Fake News Detection

Benjamin Toler
January 30, 2023

Summary

- Text classification model that identifies news articles as fake or real with 96% accuracy
- Fake articles often about political figures

Overview

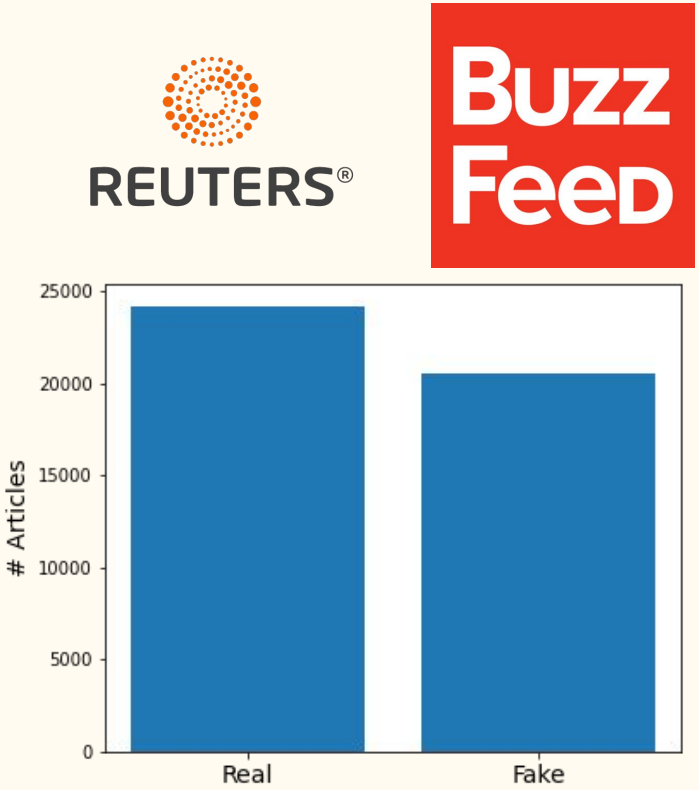
- Business Problem
- Data
- Modeling
- Results
- Conclusions
- Next Steps

Business Problem

- Facebook is trying to cut down on the spread of misinformation on its platform
- A classification model will improve their efficiency in identifying false articles to be removed
- Increase customer trust in platform
- Prevent potential legal issues and regulatory intervention

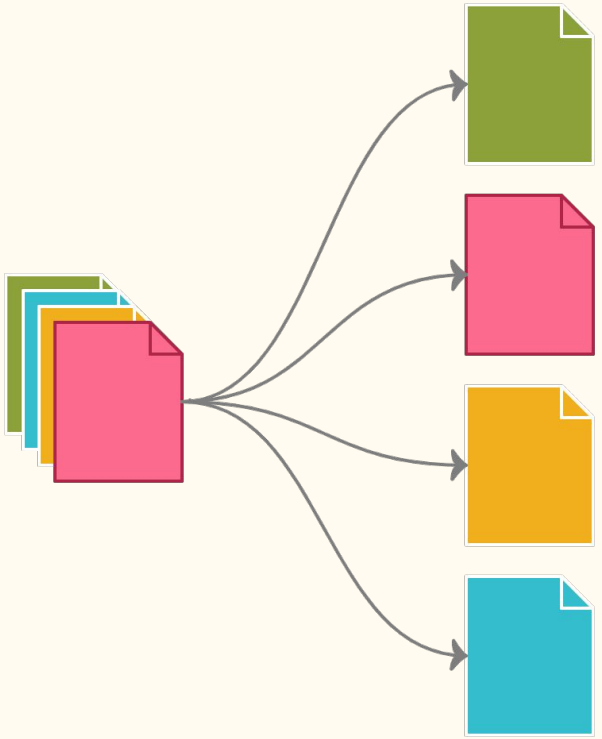
Data

- 44,706 news articles
  - Reuters/ BuzzFeed
  - Text data
- Balanced dataset:
  - 54% real news
  - 46% fake news



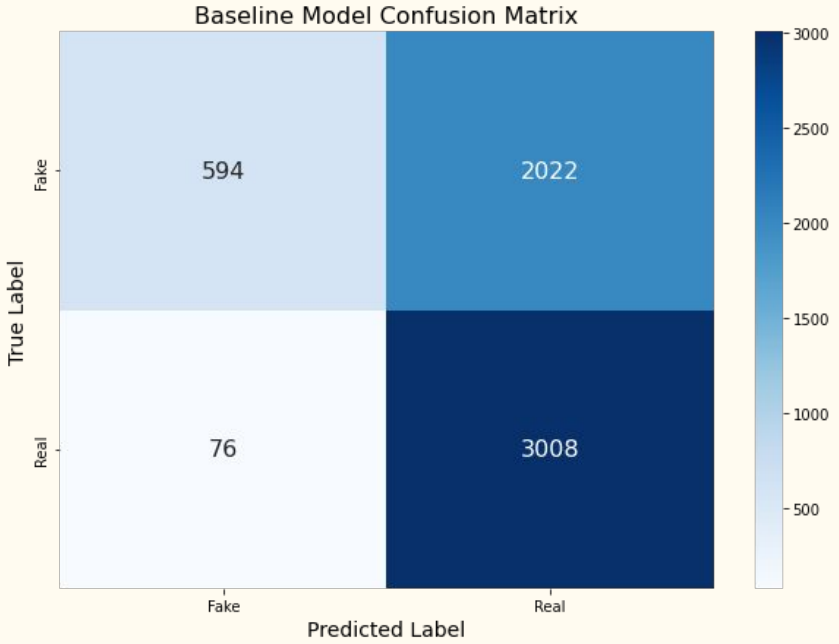
Model

- Text Classification
- Iterative Approach
  - Feature Engineering
  - Classifier Selection
  - Hyperparameter tuning
- Evaluation metrics:
  - Accuracy
- Final Model:
  - Count vectorizer
  - XGBoost Classifier

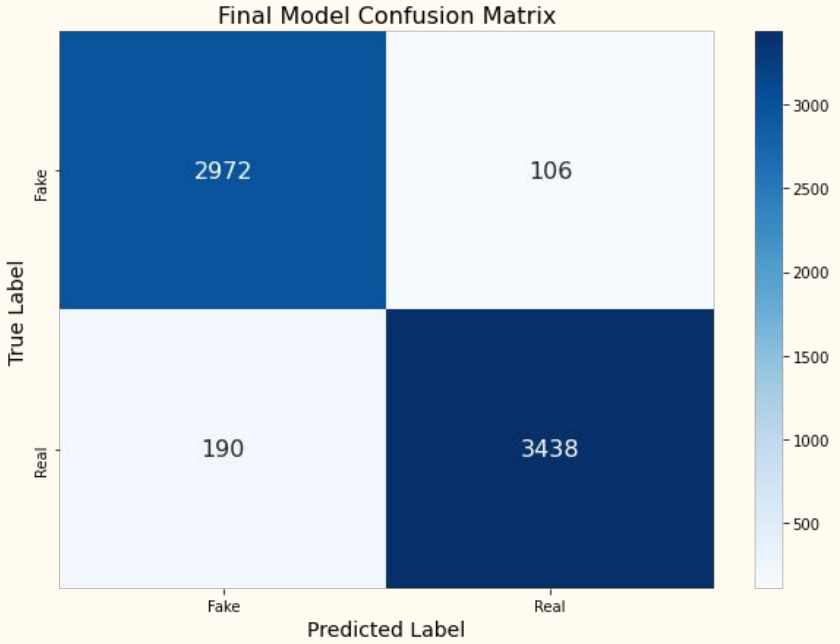


Results

- Baseline Model Accuracy: 64%

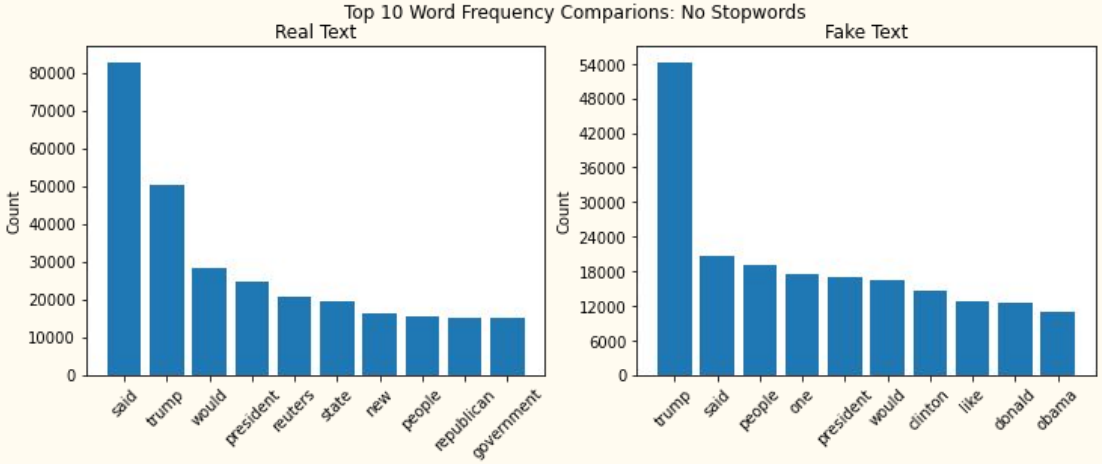


- Final Model Accuracy: 95%



Conclusion

- Use model to detect and remove fake news articles posted on platform
- Provide warnings to customers about articles with high references to political figures
- Use model to provide verification tag to real articles



Next Steps

- Expand dataset to non-political news articles
- Try other text classification techniques to further improve accuracy
- Determine if model can be developed using other article features such as title or author

Thank You!

Questions?

- Email: bentoler22@gmail.com
- Github: @bentoler22
- LinkedIn: www.linkedin.com/in/benjamintoler22