



# SPHINX

09/11/2018

---

Ewelly Fabiane

Ibukun Didier

Victor Rocha

Boa vista, RR

## Visão geral

A nossa aplicação é uma extensão que será disponibilizado pela Web Store (Plataforma de extensões e tema do Google Chrome) para detectar possíveis notícias falsas com uso de machine learning. Na aplicação o usuário pode verificar se há possibilidade da notícia mostrada na página a ser analisada seja falsa, além de mostrar várias informações sobre o conteúdo da página. Além disso, usuários competentes podem sinalizar se a notícia apresentada é falsa ou não.

## Clientes

Pretende atender pessoas que desejam obter notícias verídicas habilmente, e principalmente pessoas cujo o trabalho necessite de uma verificação ágil sobre a autenticidade da notícia. De forma a proporcionar comodidade aos usuários, que não precisarão fazer buscas em vários sites para descobrir a veracidade da notícia. E ficarão mais atentos para possível notícia falsa.

## Objetivos

Criar uma aplicação que traga comodidade para aqueles que necessitam verificar rapidamente a veracidade de notícias no meio virtual.

1. Analisar páginas de notícias;
2. Mostrar escala com chances da notícia ser falsa;
3. Mostrar polaridade da notícia;
4. Mostrar nível de relevância ao longo do tempo da notícia (Data em que a notícia teve maior repercussão);
5. Validar análise contextual, isto é, verificar se o título e a notícia condizem;
6. Verificar existência e a veracidade das fontes da notícia;
7. Possibilitar sinalização de fake news para usuários com competência para tal.
8. Desenvolver e aprimorar técnicas e algoritmos com uso de machine learning para perpetuar o treinamento dos dados com o uso das informações já validadas no banco de dados.

9. Ampliar a estrutura dos dados a serem reconhecidos, isto é, acrescentar formatos de dados como áudios, vídeos, imagens.
10. Desenvolver futuramente um algoritmo de aprendizagem não supervisionada.

## Especificações

Para a realização do projeto iremos usar:

- as bibliotecas de web crawler : request , BeautifulSoup
- processamento de texto: TextBlob
- as bibliotecas sklearn , pandas e numpy foram utilizada para iniciar o processo de machine learning mais precisamente : model\_selection, preprocessing , linear\_model, naives bayes, metric, svm,TfidfVectorizer, CountVectorizer, train\_test\_split

A análise de dados é feita processando todos os dados html de uma página, para depois tratar os dados de modo a obter o título e o texto da notícia. Com isso feito o texto da notícia é fragmentado em parágrafos usando como parâmetro do *split()* os “\n” do texto. Esses parágrafos são inseridos em uma tabela onde ficam relacionados com o título da notícia para depois serem alimentados para o treinamento do Pandas. Vamos determinar as características que vão ser analisadas pelo Machine Learning.

## Plataformas

O aplicativo funciona com o Google Chrome, utilizando as linguagens: PHP, JavaScript, HTML, CSS para o front-end da aplicação e Python, PHP para o back-end.

## Slide de apresentação

<https://drive.google.com/open?id=1j-EbNsQ7RqnmEQjWEoIFl9yfX3gKO15VhSciWNjjKwA>