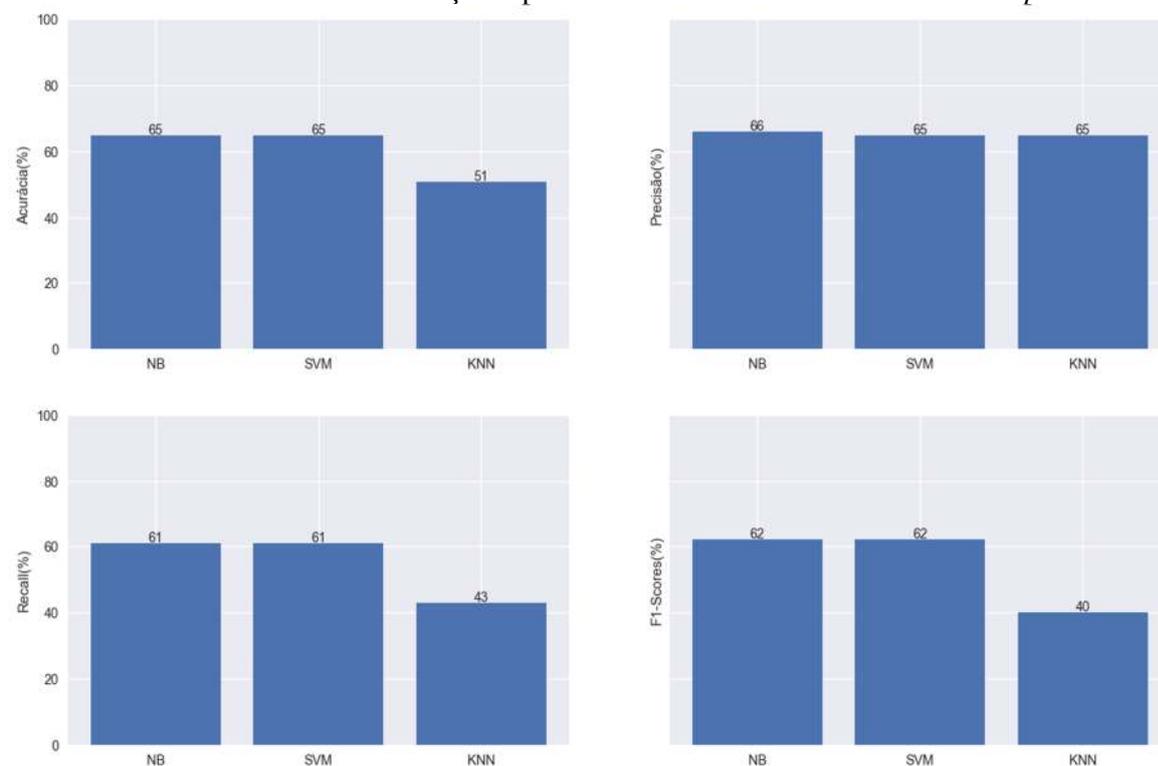


Figura 9: Métricas de avaliação da quarta etapa (Fonte: Autor).

4.5. Quinta etapa:

Na quinta etapa houve destaque para o modelo SVM que alcançou 71% de acurácia devido à redução das palavras ao seu radical (por ex., *stemming*), além de garantir boa performance nas

demais métricas. Os modelos NB e KNN tiveram performance inferior ao do modelo SVM, conforme a Figura 10, quando expostos às técnicas de pré-processamento mais robustas.

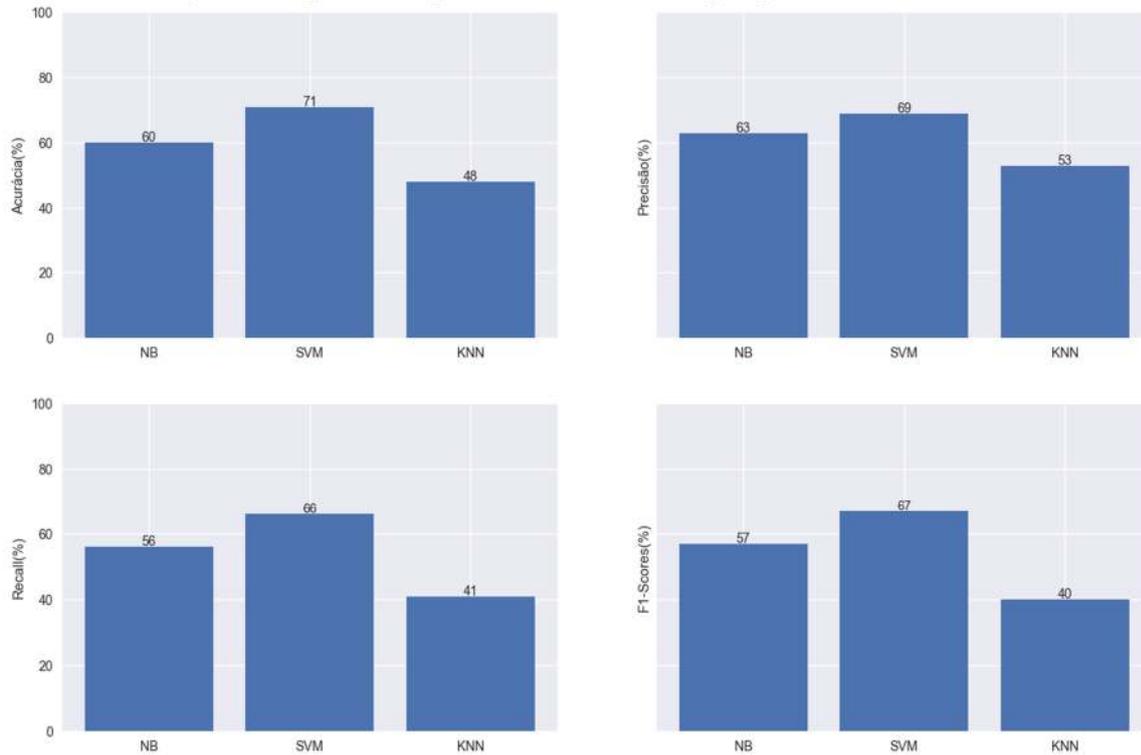


Figura 10: Métricas de avaliação da quinta etapa (Fonte: Autor).

4.6. Sexta etapa:

Na última etapa, o modelo *votting* baseado em métodos *Ensemble* alcançou 77% de precisão, sendo esta precisão a mesma alcançada pelo modelo NB na primeira etapa, enquanto o modelo *stacking* performou com maior taxa de acurácia quando considerada todas as técnicas de pré-processamento nos dados de entrada.

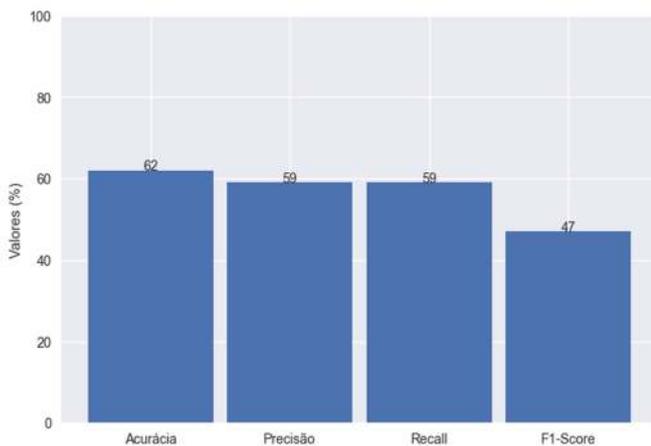


Figura 11: Métricas do método *stacking* (Fonte: Autor).

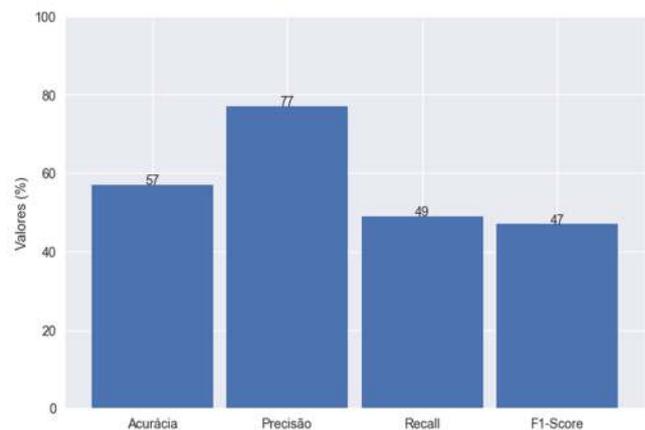


Figura 12: Métricas do método *votting* (Fonte: Autor).

5. CONCLUSÃO

Portanto, é possível afirmar através das análises dos resultados resultados na seção anterior que a escolha de modelos mais robustos para tarefas de classificação de emoções em textos curtos é diretamente proporcional à escala de processamento dos dados. O modelo Naive Bayes se baseia em métodos probabilísticos para realizar a classificação dos dados, sendo

assim, técnicas como o *stemming* influenciam na separação dos casos favoráveis do modelo, garantindo a hipótese de sua performance de 77% de precisão na primeira etapa. Por outro lado, modelos mais robustos como o SVM necessitam de dados rigorosamente processados para melhorar seu limiar de separação de emoções, justificando sua melhora de performance de forma proporcional à passagem das etapas, alcançando 71% de acurácia na penúltima etapa. Os modelos baseados em métodos *ensemble* possuem sua performance dependente dos modelos de sua entrada, sendo assim, podem ser interpretados como modelos híbridos para tarefas de classificação com baixa ou alta robustez de técnicas de pré-processamento. No geral, os modelos mantiveram suas características conforme a literatura abordada no projeto, alcançando percentuais acima de 60% nas métricas acurácia e precisão em quase todas as análises abordadas, viabilizando aplicações em tarefas de reconhecimento de emoções em textos curtos. Como trabalhos futuros, há a possibilidade da implementação dos modelos analisados em uma ferramenta com interface gráfica para inserção de texto e classificação em tempo real, considerando as 6 emoções de Ekman, e não apenas em 3 emoções como foi realizado no presente trabalho.

6. REFERÊNCIAS

- [1] Leonardo Dias Martins. 2019. *Mineração de texto para análise afetiva da interação dos usuários com jogos empáticos*. <https://doi.org/10.14210/cotb.v12.p370-377>
- [2] Gustavo de Castro Minhaneli. 2018. *Análise de dados reconhecendo emoções em textos com objetivo de obter informações úteis para vendas utilizando as ferramentas Naive Bayes e NLTK com Python*. <https://doi.org/123456789/258>
- [3] Fabio Calefato, Filippo Lanubile, Nicole Novielli, Luigi Quaranta. 2019. *EMTk - The Emotion Mining Toolkit*. Brasília, DF. <https://doi.org/10.1109/SEmotion.2019.00014>
- [4] Fernando Santos. 2013. *Mineração de opinião em textos opinativos utilizando algoritmos de classificação*. Brasília, DF. <https://doi.org/handle/10483/7711>
- [5] Paul Ekman. 2004. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life Networking (2nd. ed.)*. Henry Holt and Company, New York, NY
- [6] Savargiv, M., Masoumi, B. & Keyvanpour, M.R. *A new ensemble learning method based on learning automata*. J Ambient Intell Human Comput 13, 3467–3482 (2022). <https://doi.org/10.1007/s12652-020-01882-7>
- [7] Reddy, H., Raj, N., Gala, M. et al. *Text-mining-based Fake News Detection Using Ensemble Methods*. Int. J. Autom. Comput. 17, 210–221 (2020). <https://doi.org/10.1007/s11633-019-1216-5>
- [8] Yadollahi, A., Shahraki, A.G., Zaiane, O.R.: *Current state of text sentiment analysis from opinion to emotion mining*. ACM Comput.
- [9] Z. -q. Wang, X. Sun, D. -x. Zhang and X. Li, "An Optimal SVM-Based Text Classification Algorithm" 2006 International Conference on Machine Learning and Cybernetics, 2006, pp. 1378-1381, doi: 10.1109/ICMLC.2006.258708.
- [10] Kibriya, A.M., Frank, E., Pfahringer, B., Holmes, G. (2004). *Multinomial Naive Bayes for Text Categorization Revisited*. In: Webb, G.I., Yu, X. (eds) AI 2004: Advances in Artificial Intelligence. AI 2004. Lecture Notes in Computer Science(), vol 3339. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30549-1_43