X FICE 04 e 05 de setembro de 2019
Feira de Iniciação Científica e Extensão
INSTITUTO FEDERAL Catarinense Campus Camboriú

DESENVOLVIMENTO DE UM SISTEMA DE DETECÇÃO AUTOMATIZADA DE FAKE NEWS UTILIZANDO MACHINE LEARNING

Stefano X. Soares[1]; Roger Monteiro[2], Daniel Fernando Anderle[3]; Rodrigo Nogueira[4]

RESUMO

Com o avanço da Internet a facilidade e a velocidade no compartilhamento de notícias, o problema da disseminação de *fake news* aflige a sociedade como um todo, afetando cada vez mais o nosso cotidiano. Tendo em vista os problemas causados pela desinformação, este trabalho tem como objetivo o estudo e análise dos métodos de *machine learning* para desenvolver um mecanismo de coleta de dados de forma inteligente a partir de *datasets* de notícias e a implementação de algoritmos como filtros. Por fim, foi desenvolvido um sistema que permite a classificação de notícias em verdadeiras e fake news.

Palavras-chave: Fake news. Notícia falsa. Machine learning. Aprendizado de máquina.

INTRODUÇÃO

Desde o início da Web, o volume de dados que estão nos repositórios na rede mundial tem crescido de forma exponencial, atualmente são cerca de 200 milhões de sites ativos na Internet, dos quais, apenas a rede social Twitter gera, em média, 500 milhões de postagens por dia. Tal explosão de dados, levou a um estudo do IDC (Institute Data Corporation) que estima que até 2020 serão gerados 44 zettabytes de dados em todo mundo (IDC, 2012).

Nos diferentes nichos de redes sociais que surgiram, observou-se maneiras diferentes de redigir críticas, propiciadas pelas características das aplicações. Sites específicos, como especializados em críticas de filmes, permitem que usuários escrevam textos relativamente longos. Os microblogs, por outro lado,

[1] Estudante de Sistema de Informação - Instituto Federal Catarinense - stefano.xavier@hotmail.com
[2] Estudante de Analise e Desenvolvimento de Sistemas - UNIASSELVI - roger.o.monteiro@gmail.com
[3] Professor Instituto Federal Catarinense - Campus Camboriú - daniel.anderle@ifc.edu.br
[4] Professor Instituto Federal Catarinense - Campus Camboriú - rodrigo.nogueira@ifc.edu.br

impõem limites na quantidade de caracteres das mensagens e não são ambientes exclusivamente destinados para publicação de críticas. No processo de descoberta e pesquisa que prosseguiu nas redes sociais, surgiu a necessidade de expressar opiniões de forma mais direta (VON LOCHTER, 2015).

Segundo Nogueira (2018), os sites de notícias são o terceiro maior veículo de informação mais acessado da Internet, perdendo apenas para aplicativos de mensagens e redes sociais. Esta informação reflete a importância do uso de sites de notícias e seu impacto no cotidiano das pessoas.

Juntamente com a importância de textos de notícias e seu compartilhamento das mesmas em redes sociais, vem a ascensão e disseminação das fake news. Desde meados de 2017, a quantidade de eventos e debates acerca deste fenômeno que vem sendo chamado de fake news cresceu de forma. Fake news pode ser definida como artigos de notícias que são intencional e verificadamente falsos e podem enganar os leitores. Em nessa definição de fake news inclui artigos de notícias fabricados intencionalmente, como um artigo amplamente compartilhado do agora extinto site denverguardian.com com a manchete “FBI agent suspected in Hillary email leaks found dead in apparent murder-suicide”(Agente do FBI suspeito de vazamento de e-mail de Hillary encontrado morto em aparente assassinato-suicídio) (DELMAZO, 2017).

Dado seu destaque, tem sido realizadas diversas multidisciplinares sobre o tema. Almejando contribuir com tais pesquisas, este trabalho tem como objetivo acoplar à etapa de ETL (Extract, Transform, Load) de um Data Warehouse de Notícias o enriquecimento semântico através de classificação do tipo de notícias: real ou falsa.

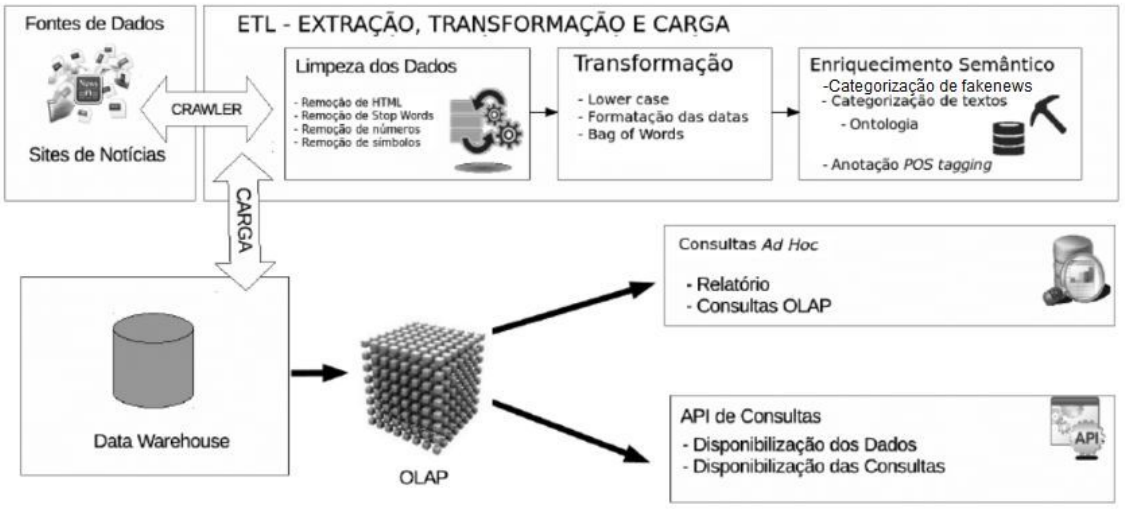
PROCEDIMENTOS METODOLÓGICOS

A primeira etapa deste projeto foi dedicado ao levantamento bibliográfico (PIZZANI et. al, 2012), onde através de artigos e livros se obteve a fundamentação teórica e estado da arte, foi de suma importância para se obter os melhores métodos de aprendizado de máquina empregados durante os experimentos.

Em uma segunda etapa, foi desenvolvido um script de coleta e análise de notícias, permitindo com que esta pesquisa também se enquadre como pesquisa tecnológica de acordo com (JUNIOR et al. ,2014), pois o produto final é conjunto de arquitetura, software, complementado de um conjunto de dados.

No que se refere à base de dados, após pesquisa bibliográfica sobre dados com fake news, pode se verificar que existem poucos recursos disponíveis no idioma Português do Brasil, no qual o dataset mais utilizado é o Fake.br (MONTEIRO et al., 2018). Tendo como em vista complementar este conjunto de dados e obter melhores resultados este trabalho também se propões a coletar dados de notícias. A metodologia de desenvolvimento prático deste trabalho é baseada na arquitetura proposta por NOGUEIRA (2018), na qual o classificador gerado será acoplado a etapa de ETL de um Data Warehouse gerando o enriquecimento semântico em uma nova dimensão.

Figura 1. Arquitetura utilizada



Fonte: Nogueira (2018).

Para realizar os experimentos foi desenvolvido um web crawler, utilizando a linguagem python, juntamente com a biblioteca *beautiful soup* para a coleta inicial dos dados. Foi construído um dataset composto por 1744 títulos e corpo de notícias falsas coletadas dos sites <boatos.org> e <g1.globo.com/fato-ou-fake>, e 3185 títulos e corpo de notícias verdadeiras coletadas do site brasil.elpais.com. Inicialmente será efetuado testes utilizando apenas os títulos das notícias,

posteriormente o corpo juntamente com título e fazer um comparativo entre ambos. Para isso, serão utilizados os algoritmos de aprendizado de máquina (Machine Learning), Regressão Logística, AdaBoost, Naive Bayes e SVM (KOSALA, 2000) .

A partir da criação de um sistema de coleta, com um algoritmo acoplado à etapa de ETL, este irá automaticamente classificar os dados coletados, aumentando assim a acurácia do classificador, e gerando uma base maior de dados para futuros trabalhos de combate a *fake news.* Também foi construído uma interface Web, onde o usuário será capaz de submeter um link e verificar se este é ou não uma notícia verdadeira, servindo este como protótipo antes de ser submetido a etapa de ETL (sendo esta, o propósito geral deste trabalho).

RESULTADOS ESPERADOS OU PARCIAIS

Após a aplicação dos algoritmos Regressão Logística (Logistic Regression), AdaBoost, Naive Bayes e SVM (kernel linear), os mesmos obtiveram a acurácia de 88,85%, 81,37%, 86,22% e 87,45% respectivamente, no modelo de testes. Como técnica de avaliação dos modelos empregados, foi utilizado a validação cruzada com o método k-fold = 10.

Novamente o dataset foi dividido entre treino e teste, juntando agora os títulos ao corpo das notícias. Receberam os mesmos tratamento acima citados, obtendo a acurácia de 90,88%, 84,23%, 91,19% e 91,16% nos algoritmos Regressão Logística (Logistic Regression), AdaBoost, Naive Bayes e SVM respectivamente. A aplicação do método de validação cruzada, revelou um overfitting em alguns casos.

Por fim, o dataset foi dividido para utilização apenas dos corpos das notícias. Foram empregados os mesmos métodos utilizados anteriormente em relação ao tratamento e limpeza dos dados. A aplicação dos algoritmos resultou em 90,88%, 94,23%, 91,19% e 91,16% de acurácia nos algoritmos Regressão Logística (Logistic Regression), AdaBoost, Naive Bayes e SVM respectivamente.

Tabela 1. Comparativo entre os datasets em relação à acurácia e validação cruzada.

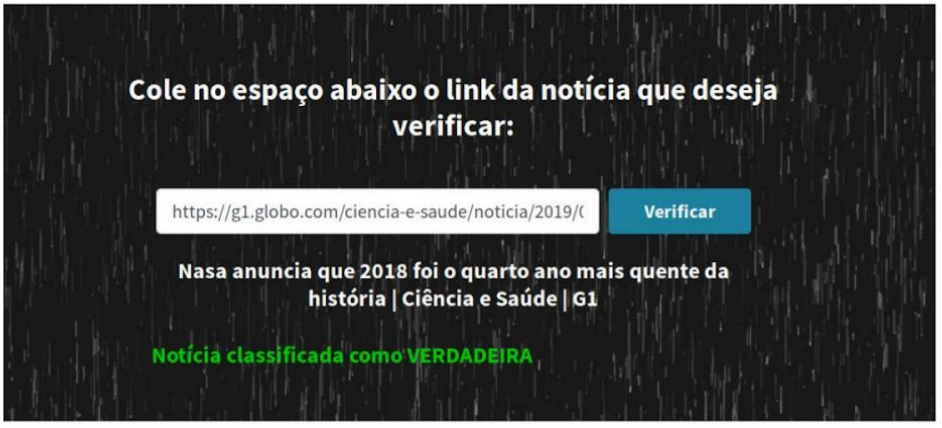
	Regressão Logística	AdaBoost	Naive Bayes	SVM (kernel Linear)
Título	88,85%	81,37%	86,22%	87,45%
K-fold	0,88	0,75	0,86	0,55
Corpo	97,40%	95,12%	97,80%	98,62%
K-fold	0,97	0,95	0,97	0,64
Título + Corpo	90,88%	84,23%	91,19%	91,16%
K-fold	0,90	0,84	0,91	0,54

Fonte: Os autores.

A partir da análise de resultados, o método de Naive Bayes foi selecionado o melhor método, pelo fato de obter uma alta acurácia, complementado de ser um método de aprendizado incremental (online).

Posterior ao acoplamento foi desenvolvido a interface de classificação de fake news, mostrada pela Figura 2. e está disponível no servidor https://detectorfakenews.herokuapp.com/. A ferramenta espera como parâmetro o link de um site de notícia, e retorna se ele é ou não uma notícia falsa (*fake news*)

Figura 2. Interface Web da Aplicação desenvolvida.



Fonte: Os autores.

CONSIDERAÇÕES FINAIS

O overfitting constitui-se um problema recorrente em bases textuais.

Alguns algoritmos chegaram a resultados bastante relevantes, mas ao aplicarmos a validação cruzada com k=10, notou-se um grande overfitting em alguns casos. Sendo assim, observou-se que o algoritmo Naive Bayes obteve além da alta acurácia, tolerância ao overfitting.

Para futuros trabalhos, tem-se como objetivo avaliar outras características técnicas de pré-processamento, aumentar a base de treino, aplicar os novos resultados a interface web, e posteriormente, o acoplamento a ETL do Data Warehouse.

REFERÊNCIAS

DELMAZO, Caroline; VALENTE, Jonas CL. **Fake news nas redes sociais online: propagação e reações à desinformação em busca de cliques**. Media & Jornalismo, v. 18, n. 32, p. 155-169, 2018.

IDC. Gantz, J., & Reinsel, D. (2012). **The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east**. IDC iView: IDC Analyze the future, 2007(2012), 1-16.

KOSALA, Raymond; BLOCKEEL, Hendrik. **Web mining research: A survey**. ACM Sigkdd Explorations Newsletter, v. 2, n. 1, p. 1-15, 2000.

JUNIOR, Vanderlei FREITAS et al. A pesquisa científica e tecnológica. Espacios, v. 35, n. 9, 2014.

MONTEIRO, Rafael A.; SANTOS, Roney L. S.; PARDO, Thiago A. S.; ALMEIDA, Tiago A. de; RUIZ, Evandro E. S.; VALE, Oto 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, Cham, 2018. p. 324-334.

NOGUEIRA, Rodrigo Ramos. **O Poder do Data Warehouse em Aplicações ed Machine Learning: Newsminer: Um Data Warehouse Baseado em Textos de Notícias.** São Paulo: Nea, 2018.

VON LOCHTER, Johannes et al. **Máquinas de classificação para detectar polaridade de mensagens de texto em redes sociais**. 2015.