Liar Liar! Pants on Fire!

Fake News Detector powered by Machine Learning

Faiz Hasmani

Shruti Bangera

Yazhini Gopal

Nida Iqbal

Syed Muddassir Raza

Nora Younes

Fake News Detector powered by Machine Learning

Reason for Topic Choice

In a time when we can spread and receive information quickly, thanks to the technology and high-speed networks , media literacy is very important.

Detecting fake news has become more important than ever for the sake of a peaceful society. A prediction tool would be a terrific way to help people make more educated decisions on news.

FNN
FAKE NEWS NETWORK

Reason for Topic Choice

In a time when we can spread and receive information quickly, thanks to the technology and high-speed networks , media literacy is very important.

Detecting fake news has become more important than ever for the sake of a peaceful society. A prediction tool would be a terrific way to help people make more educated decisions on news.

Only 26% of Americans are confident in their ability to detect fake news

THAT MEANS THAT

242,868,000

AMERICANS WOULD NOT BE ABLE TO TELL FACT FROM FAKE

Questions We Hope to Answer

✓ Is the article True or False?

✓ What is the accuracy of the prediction model?

✓ Are Fake news articles only found on less reliable websites?

Data Source

- University of Victoria

- Reuters.com (True articles)

- Politifact (Fake articles)

title	text	subject	date
Donald Trump Sends Out Embarrassing New	Donald Trump just couldn't wish all	News	31-Dec-17
Drunk Bragging Trump Staffer Started Russia	House Intelligence Committee Chair	News	31-Dec-17
Sheriff David Clarke Becomes An Internet Jok	On Friday, it was revealed that form	News	30-Dec-17
Trump Is So Obsessed He Even Has Obamaâ€	On Christmas day, Donald Trump an	News	29-Dec-17
Pope Francis Just Called Out Donald Trump D	Pope Francis used his annual Christr	News	25-Dec-17
Racist Alabama Cops Brutalize Black Boy Whi	The number of cases of cops brutali	News	25-Dec-17
Fresh Off The Golf Course, Trump Lashes Out	Donald Trump spent a good portion	News	23-Dec-17
Trump Said Some INSANELY Racist Stuff Insid	In the wake of yet another court de	News	23-Dec-17

Tech Stack

Jupyter Notebook

Machine Learning

PYTHON

Google Colab

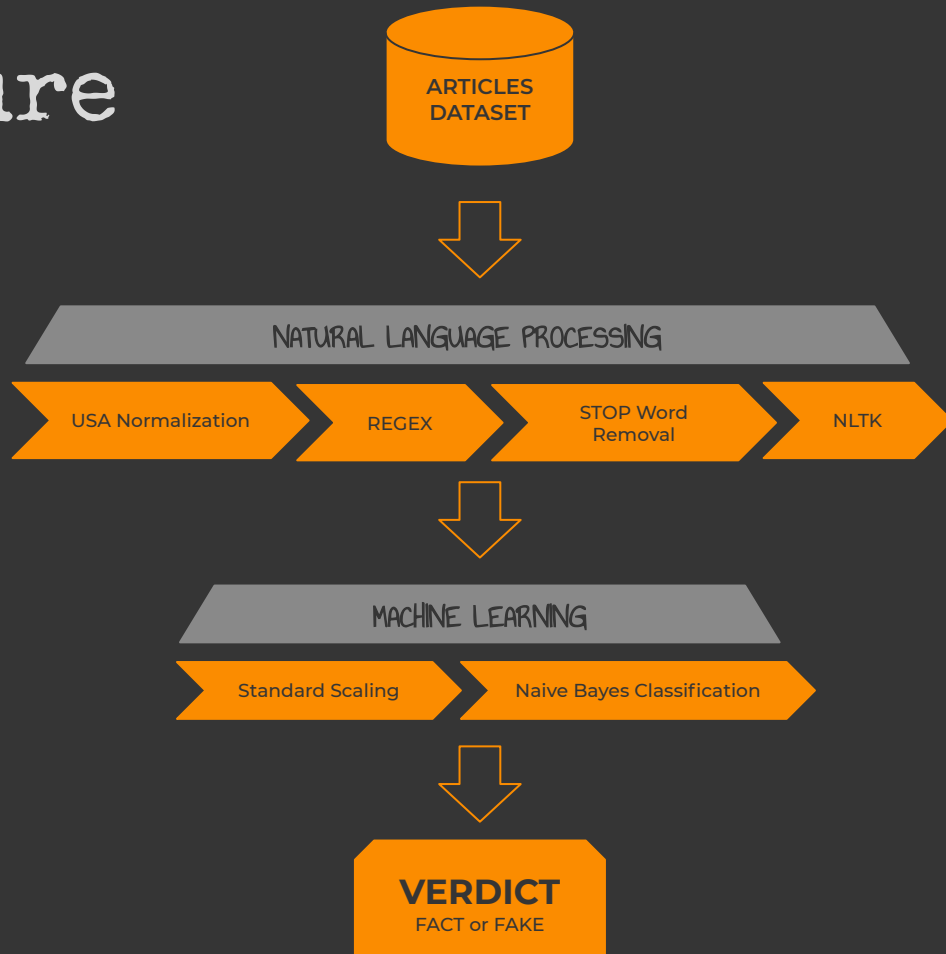Natural Language Processing

Tableau

SQLAlchemy

CSS

HTML

Flask

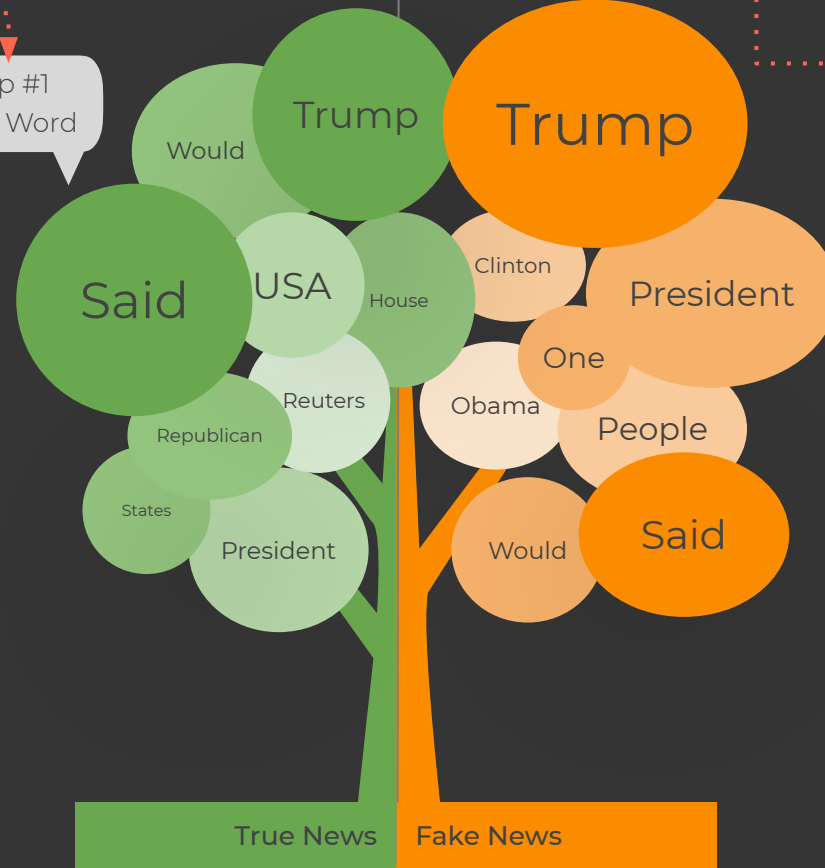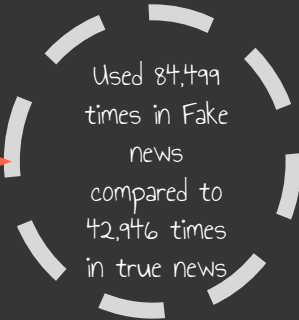PostgreSQL

Heroku

Bootstrap

Architecture

ARTICLES DATASET

NATURAL LANGUAGE PROCESSING

USA Normalization → REGEX → STOP Word Removal → NLTK

MACHINE LEARNING

Standard Scaling → Naive Bayes Classification

VERDICT
FACT or FAKE

Word Cloud

MOST COMMON WORDS

Top #1
True Word

Top #1
Fake Word

Used 54,330 times in true news, compared to 32,893 times in fake news

Used 84,499 times in Fake news compared to 42,946 times in true news

Trump

Trump

Would

Said

USA

House

Clinton

President

One

Obama

Reuters

Republican

People

States

Would

Said

President

President Names more likely to be found in Fake news
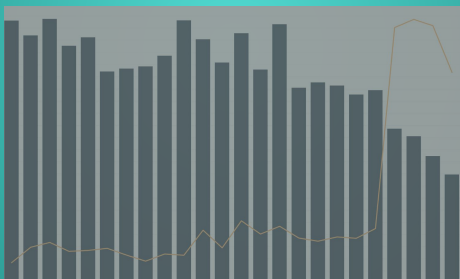
True News   Fake News

Results of Analysis

Top Features

- **Length** of Title (64 vs 88 characters)
- **Pronouns** used in articles (2.02% vs 3.90%)

SVM Model

Pre requisites
- NLP
- Standard Scaling

Pros
- 95% accuracy

Cons

- Longer training time for large datasets
- Regex Integrations issues

Naive Bayes Model

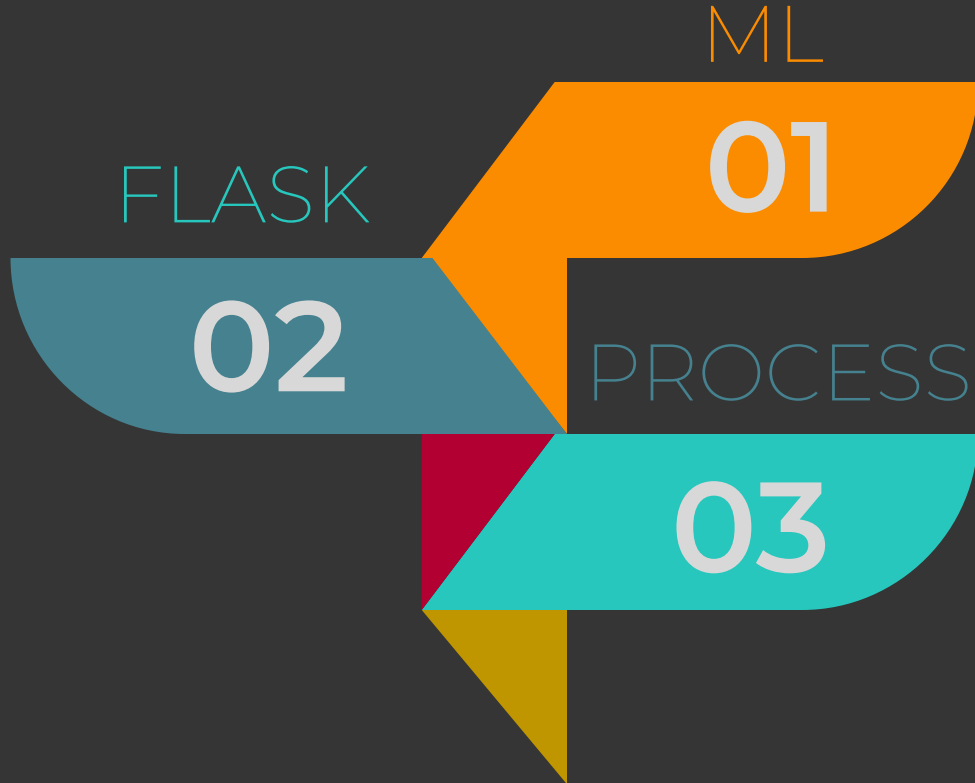**Pre requisites**
- Count Vectorization

Pros

- Lesser training time
- Easier to integrate with Flask

Cons
- 92% accuracy

What the team would have done differently?

ML

01

FLASK

02

PROCESS

03

Train ML model on a comprehensive dataset with more features

Identify integration issues early on in the development process

Build Flask Application using NLP and the most accurate ML model

Recommended Add-ons

- Determine the **accuracy** of prediction for every Test article

- Web scrape articles from different News channels and rate them based on their **reliability**

- Receive User feedback and build a **Reinforcement** machine Learning model

Fake News Detection!

How can you help?