Fake News Detection

Team 17

Zeinab Sakran Mahmoud - 2021170219

Salma Mahmoud Nasr Aldin - 2021170233

Salma Morad Said Ahmed - 2021170234

Smaher Nabil Salem - 2021170237

Marwa Ehab Fathy - 2021170503

The Fake News Detection dataset initially has four columns. The Unnamed:0 column is automatically dropped since it is unique to each news record and provides no added value to the classification.

When we printed df.head(), we noticed that some of the FAKE labels had ‘strong’ or ‘charged’ words in the title. This prompted us to ask a few questions.

1. Is there a relation between the relative amount of ‘emotional’ words in the news and its label?
2. Do fake news articles exhibit distinct patterns in word choice or sentiment (positive, negative, or emotional) compared to real news articles?

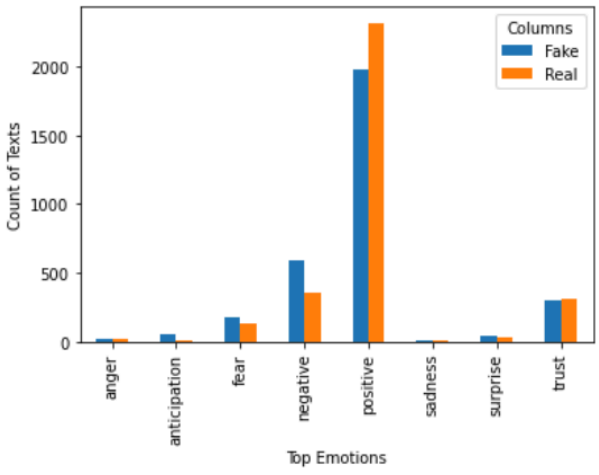
To answer these two questions, we searched and found a dataset called NRCLex that provides a list of words along with their sentiment labels: Fear, Anger, Trust, Surprise, Positive, Negative, Sadness, Disgust, Joy, Anticipation.

First we filter fake news entries, then iterate through their text content. For each article, we extract the most prominent emotion category, total word count, and the sum of emotional words. We add these metrics as new columns to the DataFrame and compute the average emotional intensity of fake news articles.

The same steps are repeated to get the average emotional intensity of the real news articles so that they can be compared.

The results of the averages were very similar, therefore we can conclude that the amount of ‘emotional’ words does not affect the label of the news.

For the second question, we group articles by their dominant emotion and count how many articles fall into each emotion category (e.g., anger, joy) for both fake and real news. By plotting the results, we can see that there is an overall trend where most of the news (fake and real) uses words in the positive category followed by the negative one, so no differences in trends were found between the fake and real news.



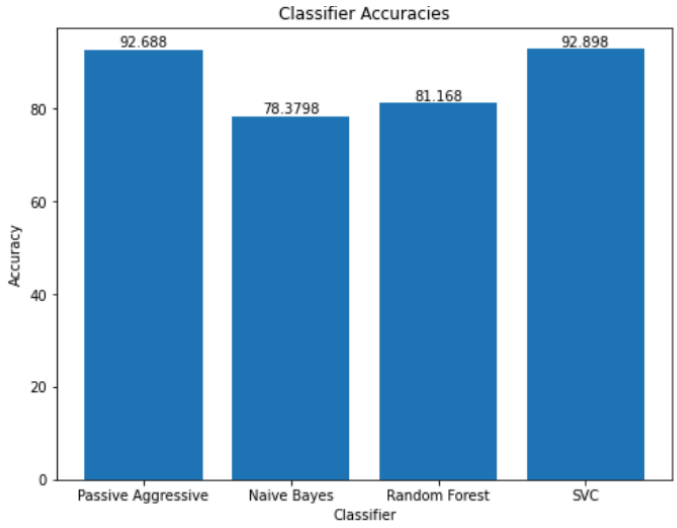
Had the answer to the previous questions been yes, we would've extracted a new feature with the frequency of 'emotional' words in each text and used it to train the models.

Next we started prepping the data to feed it to the classification models.
First, we removed all the stop words, then used an object of SnowBall stemmer to stem the news.

Next we split the data into train and test dataframes.

We then used a label encoder for the label feature and TF-IDF to 'encode' the news.

Now that the data is ready, we pass it to multiple classification models and plot their accuracies.



Finally, to predict whether unseen news is fake or real, we need to predict using a previously trained model, so we chose to dump the one with the highest accuracy: Passive Aggressive Classifier. We dump this model and the tfidf object so we can load them to be used on the new data. The rest of the preprocessing (joining and dropping columns, removing the stop words and stemming is done separately on the new dataset).