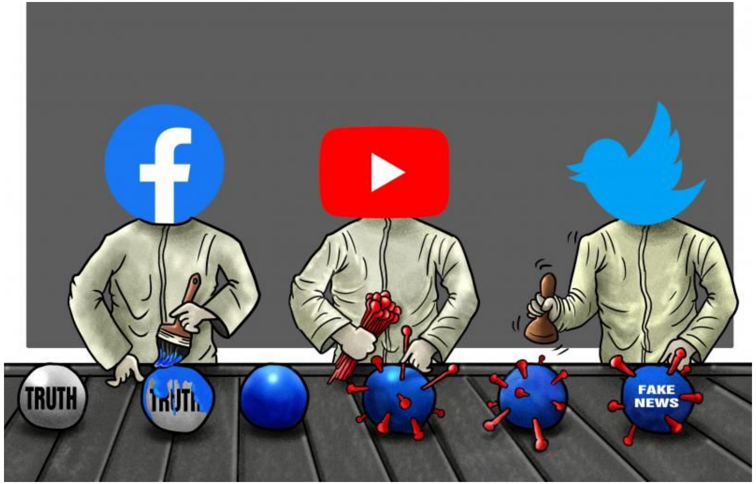
Fake News Detection Using Machine Learning

Dmytro Vremenko and Stefan Philip

Problem Description

- People rely on social media for news and current events
  - Not typically moderated for the correctness of content
- Can be weaponized to . . .
  - Sway public opinion
  - Political polarization
  - Consuming attention
- Fake news can tap into cognitive biases and spread quicker than real news
- Real life examples:
  - Outbreak of COVID-19
  - Russo-Ukrainian War



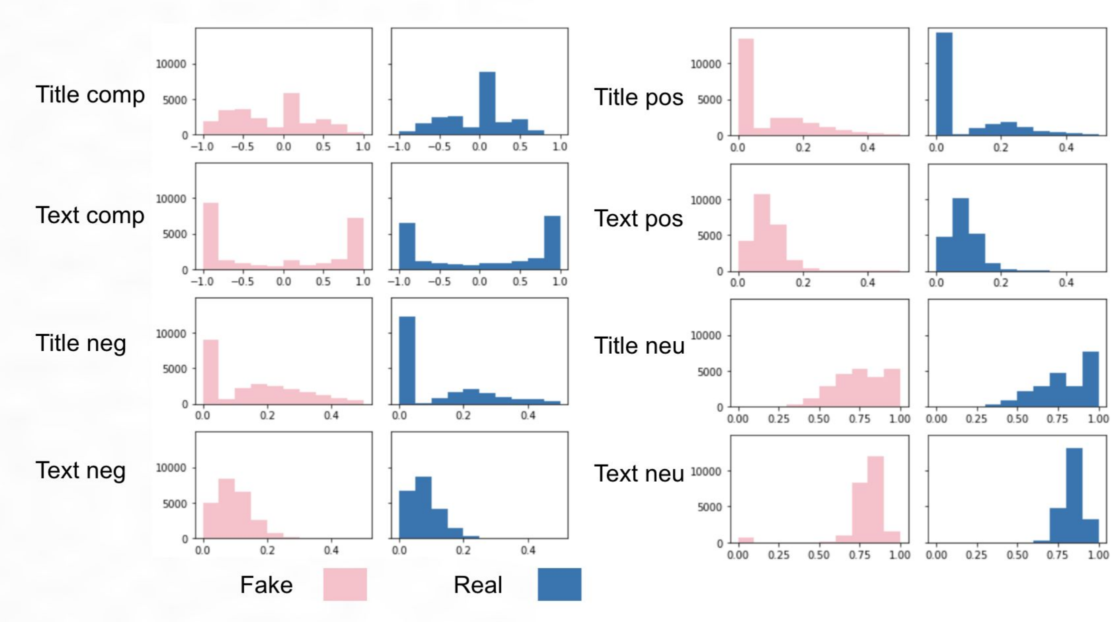
Dataset

- University of Victoria ISOT Research Lab Fake News Dataset:
  - 21,417 **True** news instances (from Reuters.com)
  - 23,481 **Fake** news instances (from sources flagged by Politifact)
  - 4 columns: date, title, text, subject
  - 8 distinct subjects: "world-news", "politics-news", "government-news", "middle-east", "us-news", "left-news", "politics", and "news"

title	text	subject	date
Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

Feature Extraction

- Sentiment Analysis:
    - Positive [0, 1]
    - Negative [0, 1]
    - Neutral [0, 1]
    - Compound [-1, 1]
    - Separately for title and text
    - Added 8 features



Feature Extraction

- Term Frequency-Inverse Document Frequency (TF-IDF) Analysis:
  - Top 2,000 weighted terms from both titles and text
  - Fitting on training data only
  - Transformed training and testing data using the weighting
  - * Removed term “reuters” from the text set
  - Added 3,999 features

Top 20 *Title* Term Weights

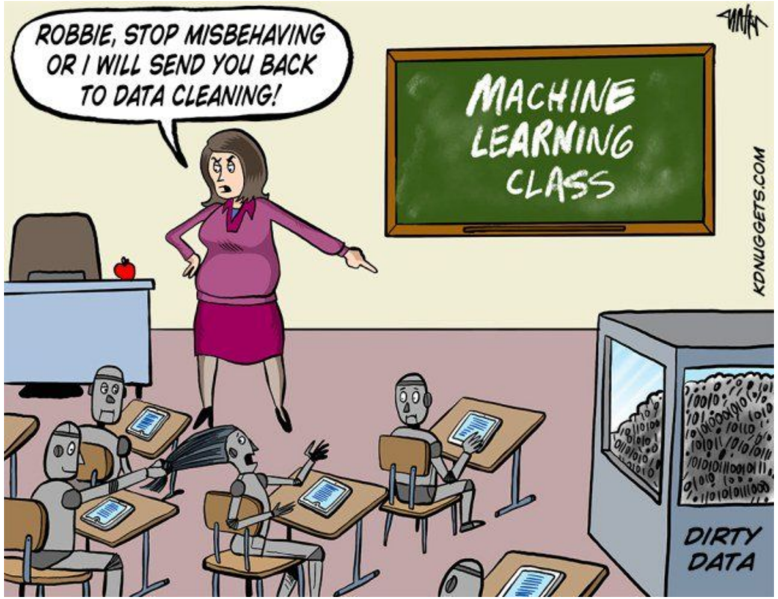
feature	weight
hat	8.057185
founder	8.057185
known	8.023284
football	8.023284
hey	8.023284
club	8.023284
ferguson	8.023284
blocking	7.990494
kurdistan	7.990494
goldman	7.990494
allegedly	7.990494
judges	7.990494
delays	7.990494
deported	7.990494
singapore	7.990494
player	7.990494
backfires	7.990494
catholic	7.990494
prisoners	7.990494
problems	7.990494

Top 20 *Text* Term Weights

feature	weight
quot	10.325869
js	6.620460
hariri	6.440875
flint	6.324615
taiwan	6.012835
catalan	5.999531
pope	5.969160
cuban	5.895052
catalonia	5.879304
christie	5.852327
epa	5.844751
ireland	5.826059
hannity	5.815010
soros	5.815010
manafort	5.811353
carson	5.771992
kushner	5.764999
parenthood	5.764999
rohingya	5.751158
id	5.747728

Model Training

- 8 Classification Models:
    - Naive Bayes
    - KNN (k=30)
    - Logistic Regression
    - Decision Tree (IG)
    - Decision Tree (GI)
    - Random Forest
    - AdaBoost
    - MLP Classifier
- Model Evaluation:
    - Accuracy
    - Error
    - Precision
    - Recall
    - F1 score
    - ROC curve
    - AUC
    - Confusion matrix

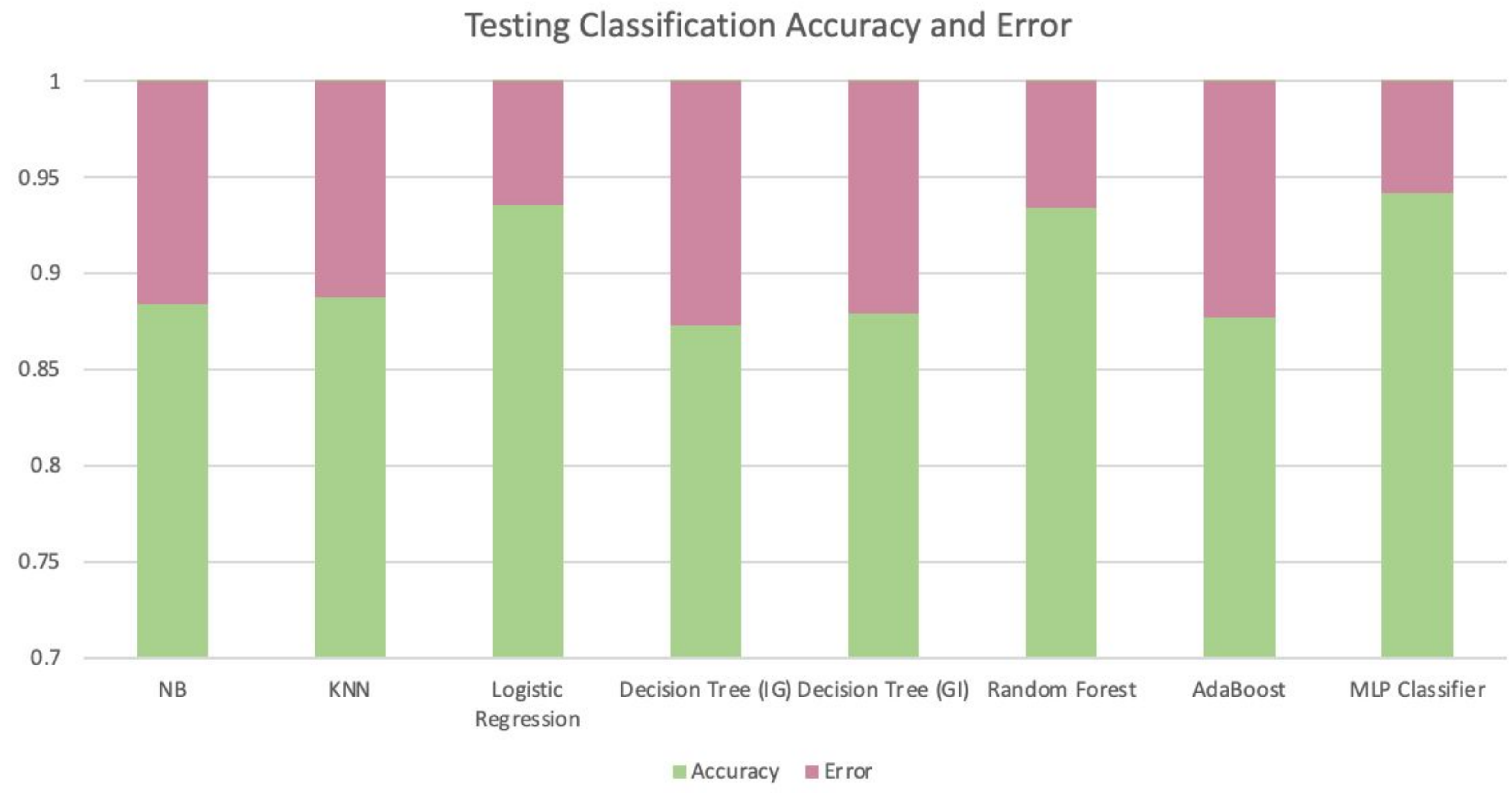


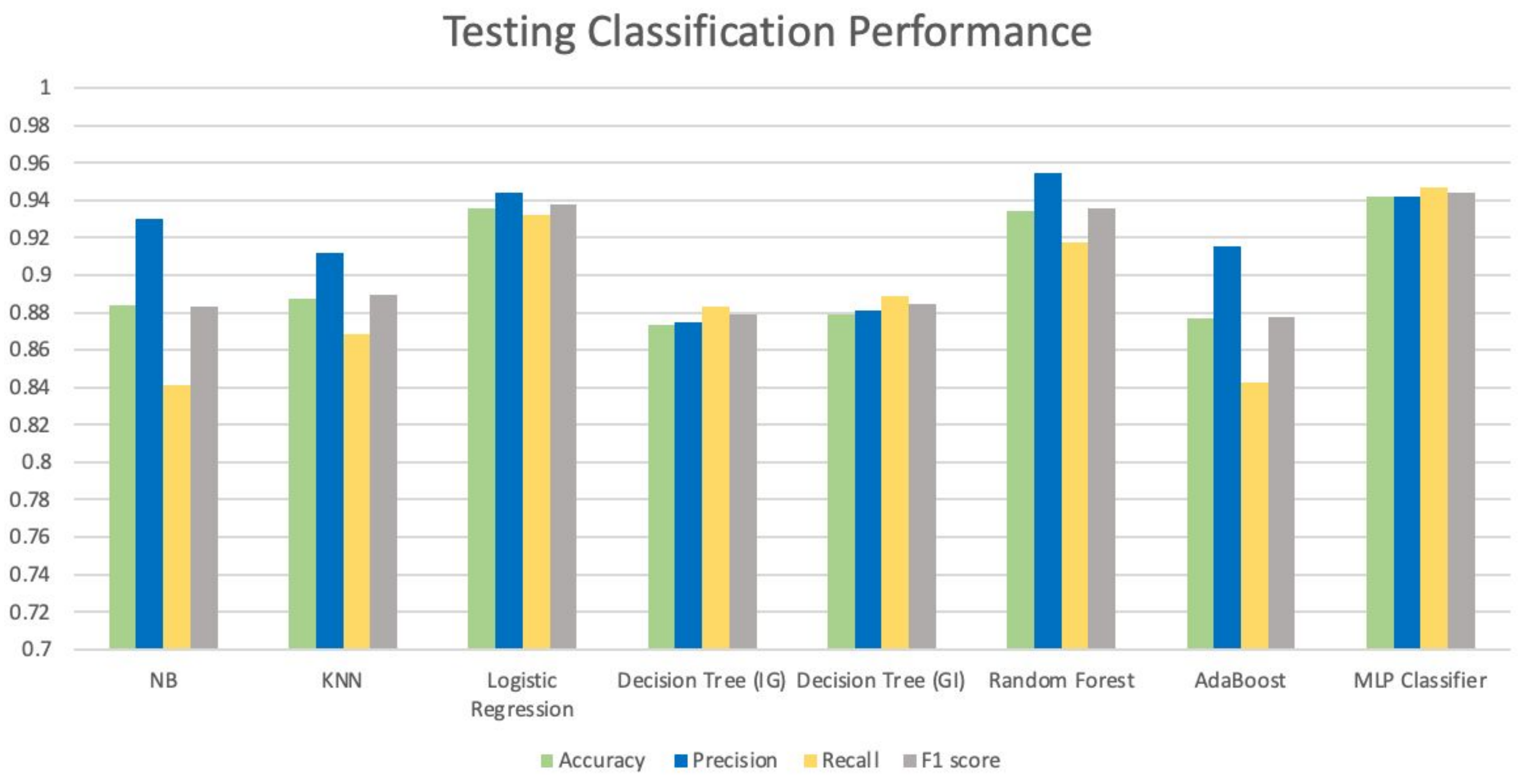
Random Forest Feature Importance

Top 30 Most Important Features

Sentiment Analysis

feature	importance	feature	importance
text: vehicle	0.062308	title: watch	0.010085
title: video	0.059252	title: trump	0.009928
title sa comp	0.025180	title: obama	0.009186
text sa neu	0.021194	title sa pos	0.007946
text: higher	0.015865	text: opponents	0.007766
text: russian	0.015847	title: breaking	0.006073
text sa neg	0.015111	title: gop	0.006062
title sa neg	0.014692	text: biggest	0.005916
title: says	0.014518	text: gets	0.005351
title sa neu	0.014181	title: donald	0.004060
title: hillary	0.013598	title: us	0.004046
text: warning	0.011733	title: north	0.003893
text sa comp	0.010907	text: hours	0.003624
text sa pos	0.010749	title: house	0.003508
text: town	0.010604	title: black	0.003359


Testing Classification Accuracy and Error
1
0.95
0.9
0.85
0.8
0.75
0.7
NB
KNN
Logistic Regression
Decision Tree (IG)
Decision Tree (GI)
Random Forest
AdaBoost
MLP Classifier
Accuracy
Error


Testing Classification Performance
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
NB
KNN
Logistic Regression
Decision Tree (IG)
Decision Tree (GI)
Random Forest
AdaBoost
MLP Classifier
Accuracy
Precision
Recall
F1 score

Limitations

- Single Source of True Articles
  - Maybe the models pick us the writing styles of people who work there

Future Directions

- Diversify source of real news
- Expand to other languages
- Expand to audio/video