Can we detect fake news online?

Team 28 - Star Silva, Hui Qi, Mingcan Li, Tongzhao Liu, Rebekkah Ismakov

Reference: https://github.com/AIRLegend/fakenews/tree/master

Background

According to a 2019 Pew study, most Americans see made-up news as detrimental, negatively impacting trust in societal institutions, public health, and democratic stability.

We wanted to identify fake news by only looking at text - that is, without images or social network analysis. We built classification models using machine learning methods and try to improve the accuracy of the prediction of the models.

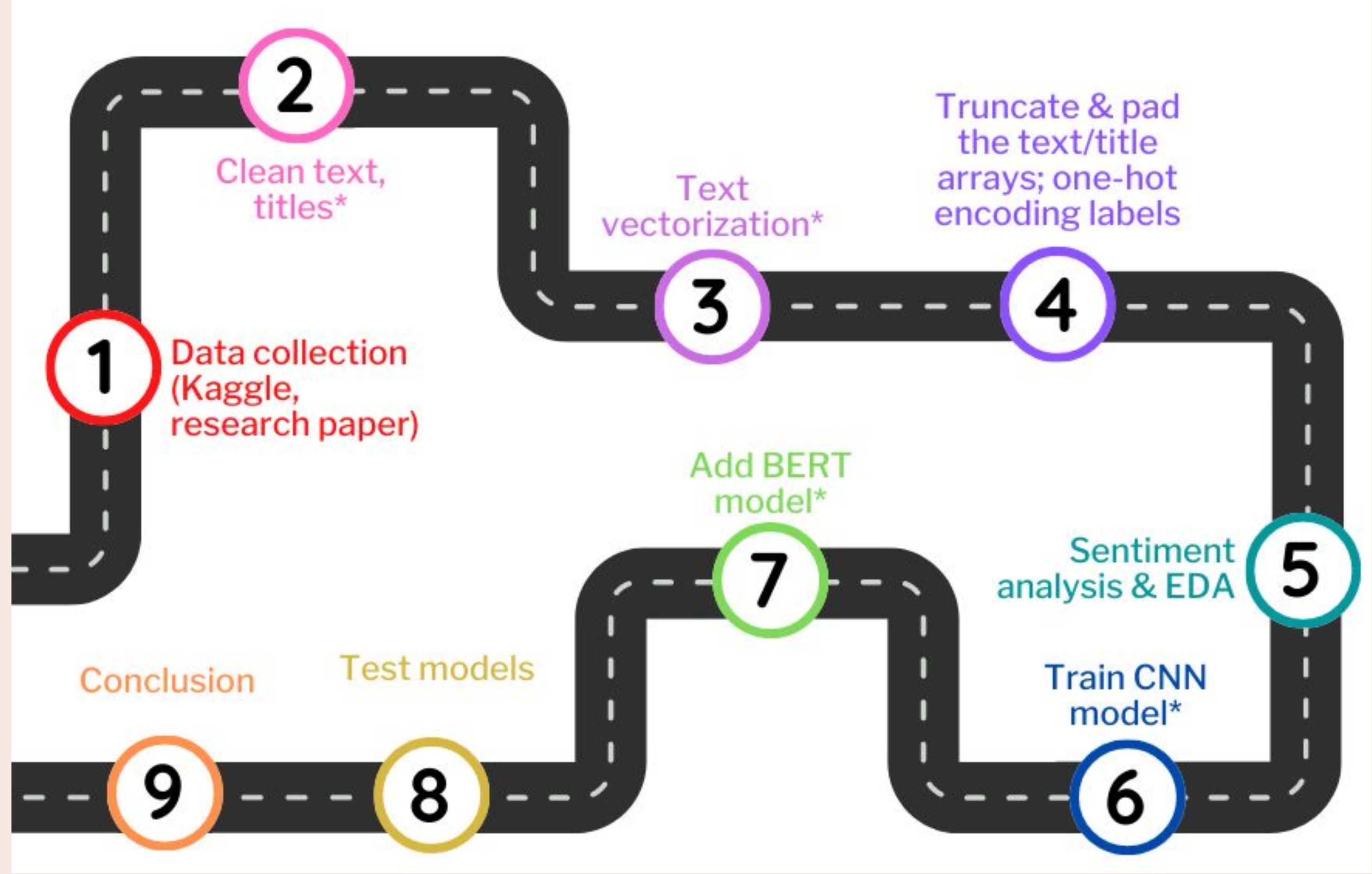
Data

We merged 3 datasets for training models and 1 dataset to test the models:

- DATASET 1. **True & fake** news, 20k rows, from 240+ websites
- DATASET 2. **All fake**, 13k rows, from 244 websites
- DATASET 3. **All true,** 38k rows
- DATASET 4. **TI_CNN**, 20k rows, true & fake, only for testing

Process

* = where we improved upon original



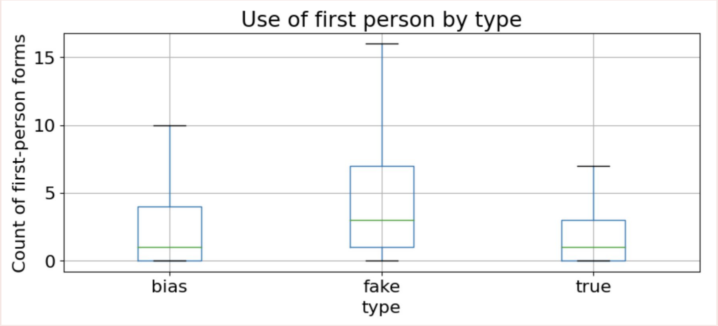
Library/Model used in process:
Step 3: KeyedVectors model vocabulary
(GoogleNews-vectors-negative300.bin)

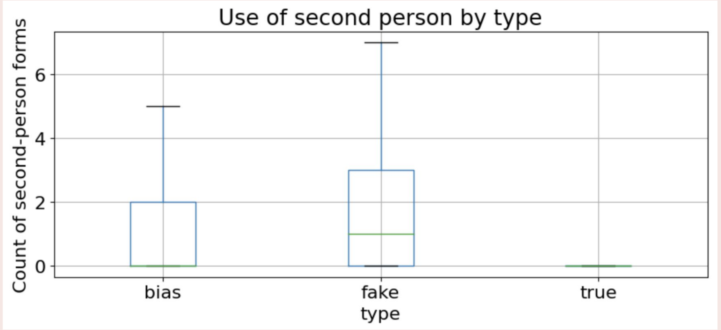
Improvements from GitHub group:
Step 1: larger dataset used for training
Step 2: lower text, translate emoji to its corresponding text
Step 3: nltk library
Step 6: Bayesian Optimization
Step 7: Bert Model

Exploratory data analysis

We used sentiment analysis to discover links between the narrative point of view, mood and whether there is any difference in these fields for true/fake news.

One key insight: Fake news tends to express the news using the first and second person. But true news almost never uses the second person.

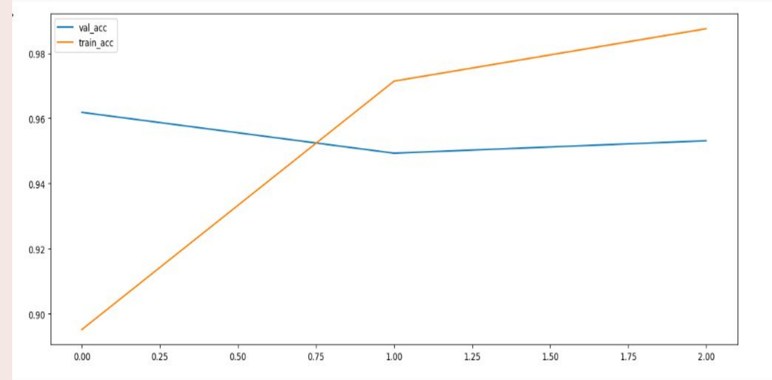




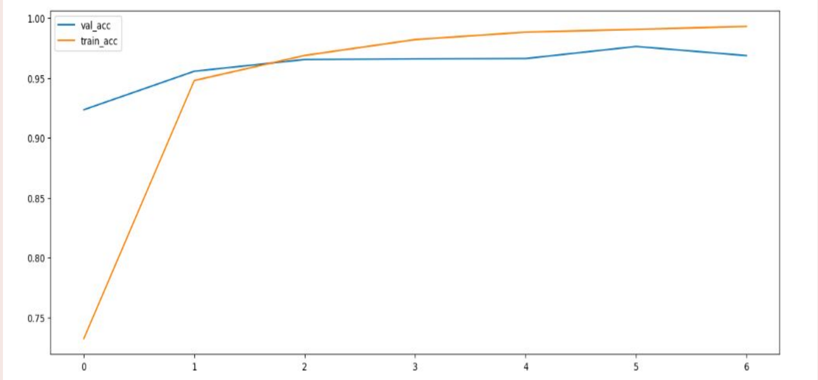
Models

1. CNN Based architecture

- **Dataset Split: 72-8-20**
- **3 training models:** initial model from GitHub, model with hyperparameters from Bayesian Optimization, retrain model with expending training data
- **Hyperparameter Tuning:** learning rate, momentum, dropout1, dropout2, filtersTitle, denseTitle, filtersContent, denseContent
- **Performance Metrics**: old_cnn_model achieved a test accuracy of 94.65%, new_cnn_model achieved a test accuracy of 97.61%. AUC-ROC score: 1



old_cnn_model accuracy

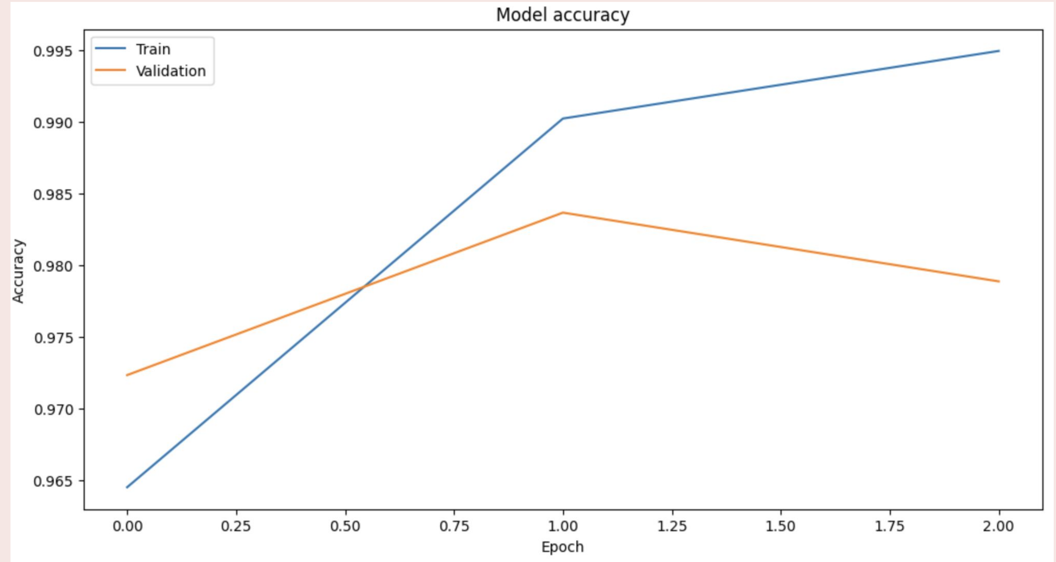


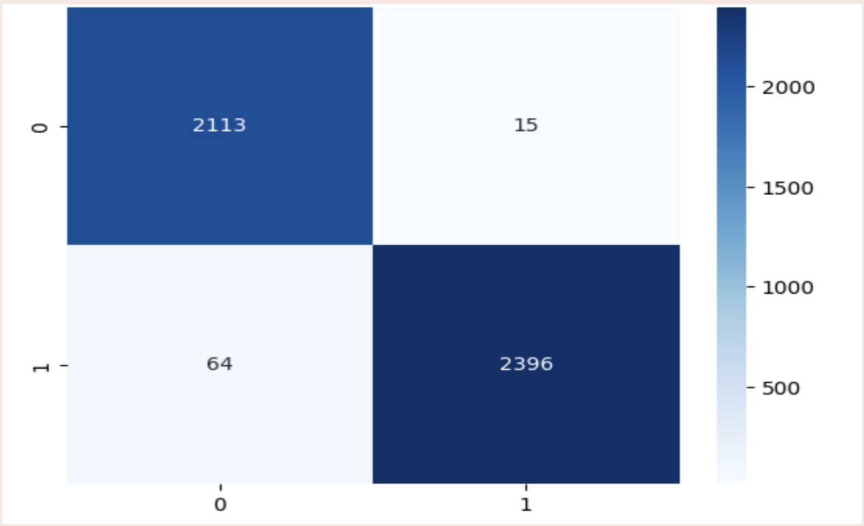
new_cnn_model accuracy

Test on TI_CNN dataset	old_cnn_model	new_cnn_model (Bayesian)	retrained_cnn_model (expand training data)
Accuracy	72.90%	77.41%	77.59%

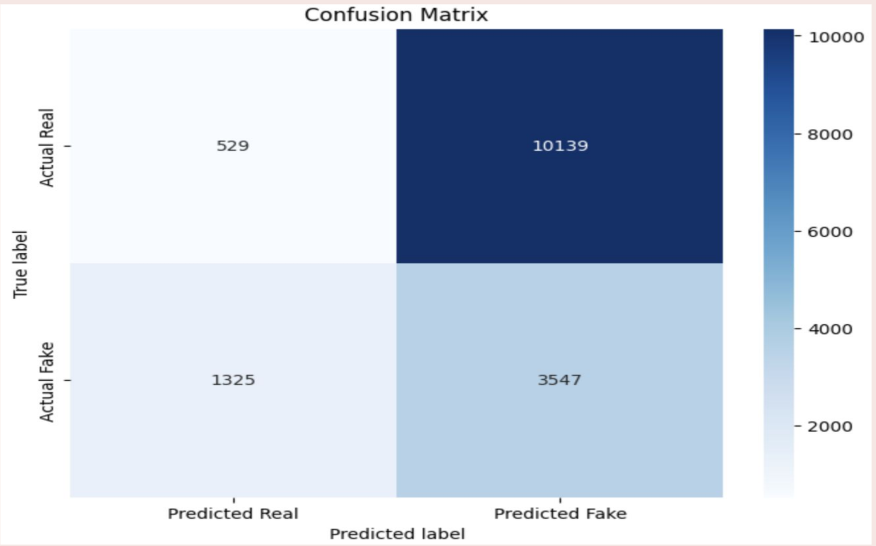
2.. BERT Based architecture

- **Dataset Split**: 80-10-10 (merge dataset) --> 24,500 fake, 21,400 real
- **Hyperparameter Tuning**: train model with learning_rate =2e-05, batch_size =16, epochs =3, max_length=128
- **Performance Metrics**: F1 score :**98.7%**





- **Test on TI-CNN Dataset:** Highly skewed: 10,668 fake, 4,872 real
- **Performance Metrics**: AUC-ROC score: 0.2625
- Model exhibited bias towards classifying news as fake.



Limitations

- Binary classification is may not accurately catch partially true articles
- Prioritizing for recall would result in more false positives

Conclusion:

Our CNN model achieved 97.6% accuracy over 94.7% on merging datasets
Our CNN model achieved 5% accuracy improvement on TI_CNN dataset