

Supervised Machine Learning Ensemble for Fake News Classification of Social Media



[10]

Mohammed Mosuily | MSc Artificial Intelligence | Supervisor: Dr Stuart Middleton | **UNIVERSITY OF Southampton**

Introduction

What is Fake news?
False information or stories that appear to be news, typically spread using social media outlets. Which could be influencing political views or misleading users opinions. Types of fake news, satire, clickbait, propaganda, or mistakes. [3][4]
Two categories for false news:
• Disinformation –shared or created deliberately [3]
• Misinformation –might have been shared without the intention to mislead [3]
Why this project matters?
As figure 1 shows, fake news possibly will intensify the conflict between people and cause a rift in society. Also, false information could lead to endangerment of public health such as the case with the COVID-19. pandemic. [1][4][5]

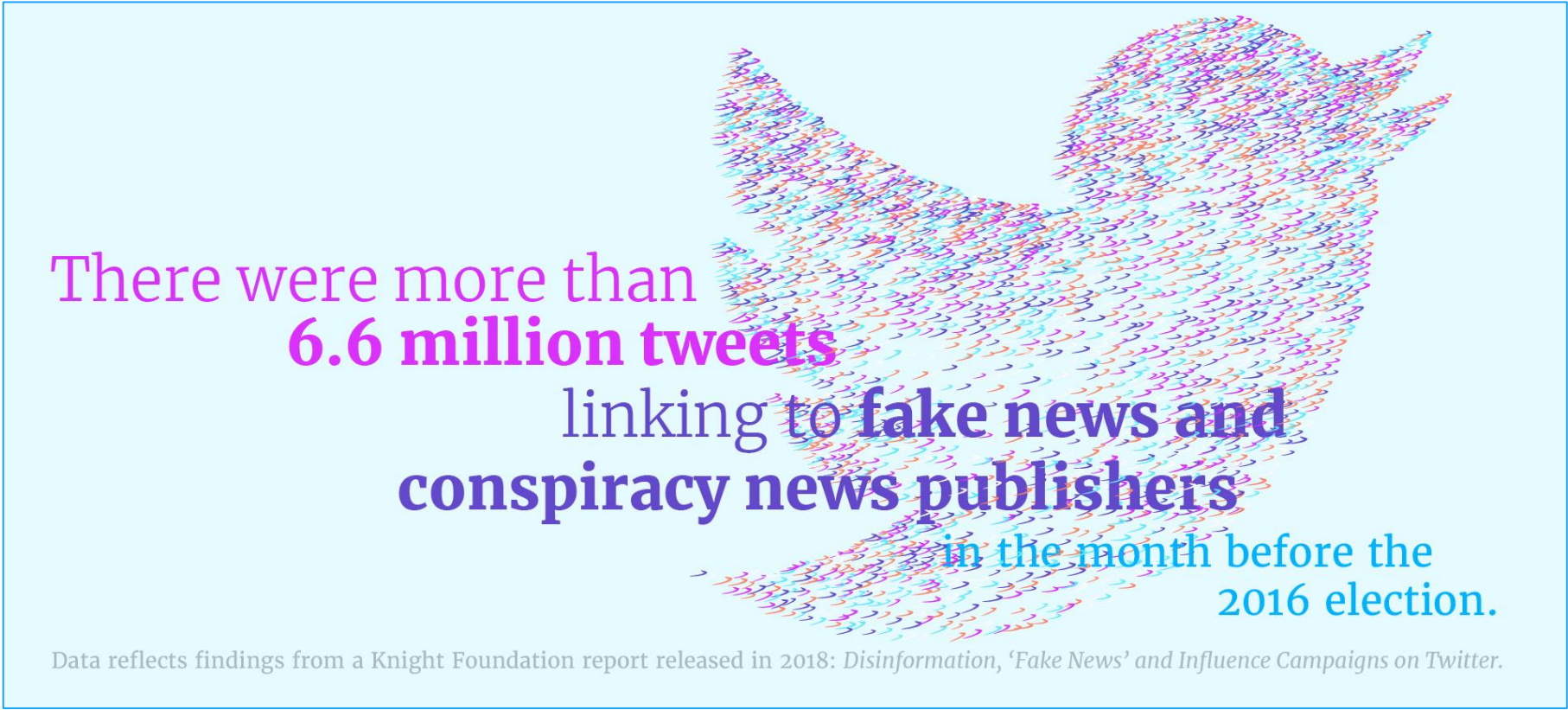


Figure 1: 2016 US election fake news tweets estimation[3]

Ensemble methods in machine learning:
Instead of only using one model to predict an outcome, ensemble methods takes multiple models into account. For example, using Support Vector Machines (SVM) and Logistic Regression (LR) together. Creating multiple subsets for each model and then aggregate them together at the end. Therefore, the prediction usually has higher accuracy than only using one model. [9]

Research Question

Which feature types are most important to current ensemble models for Fake News user classification? How robust are current ensemble models across different news topics and different user posting styles?

Target Re-implementation Paper

An Ensemble Model Using N-grams and Statistical Features to Identify Fake News Spreaders on Twitter [8]

Background

Major themes:

- Logistic Regression and Support Vector Machines or both [6]
- Supervised machine learning methods tend to perform better than deep learning methods [6]
- Automatically identifying fake news spreaders on Twitter is feasible with high precision. However, false positives should be considered while developing a method to detect false news [6][8]

Major features used:

- n-grams (a contiguous sequence of n items from the text) [6][8]
- Stylistics (Number of mentions, URLs, hashtags, punctuation)[6]
- Personality and emotions of the author [6]
- Word embeddings

Target re-implementation paper:
Calculating the statistics from a tweet like finding the lexical diversity or the average length, combined with n-grams to detect fake news spreaders. While also using more sophisticated machine learning methods.

Evaluation Plan

Proposed Experiment/Metrics setup:

- Data: PAN 2020 Profiling Fake News Spreaders on Twitter table 1 shows the dataset [6]
- Programming Language: Python with different statistics, machine learning, and data manipulation libraries
- Ensemble with multiple classifiers, test multiple combinations of supervised machine learning classifiers using cross-validation. Figure 3 shows a proposed system architecture that should be the blueprint for the project
- Measure the Aggregate confusion matrix to find out about false positive. It is important to understand the legal and ethical consequences of accusing a truthful user of being a fake news spreader.
- Model hyperparameters manipulation

Expected result :

- Fake news account spreader detected with high accuracy, generate two lists for fake/truthful authors
- Figure 2 shows the Aggregated confusion matrix for English-Dataset almost 30%-35% false-positive, which should be reduced wit the expected results. This figure is based on the PAN 2020 competition outcome [6]
- Explore the statistical features model and find the Logistic regression coefficients for the predicted probabilities

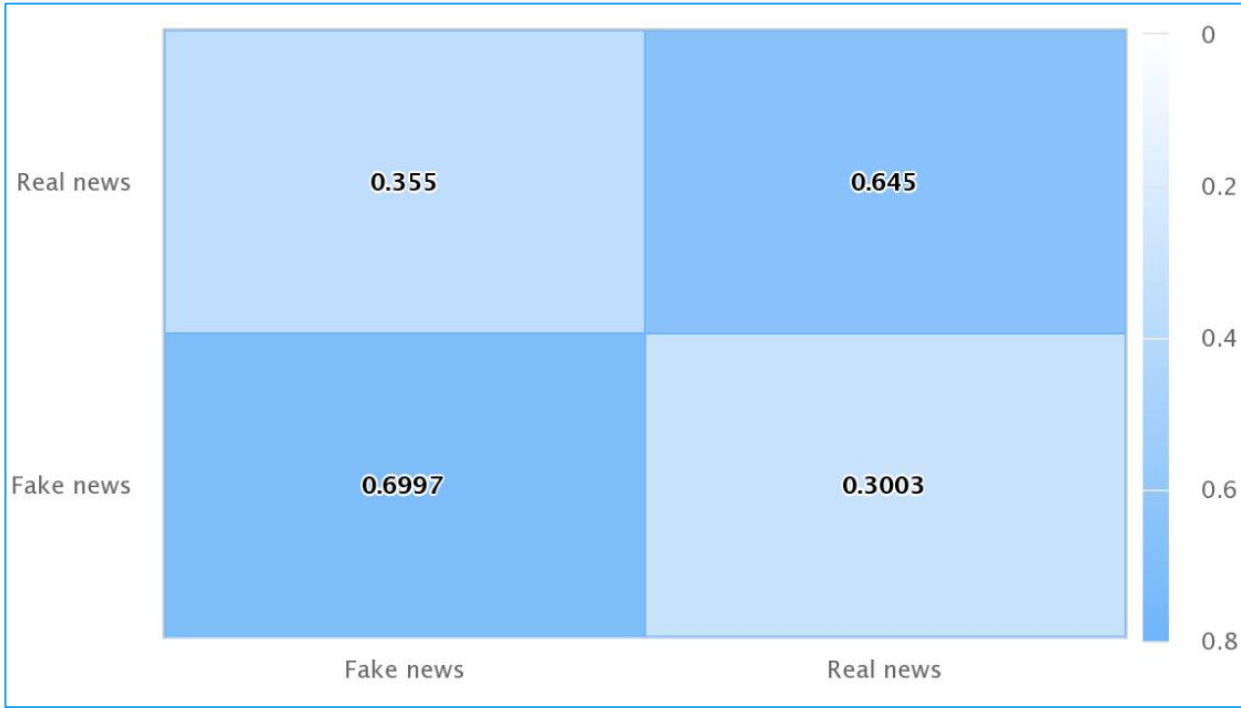


Figure 2: Aggregate confusion matrix (English-Dataset)[6]

Language	Training	Test	Total
English	300	200	500
Spanish	300	200	500

Table 1: Dataset-Number of tweet authors [6]

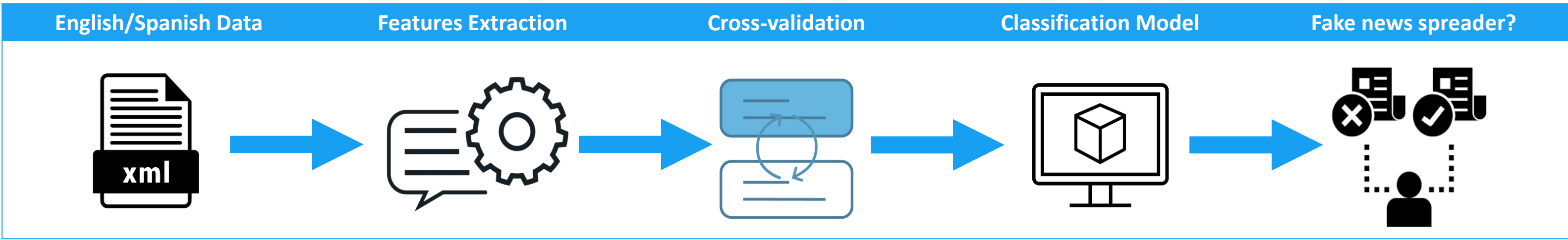


Figure 3: Proposed system architecture

Aims – Outcome – Objectives

Aim:
Detect fake news spreaders on social media using a new ensemble design
Objectives:

- Experiment and explore the performance of current ensemble methods
- Explore the trade-off between performance, accuracy, and false-positive
- Test the robustness of current ensemble models

Outcomes:

- Use an ensemble that achieves high accuracy in detecting fake news spreaders, while taking false-positive into consideration
- Combat fake news on social media platforms
- Provide a safer and truthful online environment for users

References

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, 2017, doi: 10.1257/jep.31.2.211.
[2] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8885861.
[3] J. Lionetti, "Fake News – Can it be stopped?," *Pepperdine Graphic Media*, 2019. <https://pepperdine-graphic.com/fake-news-can-it-be-stopped/>
[4] K. Foundation, "Seven ways misinformation spread during the 2016 election," *Knight Foundation*, 2018. <https://knightfoundation.org/articles/seven-ways-misinformation-spread-during-the-2016-election/> (accessed Mar. 17, 2020).
[5] O. D. Apuke and B. Omar, "Fake news and COVID-19: modelling the predictors of fake news sharing among social media users," *Telemat. Informatics*, vol. 56, no. July 2020, p. 101475, 2021, doi: 10.1016/j.tele.2020.101475.
[6] F. M. R. Pardo, A. Giachanou, B. Ghanem, and P. Rosso, "Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter," *CLEF 2020 Labs Work. Noteb. Pap.*, pp. 22–25, 2020, [Online]. Available: CEUR-WS.org.
[7] M. Lichouri, M. Abbas, and B. Benaziz, "Profiling Fake News Spreaders on Twitter based on TFIDF Features and Morphological Process Notebook for PAN at CLEF 2020 Profiling Fake News Spreaders on Twitter based on TFIDF Features and Morphological Process," no. September, 2020.
[8] J. Buda and F. Bolonyai, "An Ensemble Model Using N-grams and Statistical Features to Identify Fake News Spreaders on Twitter Notebook for PAN at CLEF 2020," no. September, pp. 22–25, 2020, [Online]. Available: <https://github.com/pan-webis-de/bolonyai20>.
[9] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, pp. 1–18, 2018, doi: 10.1002/widm.1249.
[10] HT Correspondent, "Govt withdrawal of fake news press release victory of democracy, media: Congress," *Press Trust of India*, 2018. <https://www.hindustantimes.com/india-news/govt-withdrawal-of-fake-news-press-release-victory-of-democracy-media-congress/story-CU4s7FzdxU6Fed0JtZse6K.html>.