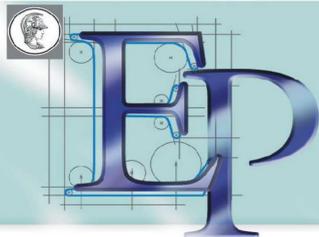


Projeto de Formatura – Turmas 2019



PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia Elétrica – Ênfase Computação

Tema: **SOFTWARE DE CHECAGEM DE NOTÍCIAS EM LÍNGUA PORTUGUESA PARA CLASSIFICAÇÃO DE FAKE NEWS**

MOTIVAÇÃO

O IBGE realizou um estudo em 2016 que aponta que existem mais 100 (cem) milhões de internautas no Brasil e, segundo o Reuters Institute (2018), mais de 66% dos brasileiros usam redes sociais como fonte de notícias. Por mais benéfico que isso seja, deve haver uma preocupação com que tipo de conteúdo essas pessoas estão consumindo, pois está se tornando cada vez mais comum o compartilhamento de *fake news*, impactando a vida de milhões de pessoas.

18/10/2013 04h06 - Atualizado em 18/10/2013 12h15

Após denúncia de maus-tratos, grupo invade laboratório e leva cães beagle

Ativistas invadiram laboratório de pesquisa em São Roque nesta sexta (18). Empresa alega que realiza testes dentro de normas e exigências da Anvisa.

Como funciona a engrenagem das notícias falsas no Brasil

RESUMO Reportagem mostra como funciona a fábrica de títulos sensacionalistas e inverdades que se

PUBLICIDADE

Figura 1 – Consequências negativas de fake news

Assim, essa tese surgiu do incômodo causado pelas dimensões tomadas pela difusão de notícias falsas na Internet. Fato acentuado durante as eleições presidenciais no Brasil em 2018, as quais foram marcadas pela veiculação em massa de informações falsas.

OBJETIVO

Tem-se por objetivo utilizar de algoritmos de Inteligência Artificial para classificar notícias veiculadas na Internet quanto à sua veracidade baseada em características como:

- Análise dos padrões de escrita
- Análise da fonte
- Análise de sentimento

Dessa forma, a partir do link de uma notícia, vindo tanto de redes sociais como de sites jornalísticos e blogs, classificar o texto em Verdadeiro ou Falso.

Utilizando Python com a biblioteca *sklearn* e um servidor *Flask* para suportar as interações com o usuário.



Integrantes: Bruno Henrique Vasconcelos Lemos
João Raphael de Souza Morales

Professor Orientador: Tereza Cristina Melo de Brito Carvalho

METODOLOGIA

Primeiro buscou-se entender quais são as características mais relevantes que definem uma notícia falsa e que podem ser capturadas por software, como: escrita, proporção de erros, número de substantivos, adjetivos, pronomes, etc, além de características como o autor, plataforma veiculada e uma análise de sentimento na mensagem do texto.

Então, de posse das principais que definem as *fake news*, procurou-se analisar qual algoritmo é mais adequado para classificar esse tipo de dado. Para tanto, foram testados Naive Bayes, Lasso, K-Nearest Neighbors, Neural Networks e Random Forest.

De forma mais macro, a arquitetura da aplicação foi definida como:

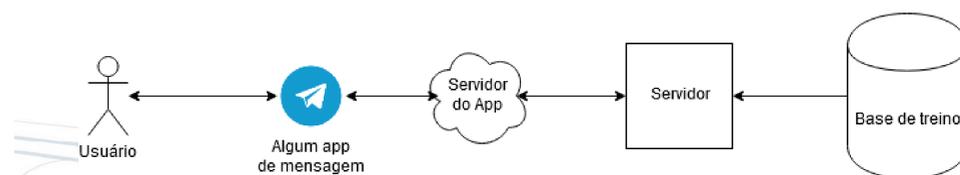


Figura 2 – Arquitetura da aplicação

RESULTADOS

Verificou-se que o problema de fato pode ser descrito de forma satisfatória por variáveis linguísticas, aquelas relacionadas à forma como a notícia foi escrita. Porém, para variáveis relacionadas à fonte da notícia (canal de origem, autor, mídia veiculada) notou-se um aumento substancial do viés.

Os resultados de acurácia da predição por diferentes modelos aplicados em uma base classificada manualmente de mais de 7.000 notícias podem ser verificados:

Modelo	Split Base	Acurácia(%)
KNN	Simple Split	78-80
KNN	KFold	78-80
NAIVE BAYES	Simple Split	90-93
NAIVE BAYES	KFold	90-92
NEURAL NETWORK	Simple Split	90-93
NEURAL NETWORK	KFold	90-94
RANDOM FOREST	Simple Split	92-95
RANDOM FOREST	KFold	92-95

Figura 3 – Modelos de algoritmos de classificação aplicados

Assim, a aplicação mostra grande potencial no auxílio da formação de conhecimento pelo público que consome conteúdo por meio da Internet, principalmente pelas redes sociais.