news_Buster ©

FAKE NEWS DETECTOR

STOP THE SPREAD OF MISINFORMATION WITH FACT-CHECKING

DECEMBER 3, 2019

Alex Kwan

Aaron Tan

Elise Lagacé

Gurtej Bhasin

Jacob Bulir

OBJECTIVE & SCOPE

Issues with Fake News:

- Plummeting stock price

- Destroyed reputation

- Unreasonable customer expectations/weak engagement

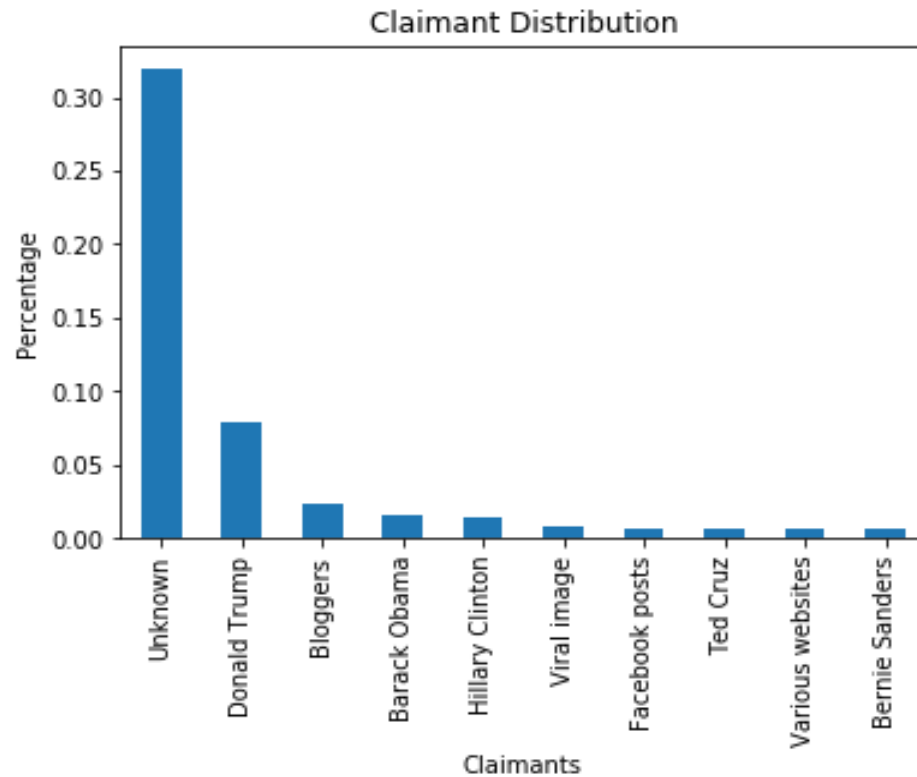## The Solution – "Fake News Detection" Algorithm

- Combined with your current content-management algorithms

- Significant reduction in fake news content

- Build trust with customers – build a long lasting reputation!
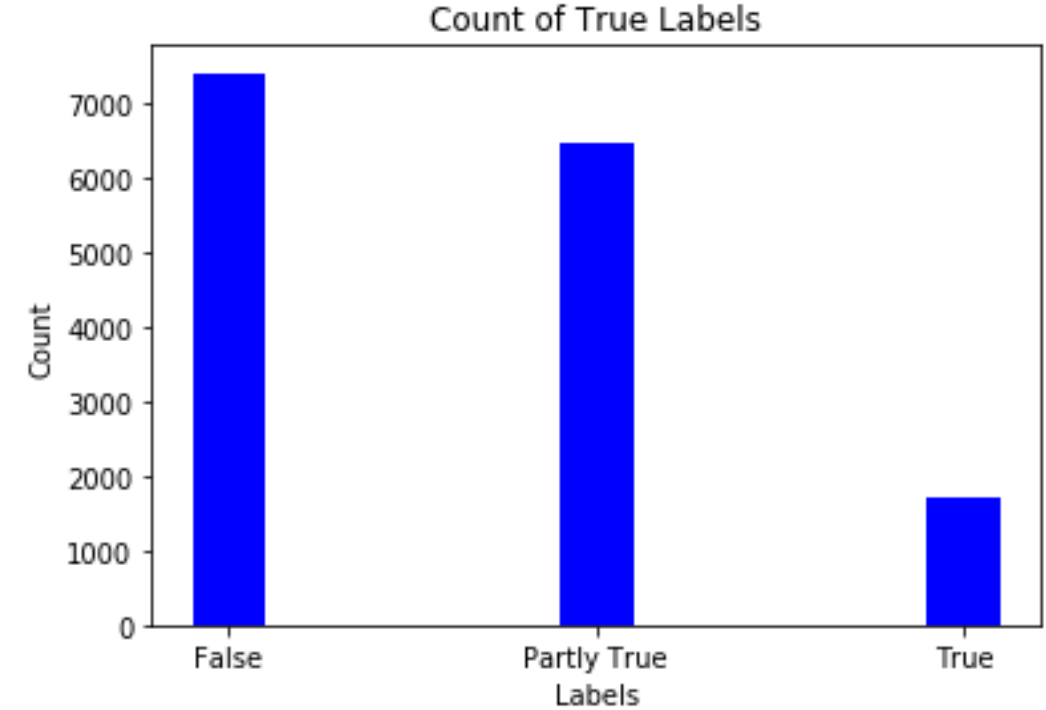
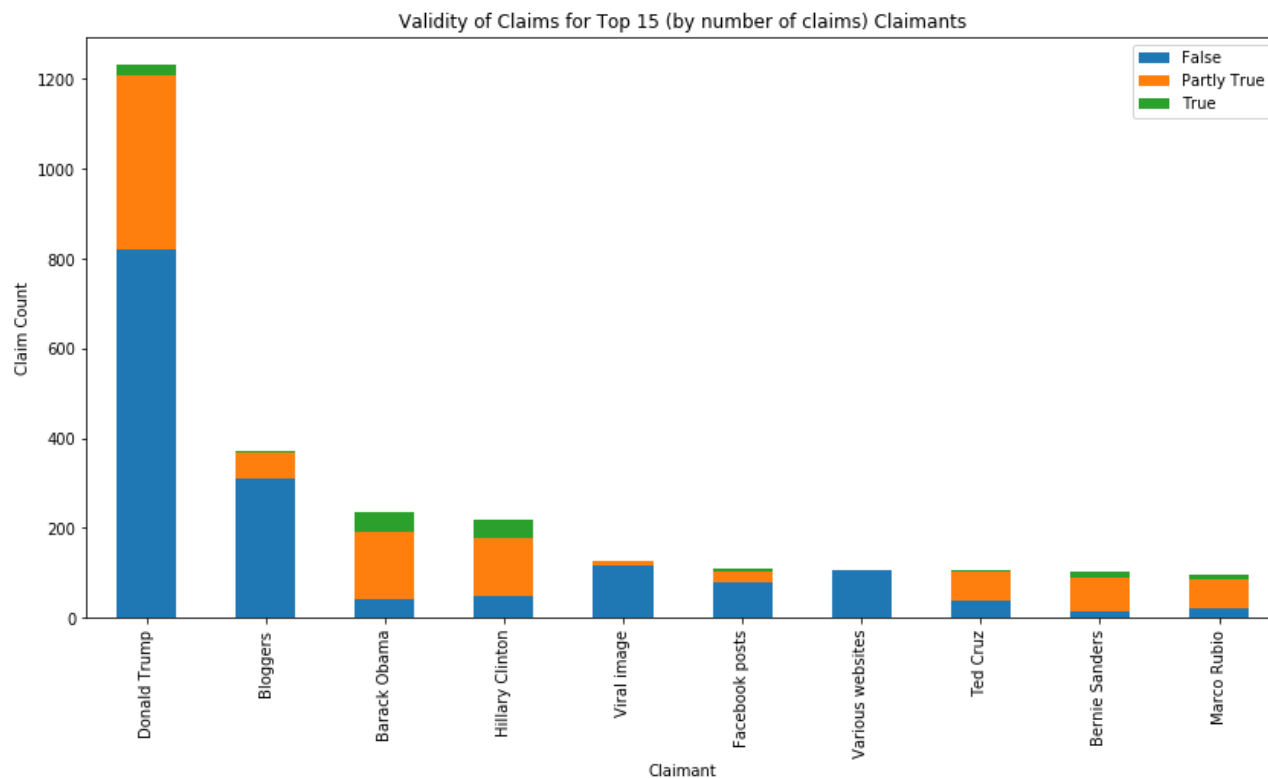**How can you be protected?!**
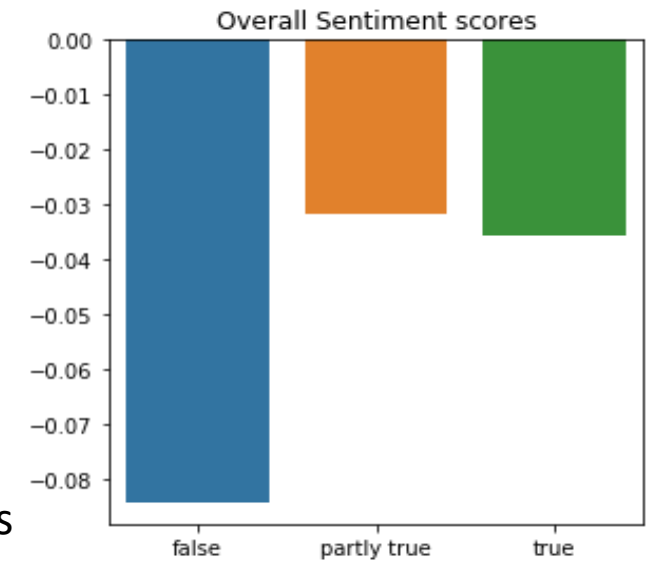
DATA VISUALIZATION

Truth Rating Distribution
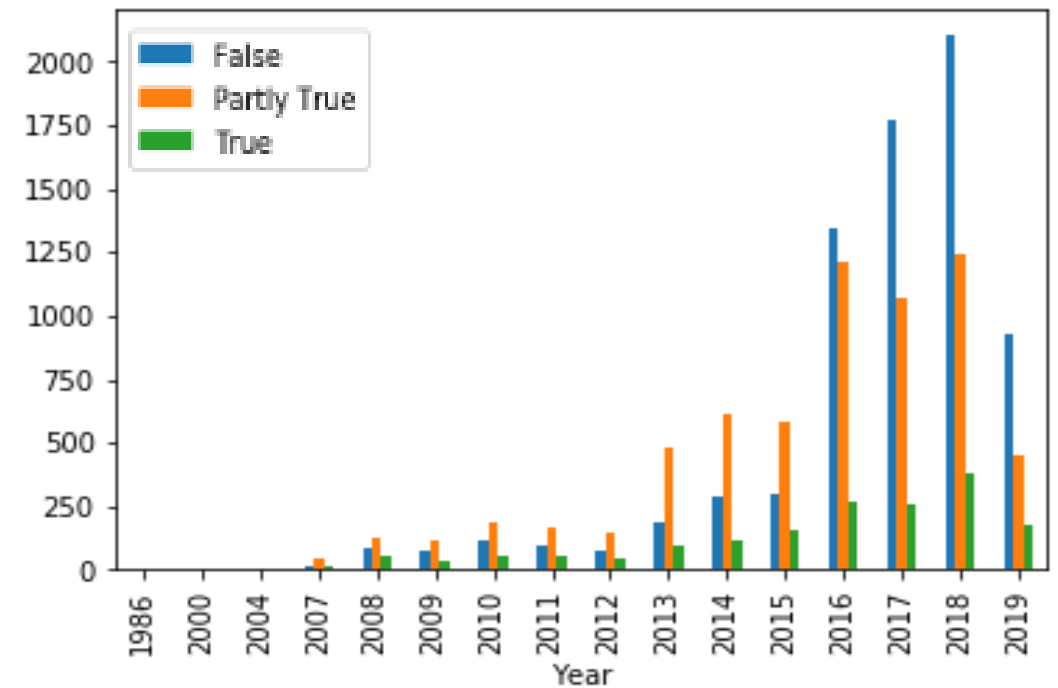

Raw Claimants
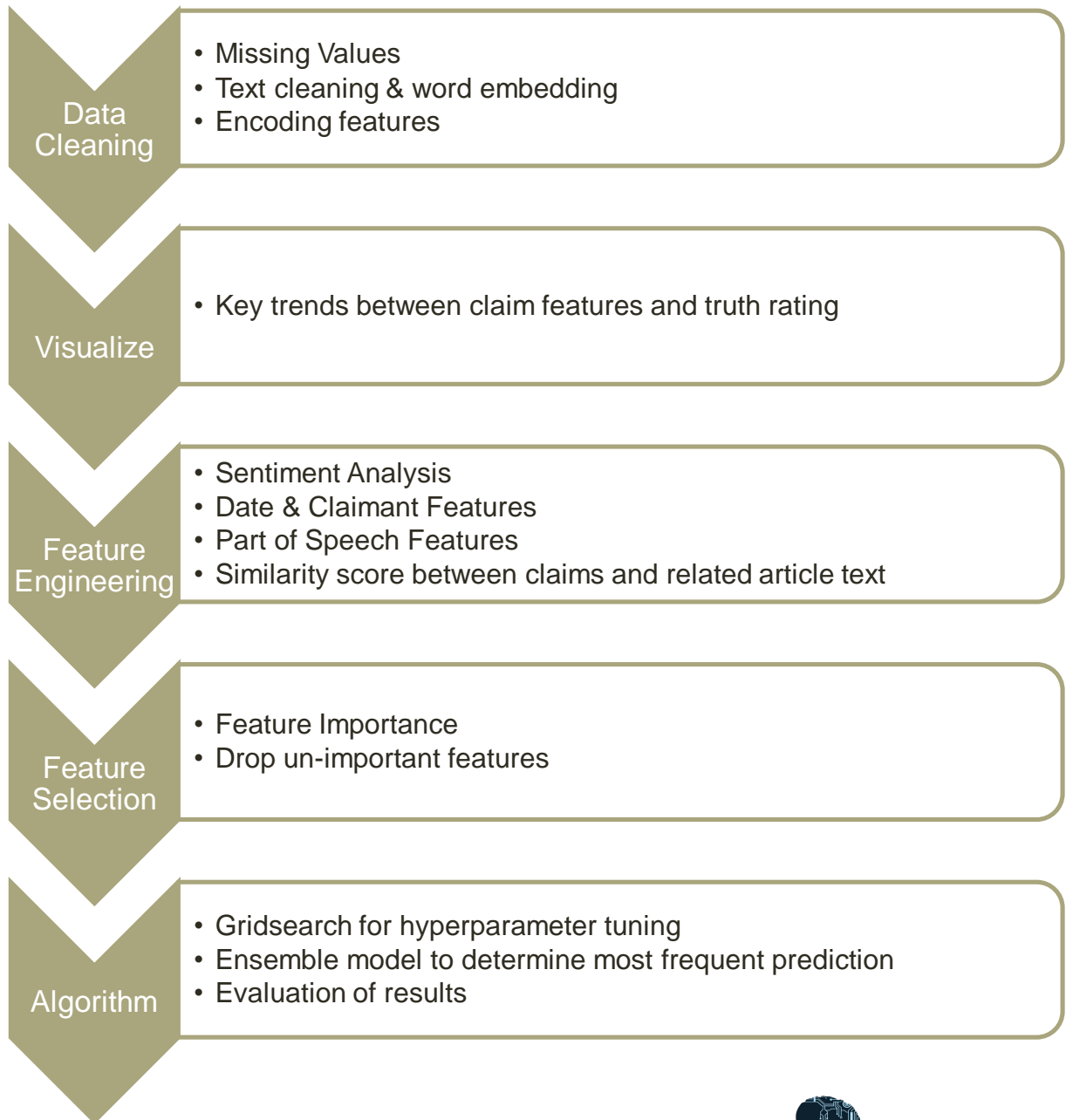
DATA VISUALIZATION

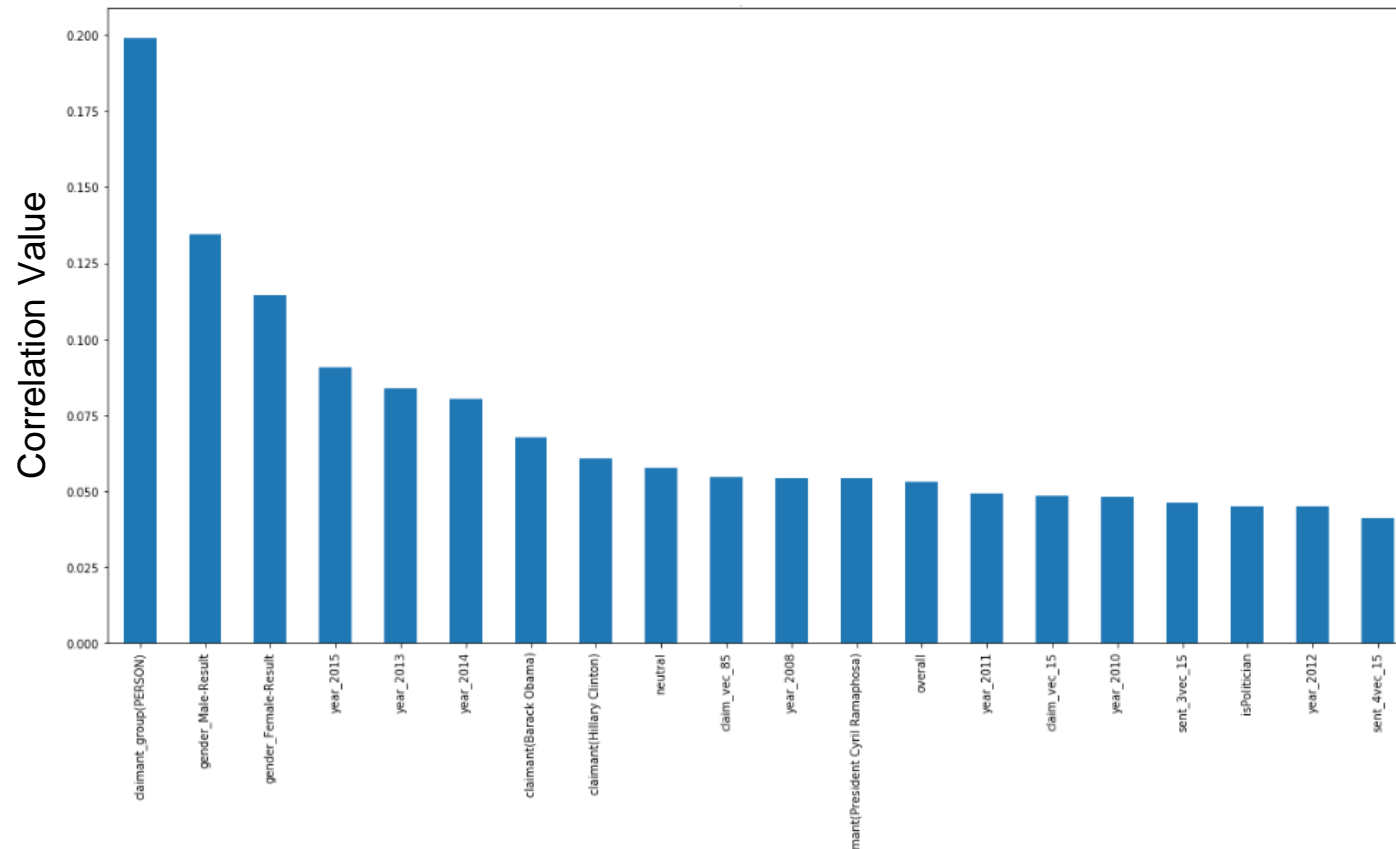Sentiment Analysis

Claimants

Dates Trends

METHODOLOGY

Features Created:

➢ Claimant group (individuals vs organization)
➢ Claimant gender (male/female)
➢ Year, month, days of the week
➢ Sentiment scores (neutral, positive, negative, overall)
➢ Similarity scores between claim and articles
➢ Parts of Speech

Data Cleaning
- Missing Values
- Text cleaning & word embedding
- Encoding features

Visualize
- Key trends between claim features and truth rating

Feature Engineering
- Sentiment Analysis
- Date & Claimant Features
- Part of Speech Features
- Similarity score between claims and related article text

Feature Selection
- Feature Importance
- Drop un-important features

Algorithm
- Gridsearch for hyperparameter tuning
- Ensemble model to determine most frequent prediction
- Evaluation of results

FINAL FEATURES

Correlation for the top 20 features

Top 20 features

Top Features

- Claimant as Individuals
- Claimant Gender
- Years
- Specific Claimants
- Sentiment scores

THE ENSEMBLE METHOD

Models

- Logistic Regression
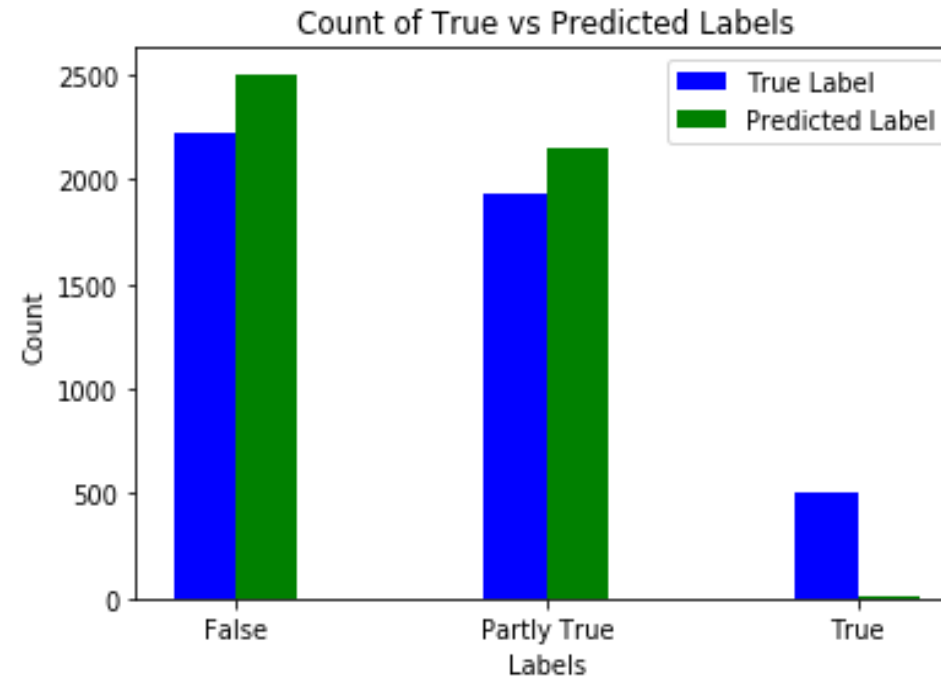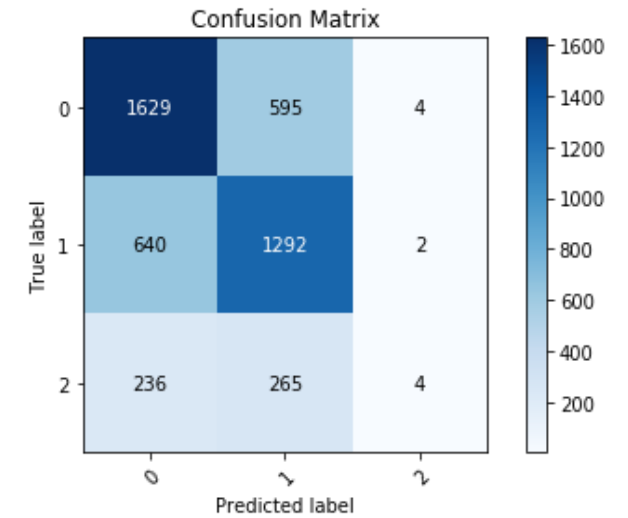
- K-NN

- Decision Trees

- SVM

Benefits

- Combine results from multiple models
- Advantage from multiple methods and mitigate pitfalls from one single model
- Improved prediction!

THE ALGORITHM

Performance

Performance Metric	Algorithm Score
Accuracy	63%
Precision	70%
Recall	47%
f1 Score	45%
Run time (cost)	85min (Data Cleaning); 400 min (Model Implementation)

Outputs

CONCLUSION & INSIGHTS

- Improved content management:
  - at least 63% reduction in fake news content

- Higher rate of others sharing/recommending your platform

- Growing and robust reputation and market value!

FURTHER WORK

- Optimize to improve precision

- Additional hyperparameter tuning for each model

- Additional Feature Engineering
  - Populate unknown claimants

- Deep Learning

We look forward to discussing further developments and optimizing this algorithm further to suit your needs.

Questions?

DATA CUP

54	news_Busters	0.313396