Brasil Sem Fake

Gabriel A. M. de Sá
Ciência da Computação
IESB
Brasília, Brasil
gabriel.sa@iesb.edu.br

Daniel L. O. Lucena
Ciência da Computação
IESB
Brasília, Brasil
daniel.lucena@iesb.edu.br

Eduardo C. P. Fernandes
Ciência da Computação
IESB
Brasília, Brasil
eduardo.fernandes@iesb.edu.br

Alan M. Nascimento
Ciência da Computação
IESB
Brasília, Brasil
alan.nascimento@iesb.edu.br

***Abstract*—O combate à desinformação vem se tornando um tópico de extrema importância. No contexto atual, a disseminação de notícias falsas ocorre em grande velocidade, onde redes sociais contribuem a este rápido alastramento. Levando em consideração a dificuldade em categorizar tais notícias como fidedignas ou não de forma manual, esforços para a detecção automática destas estão sendo constantemente realizados. Neste artigo, será feito a investigação de quais algorítmos performam melhor utilizando as métricas de acurácia e f1-score ao usarem as palavras de cada notícia como features em seu treinamento.**

***Index Terms*—NLP, fake news, características, classificação, Svm, Naive Bayes.**

I. Introdução

Caracterizada como notícias falsas que são compartilhadas deliberadamente com a intenção de enganar o leitor, as notícias falsas (Frequentemente chamadas de fake news) são uma grande ameaça. De acordo com um estudo feito em 2017 [1], foi possível verificar que havia uma quantidade maior de fake news que favoreciam o canditato à presidência Doanld Trump quando comparado com as compartilhadas em favor da Hillary Clinton, outra candidata ao cargo. Este fenômeno não é exclusivo ao Estados Unidos, posto que também foi observado, ao ser feito a análise de 346 notícias falsas [2] compartilhadas durante o periodo eleitoral em 2018, que 45% dessas eram diretamente favoráveis ao candidato Jair Bolsonaro, que eventualmente foi eleito como presidente.

Ao observar este fenômeno, torna-se claro o impacto das notícias falsas na política. Enquanto este é um exemplo pertinente, suas consequências vão além deste meio, afetando até mesmo a área da saúde. Em notícias, o presidente Jair Bolsonaro defendia o uso da cloroquina e hidroxicloroquina como ”tratamento precoce” para a covid-19, mesmo após a divulgação de evidências científicas que mostravam que estas não traziam benefícios aos pacientes que combatiam o vírus.

Ao entender o impacto que as fake news tem em diversos setores, podemos também ver porque a sua detecção é tão importante. Porém, para combater sua propagação, primeiro há a necessidade de detecta-las. Como há um volume grande de notícias circulando constantemente, não é conveniente dar o trabalho de verificar sua veracidade a uma pessoa, consequentemente, a detecção automática usando técnicas de Natural Language Processing (NLP) e Machine Learning podem suprir esta lacuna.

Para chegar ao resultado esperado, isto é, a detecção de fake news de forma automática baseada em NLP, deve ser definido quais algorítmos performam melhor ao serem treinados com as palavras presentes nas notícias, isto é, qual apresentará os valores de acurácia e f1-score melhores.

II. Objetivos Gerais

O objetivo neste artigo é, encontrar um modo de classificar notícias em português como verdadeiras ou falsas, utilizando metodos de treinamento de maquina, como o SVM, Nayve Bayes, Bag of Words e NLP, além de técnicas como a de Lemming. Para isso, sera necessário tratar os texto de entrada, para padronizar e só então permitir que o algoritimos analise o mesmo. Tendo como resultado a acuracia que o algoritimo classificou a notícia, teremos estes dados enviados por meio de uma API.

III. Objetivos Especificos

Realizar a tratação do texto passado pelo usuário, isto é, são feitas remoções de elementos considerados dispensáveis, e muitas vezes, prejudiciais para o processamento de um computador, a fim da criação de uma matriz de termo de documento (document-term matrix), e posteriormente o ato de inferir dela os dados que julgamos relevantes. Após realizado, calcular as métricas de avaliação para cada algoritmo e então concluir qual melhor se adequa para a classificação dessas notícias.

IV. Metodologia

A. Remoção das pontuações

Inicialmente, são realizadas algumas técnicas que buscam garantir que todas as palavras não contenham caracteres especiais e afins. Primeiro, são substituidos todos os caracteres maiúsculos por minúsculos, e então é feita a remoção de todas as pontuações e dígitos do texto.

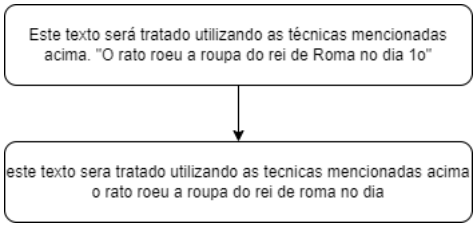

Fig. 1. Primeira etapa de limpeza

Este passo é importante para garantir que não ocorram problemas como a definiçaõ de ”também” e ”tambem” como duas palavras diferentes.

B. Remoção de stopwords

Agora, o próximo passo servirá para que não haja a presença de stopwords (palavras irrelevantes), isto é, palavras que não estão contribuindo para o entendimento da informação principal apresentada no texto. Estas palavras aparecem em abundância em qualquer linguagem humana, e para chegar a uma lista que abrange o máximo de stopwords foram realizadas etapas de análise exploratória das notícias, além da busca em bibliotecas que forneciam esses dados (E.G. NLTK [6])

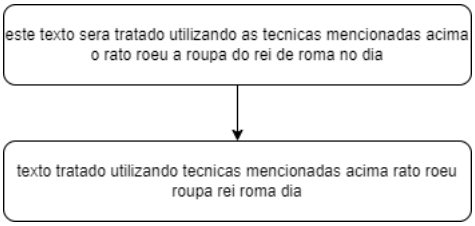

Fig. 2. Remoção das stopwords do texto

C. Lemming

Finalmente, será aplicada a técnica de lemming, ou seja, as palavrãs serão reduzidas a sua raiz, o lemma.

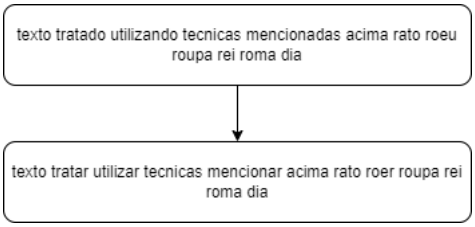

Fig. 3. Reduzindo as palavras ao seu lemming

Ao decidir se seria feito o stemming ou lemming, foi levada em conta que a prioridade é reduzir as palavras a outras que também são gramaticalmente correta, a fim de evitar problemas como stems que foram cortados demais e resultaram numa palavra que não tem mais sentido, ou palavras que apresentavam significados diferentes mas foram reduzidas ao mesmo stem.

D. Naive Bayes

Um dos algoritmos classificadores escolhido foi o Naive Bayes, em virtude de sua implementação ser prática e também boa performance mesmo com um dataset pequeno. Para chegar no featureset usado em seu treinamento, foi realizada o tratamento descrito acima em todos os textos do dataset para então definir o vocabulário usado. Com este vocabulário em mãos, é usado o modelo bag-of-words levando em conta a presença ou não de palavras do vocabulário em determinada notícia. Caso a palavra estivesse presente, esta passava a ter o valor booleano True, caso contrário, permaneceria como False.

Ao realizar este processo em todas as notícias, foi criado um featureset com 7200 linhas, que então foi utilizado para o treinamento. Sua implementação utilizou a biblioteca nltk do python.

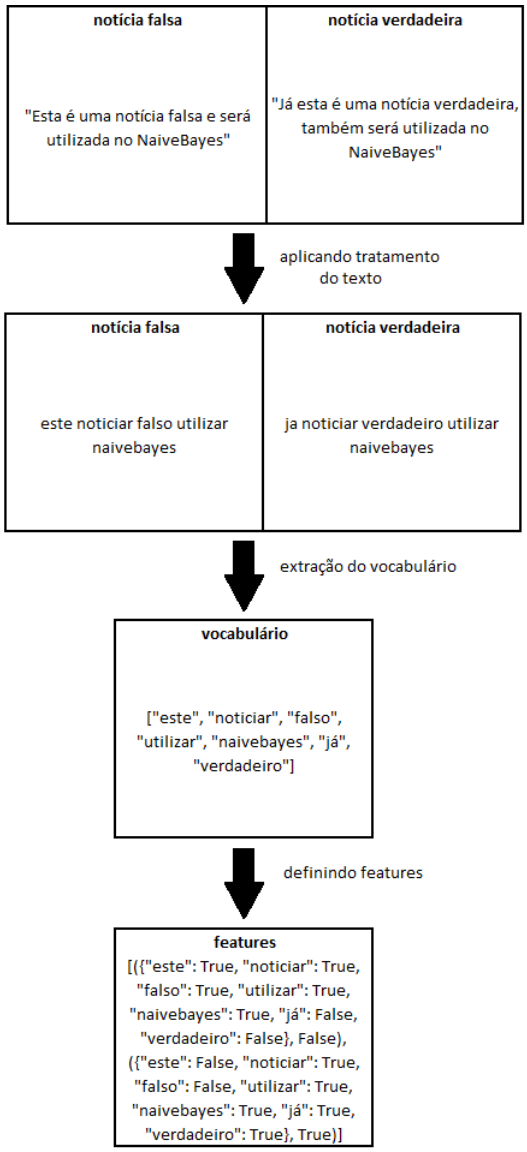

Fig. 4. Definindo features para o Naive Bayes

E. Support vector machine

O outro algoritmo classificador escolhido foi o Support Vector Machine(Svm), que é um algoritmo de fácil implementação e apresenta bons resultados. Assim como para o Naive Bayes, as informações foram tratadas de acordo com os passos citados anteriormente, porém para que os dados pudessem ser utilizados foi necessário transforma-los em um document-term matrix, para isso foi utilizada o método fit_transform() disponível com a biblioteca sklearn do python.

Após realizar tal transformação, o texto seria transformado em duas séries de informações, uma contendo de forma numérica as palavras presentes no corpo do texto e a outra sendo um número apresentando a quantidade de aparições de tais palavras. Com o dataset transformado em tais informações, ele foi usado para treinar o Modelo Svm por meio do uso da biblioteca sklearn em python.

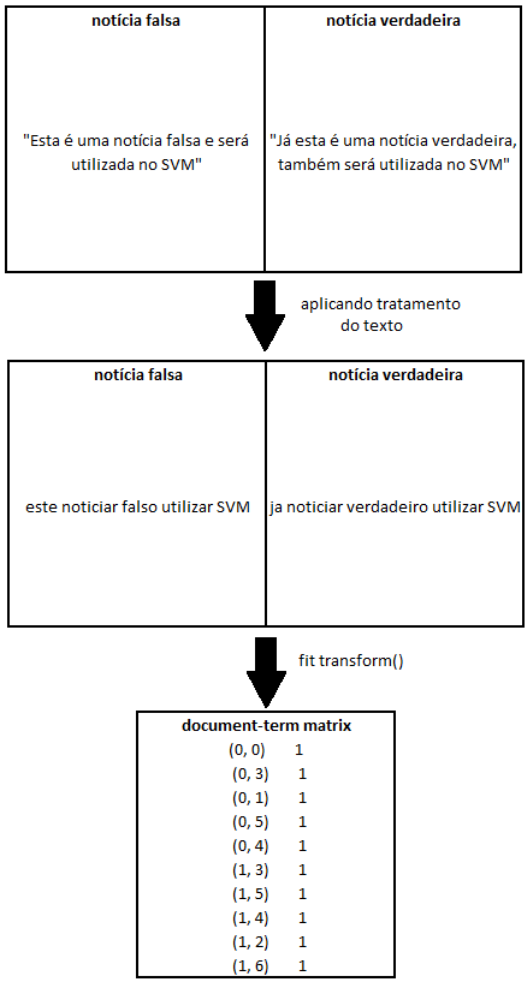

Fig. 5. Definindo features para o Support Vector Machine

V. Trabalhos correlatos

Em seu artigo, Monteiro et al. [3] faz a análise das notícias utilizando como algorítmo de machine learning o support vector machine (SVM), e como características podemos citar algumas: A quantidade média de sentenças, verbos, adjetivos, advérbios, e além destes valores de quantidade, também é feita a análise de sentimentos. Para evitar algum tipo de viés, os resultados quantitativos foram analisados com base na quantidade total de tokens, posto que foi observado que as notícias verdadeiras eram, em sua maioria, bem maiores que as notícias falsas. É importante ressaltar que este artigo é o único citado neste trabalho que utiliza as notícias na língua portuguesa, além de ter realizado a criação de uma corpus com mais de 7000 notícias, todas categorizadas, na chamada fake.br corpus.

Enquanto existem sites dedicados a exposição de notícias falsas (Como o https://www.boatos.org), a obtenção destas ainda se mostra um desafio, especialmente num contexto em que datasets vastos são um requisito na composição de um classificador confiável. Então, Hauch et al. [4] fala sobre como a maioria das descobertas na área foram feitas em inglês, além de trazer a tona um importante detalhe: Línguas diferentes apresentam caracteristicas diferentes que devem ser levadas em conta. Com isso em mente, além das diferenças culturais, apenas realizar a tradução de notícias feitas em outra língua para obter uma maior quantidade de dados se mostra inviável. Hauch et al. também traz algumas características interessantes, e.g. causa, onde é feita a verificação da porcentagem de palavras que buscam atribuir causa ao que quer que esteja sendo descrito, também a certitude, como palavras do tipo ”porque” e ”nunca”.

Em sua pesquisa, Rashkin et al. [5] divide as notícias em três grupos: Sátira, Hoax (Farsas) e propaganda (Em que engana o leitor em prol de uma agenda política). Sátiras também mimetizam notícias verdadeiras, porém costumam apresentar deixas que mostram que não devem ser levadas a sério. Em virtude disso, neste trabalho, serão ignoradas as notícias de natureza satírica. Já em seus resultados, Rashkin et al. afirma que palavras podem ser usadas para exagerar algo, como superlativos e advérbios modais, são utilizadas com maior frequência em notícias falsas, já palavras que indicam valores concretos, como comparativos e números, estavam mais presentes em notícias verdadeiras.

VI. Resultados esperados

Em virtude das limitações impostas pelo dataset, um resultado possível é o overfitting. Basicamente, a obtenção de notícias atuais se mostra um fator importante, posto que os assuntos abordados nelas podem variar dramaticamente em um curto período [7]. Levando isso em conta, e o fato de que a maior corpus publicada em português conta com apenas 7200 notícias já categorizadas, o primeiro desafio se encontra já na coleta de notícias.

Além disso, utilizaremos o f-score como principal métrica de avaliação, pois no contexto atual, notícias falsas que são categorizadas erroneamente como verdadeiras e vice-versa são igualmente perigosas, e o f-score é o mais indicado para esse caso, já que leva em consideração tanto as avaliações falso-poisitivas e falso-negativas.

VII. Resultados Obtidos

Ao realizar o treinamento dos algoritmos citados acima (Naive Bayes e Support Vector Machine), foi possível analisar os modelos utilizando as métricas de acurácia, precisão, recall e f1-score. tais métricas indicam: A porcentagem de classificações corretas (Acurácia), dentre as classificações positivas feitas, quantas estão corretas (Precisão), dentre as classificações positivas como resultado esperado, quantas estão

corretas (Recall), e a média harmonica entre precisão e recall (F1-Score).

Resultados		
	Naive Bayes	Support Vector Machine
Acurácia	84.56%	88.14%
Precisão	91.78%	85.39%
Recall	76%	90.55%
f1-score	83.15%	87.90%

Fig. 6. Métricas dos algorítmos NB e SVM

VIII. Conclusão

Tendo em mente o exposto, foi possível concluir ao analisar as métricas apresentadas acima, que o algoritmo SVM apresentou resultados superiores ao do Naive Bayes, isto graças a seus valores de acurácia e f1-score maiores, mostrando adaptar-se melhor ao dataset e objetivos citados anteriormente ao utilizar as palavras das notícias como features. Logo, esse será o algoritmo usado para nossa plataforma de reconhecimento de fake news.

References

[1] Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." Journal of Economic Perspectives, 31 (2): 211-36.

[2] T. M. S. Galvão, "Fake News na eleição presidencial de 2018 no Brasil", tese, doutorado, comunicação e cultura contemporâneas, UFBA CONTEMPORÂNEAS, 2020

[3] Monteiro, R.A.; Santos, R.L.; Pardo, T.A.; de Almeida, T.A.; Ruiz, E.E.; Vale, O.A. Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. In International Conference on Computational Processing of the Portuguese Language; Springer: Berlin, Germany, 2018; pp. 324–334.

[4] Hauch, V.; Blandón-Gitlin, I.; Masip, J.; Sporer, S.L. Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. Personal. Soc. Psychol. Rev. 2015, 19, 307–342.

[5] Rashkin, H.; Choi, E.; Jang, J.Y.; Volkova, S.; Choi, Y. Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 2931–2937.

[6] S. Bird, Natural language processing with python. O'Reilly Media, 2016.

[7] N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. de Medeiros, and D. M. F. Mattos, "Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges," Information, vol. 12, no. 1, p. 38, Jan. 2021, doi: 10.3390/info12010038.