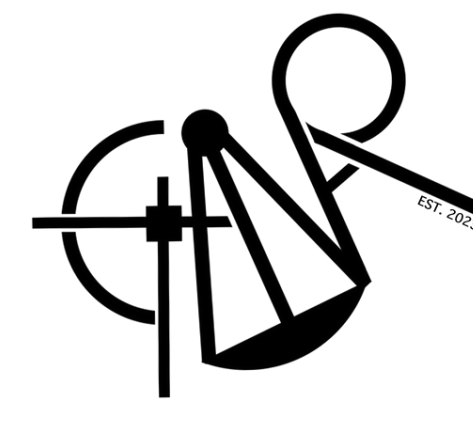
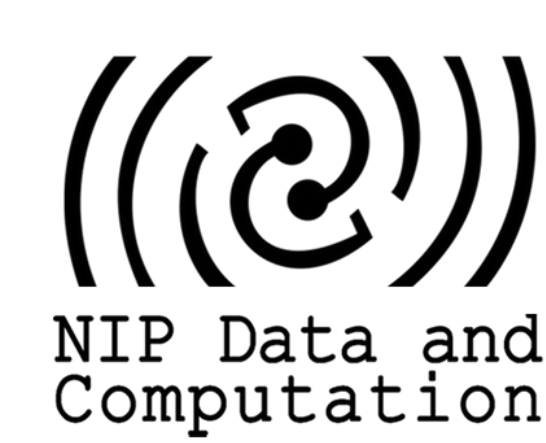


Fake News Detection in Philippine News Corpus using LDA and Sentiment Analysis with Machine Learning

Renz Jamuel B. Jamen and Reinabelle C. Reyes

Data and Computation Research Group, National Institute of Physics
College of Science, University of the Philippines Diliman, Quezon City, Philippines



MOTIVATION AND OBJECTIVES

The massive societal impacts brought by Fake News necessitate continuous research and development to combat it with available machine-learning techniques. In this work, we present an alternative approach to classify fake news articles from a Philippine-based dataset. We evaluate the impact of the addition of **dimensionality reduction and sentiment ratio extracted using VADER (Valence Aware Dictionary for sEntiment Reasoning)** on the performance of selected supervised machine-learning models.

METHODOLOGY

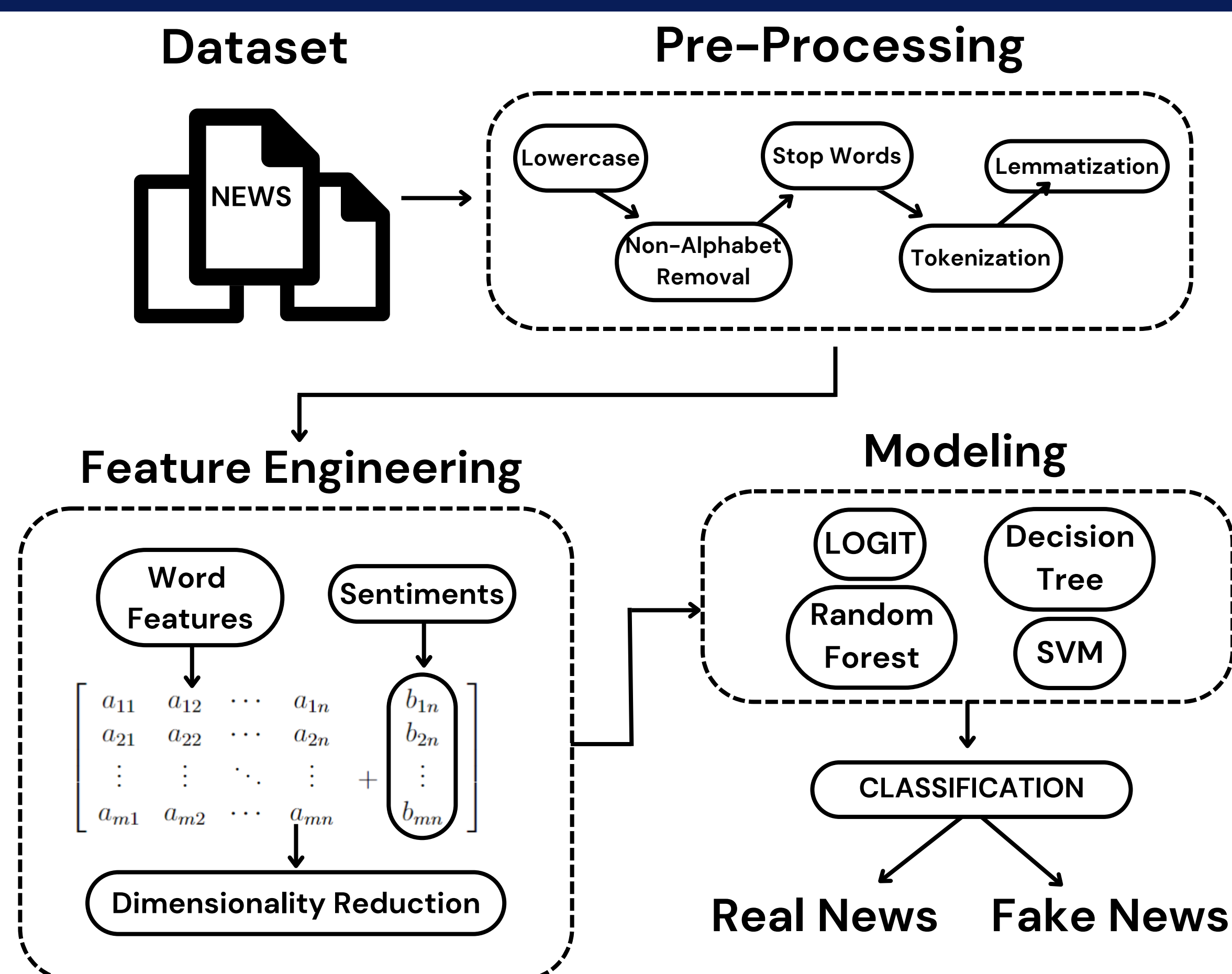


Figure 1. Flowchart illustrating development of machine learning models for Fake News classification

- The text data were cleaned and divided into a **70:30 train-test ratio**.
- Word features were represented as unigrams using the **term frequency-inverse document frequency (TF-IDF)**. We considered the top 10,000 word features and then applied filtering methods such that constant features were dropped. This resulted in a reduction to **1035 relevant features**, which served as our baseline set of features.
- Existing literature, e.g.[1], successfully utilized polarity-based sentiment analysis for fake news detection. In this study, we used **VADER** [2] to assign each article a **sentiment ratio**, which we defined as the sum of the negative and positive scores divided by the sum of the neutral scores in each article.
- We employ **Linear Discriminant Analysis (LDA)** to reduce the high feature vector dimension to a **single feature**, equal to one (1) less than the number of classes.
- We selected the four simple and most commonly-used supervised machine-learning models: **Logistic Regression (LR)**, **Support Vector Machines (SVM)**, **Decision Tree (DT)**, and **Random Forest (RF)** to facilitate faster computation and easier interpretability while also ensuring straightforward replication of our study.
- We performed **five-fold cross-validation** using **GridSearchCV** on the training set. The process was repeated for all values in the grid and the optimal values for each model with the highest cross-validation accuracy were selected.

PHILIPPINE FAKE NEWS CORPUS

We utilized this dataset compiled by Aaron Carl Fernandez from Mapua Institute of Technology, Philippines [3]. We randomly sampled around 65% or 14,725 (**7,625 Real and 7,100 Fake**) **local English** articles for model development. All news samples from both classes belong to the “**Nation**” news category from **January 1, 2016 to October 31, 2018**.

RESULTS AND DISCUSSION

The **final SVM model** achieved the highest accuracy of **93.78%** using the **single LDA feature** from the combination of **TF-IDF and sentiment ratio** from the PH dataset.

Table 1. Performance metrics of SVM model using different feature combinations from the PH Fake News dataset

Model	Features	No. of Features	Hyperparameters	Cross-validation Accuracy	Test Set			
					Accuracy	Precision	Recall	F1-score
SVM	TF-IDF (Baseline)	1035	C = 0.1	88.528 (0.003)	87.87	85.48	87.89	86.67
	TF-IDF + sentiment ratio	1036	C = 0.1	88.409 (0.003)	87.89	85.49	87.95	86.70
	LDA feature (applied to TF-IDF)	1	C = 1	93.835 (0.007)	93.35	91.05	94.45	92.72
	LDA feature (applied to TF-IDF + sentiment ratio)	1	C = 1	94.171 (0.007)	93.78	91.77	94.61	93.17

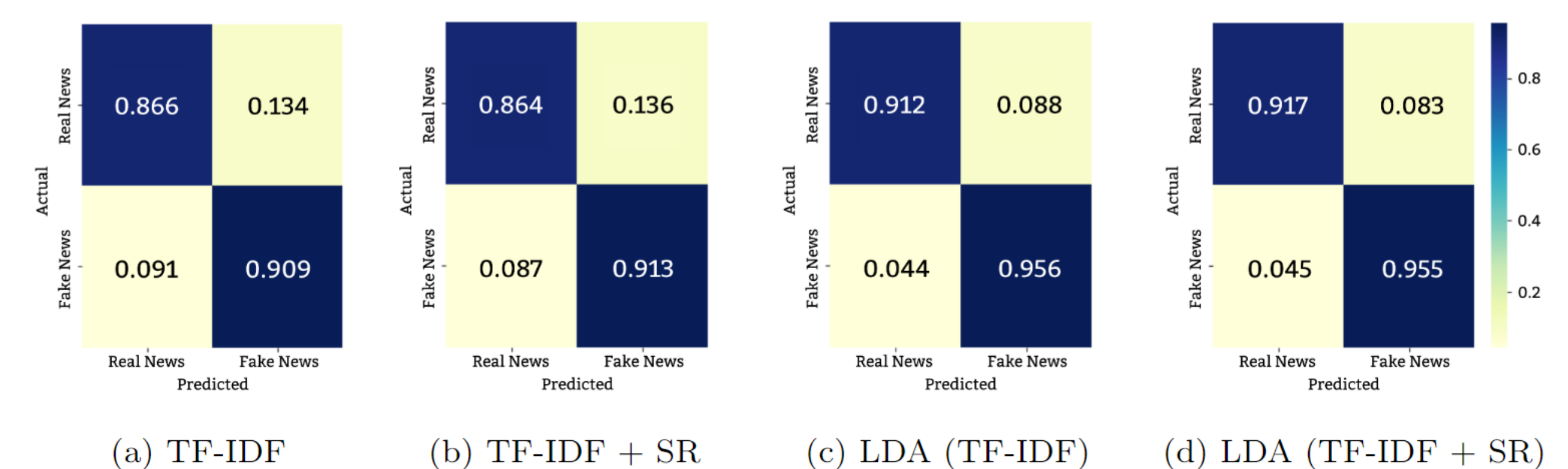


Figure 2. Confusion matrices for the testing set using different feature combinations with the SVM model

Using a single LDA feature led to significant improvements in model performance shown by the **significant reduction in false positives and false negatives**. Figure 3a illustrates the discriminative power of LDA in effectively separating Real and Fake News articles within the test dataset, with **well-separated Gaussian curves and distinct means**. Figure 3b shows the sentiment ratio distribution for Fake News articles is skewed right with a slightly higher mean as compared to Real News.

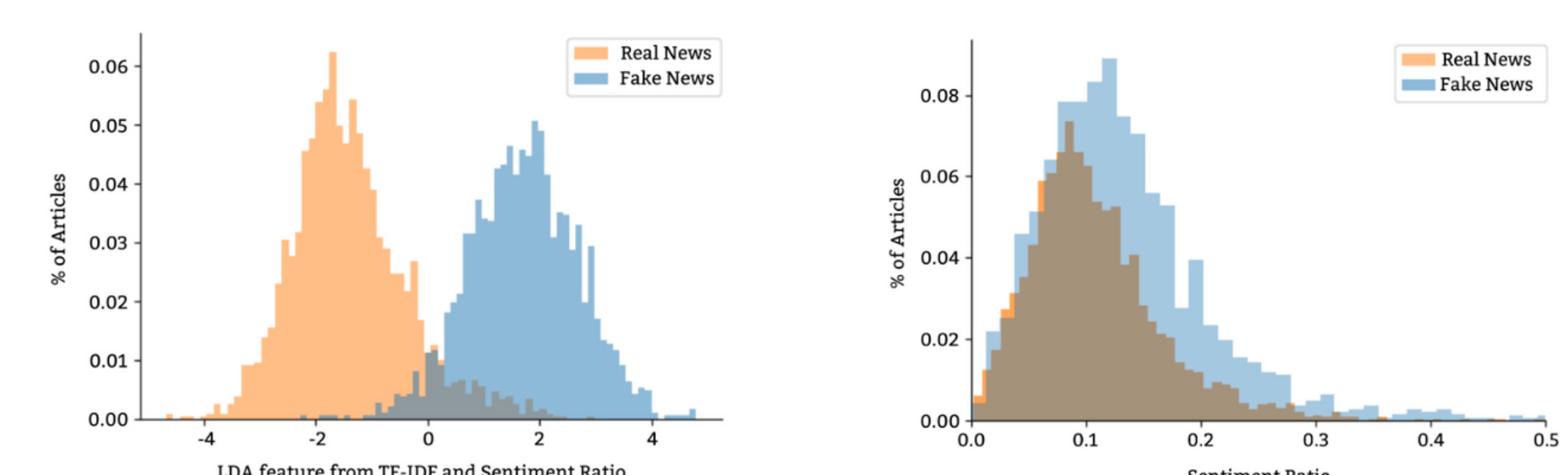


Figure 3. Histograms showing class separability between Real and Fake News through the (a) LDA feature from TF-IDF and sentiment ratio combined and (b) sentiment ratio

CONCLUSION

We leveraged **LDA and sentiment analysis** as additional feature engineering techniques for **supervised machine-learning models** to effectively classify Philippine-based online news articles as either real or fake. Our final model that uses only **one reduced feature** achieved **accuracy significantly higher than the baseline model** that relied on thousands of features and **on par with an existing model** relying on linguistic-based cues derived from the text [3].

SELECTED REFERENCES

- [1] M. A. Alonso, D. Vilares, C. Gomez-Rodriguez, and J. Vilares, Sentiment Analysis for Fake News Detection, Electronics 10, 13 (2021).
- [2] C. Hutto and E. Gilbert, VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text, in Proceedings of the 8th International AAAI Conference on Web and Social Media (Michigan, USA, 2014), 216-225.
- [3] A. Fernandez and M. Devaraj, Computing the Linguistic-Based Cues of Fake News in the Philippines Towards its Detection, in Proceedings of the 9th International Conference on Web Intelligence, Mining, and Semantics (New York, USA, 2019), 1-9.

Got questions?
Join me [here](#)

