

Figure 1: Dataset 1 article count by label and subject

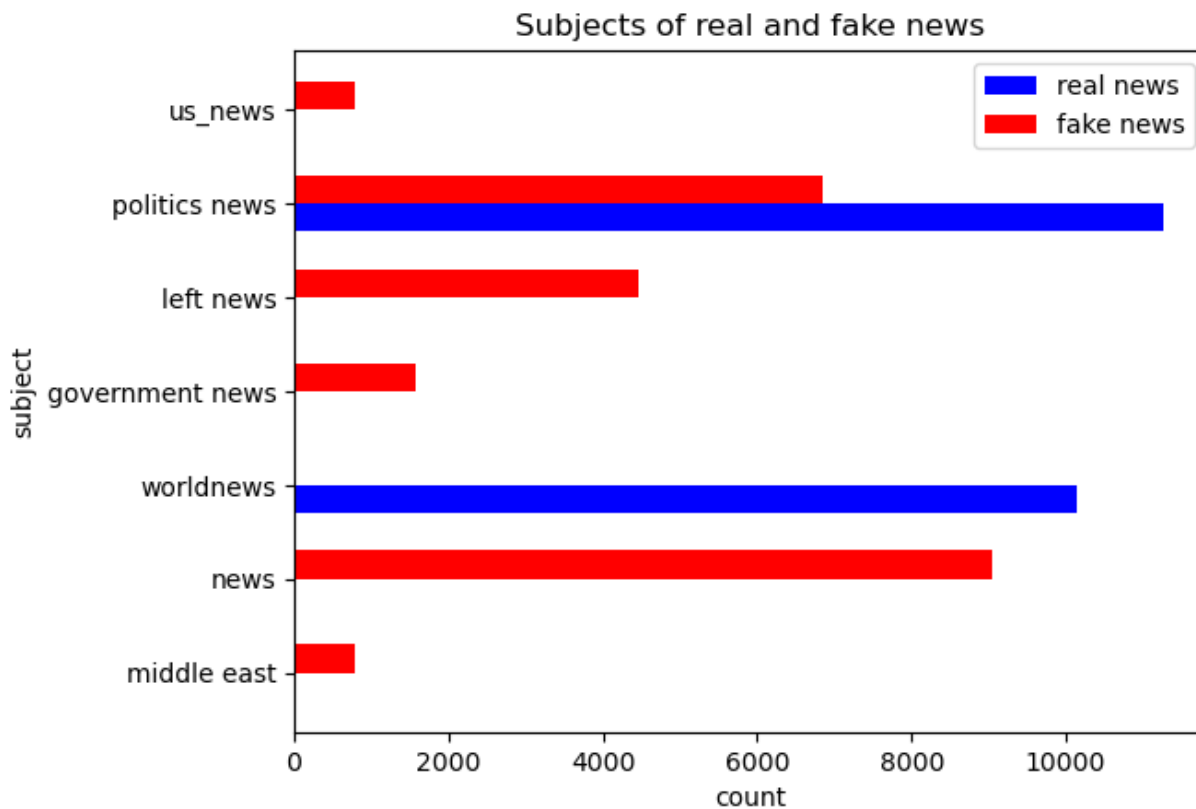


Figure 3: Frequency distributions of word counts in real and fake news articles in Dataset 1. The inner panel zooms in on a limited range for a non-compressed visual.

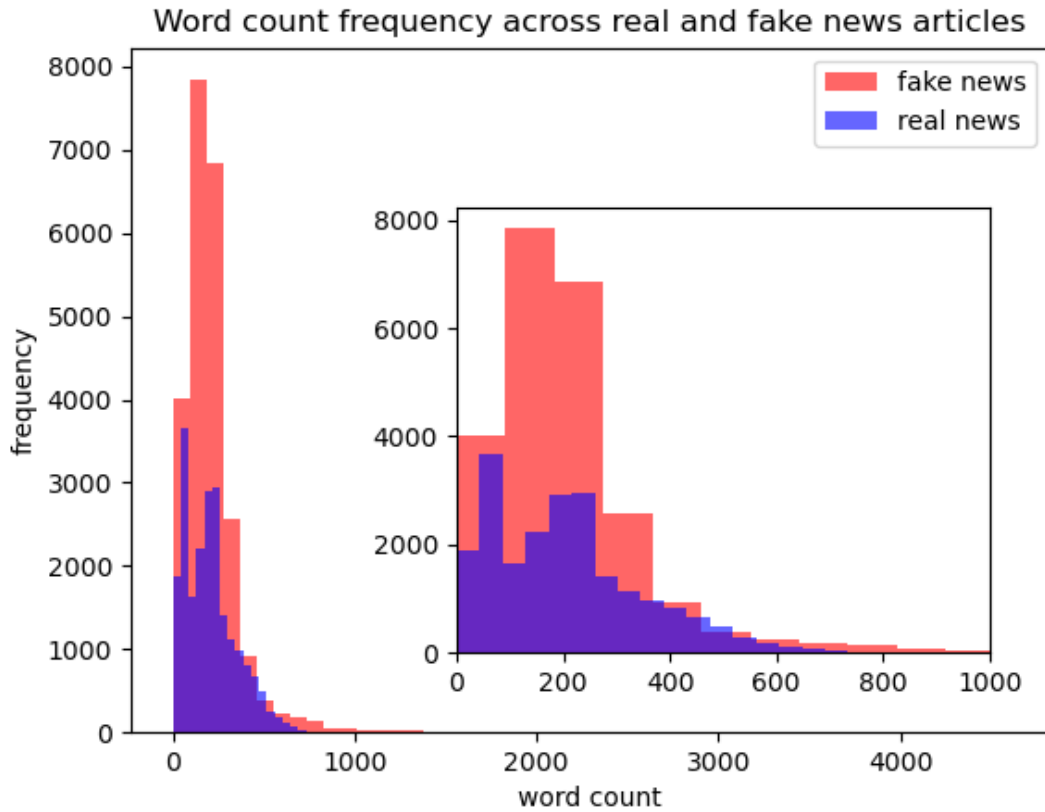


Figure 4: Part 1 LSTM accuracy scores for three models on each of three datasets. The training portion of one dataset was used to fit a tokenizer and model (x-axis), then the testing portions of each dataset, processed with the same tokenizer, were used to generate an accuracy score (y-axis).

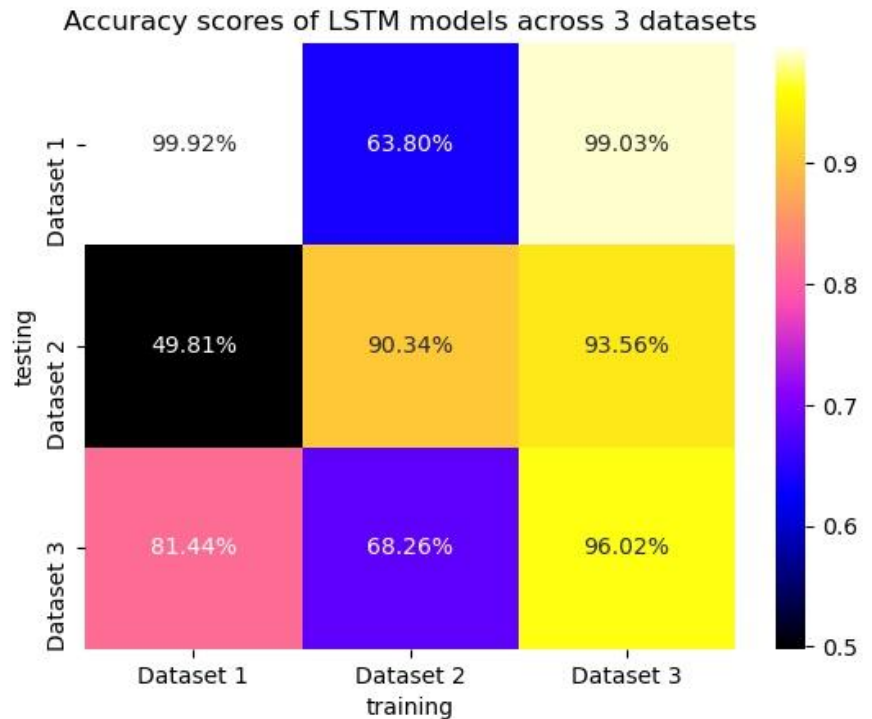


Figure 5: Part 2 accuracy scores. The y-axis labels represent (first digit) the training portion of which dataset was used to fit the vectorizer and model, (middle character) the vectorizer type: count (c) or TF-IDF (t), and (second digit) the testing portion of which dataset, processed with the same vectorizer, was used to generate the accuracy score. The model types (x-axis) are Random Forests (RF), Gaussian Naïve Bayes (GNB), Linear Support Vector Classifier (SVC), k-Nearest Neighbors (kNN), and Logistic Regression (LR).

Accuracy scores across classifiers, vectorizers, and datasets

