FENCE: Fake News Classifier

Vedant Khandelwal

I. Problem Statement

As of 2021 about five billion active internet users [1] out of which about four billion are active social media users [2], making social media becomes an easy platform to share and obtain information about anything. Hence, making social media a platform to share news with a large population. obtaining news from social media is a double edged sword, on one hand it is easy, low cost and faster, on the other hand it also becomes an easy target to spread fake news, i.e, low quality news with false information. This spread of fake news can also cause issues and can prove dangerous for family and society. Based on a survey [3], 62 percent of US adults in US were reported to obtain news information from social media, while only 38 percent obtain it from news media websites.

As people are increasingly getting indulged more and more time of their daily life in social media, it becomes an habit for them to consume information from social media. One of the main reasons are the features that provide to easy share of news, comment and discuss about the news with friends attracts more and more people. And fighting the issue of fake news during a pandemic can be very harmful for the society [4]. Hence, it is important to handle this issue for infodemic right from the start of it.

Most of the Deep Learning Algorithms have been performing very well in handling fake news with a large amount of data given about a specific topic. Because of which, those DL algorithms are unable to handle the issue for a different subject/topic and hence, unable to handle infodemic right from its start.

II. Intended Users

The intended users for the proposed system are all the social media platforms. Some of the recent statistics related to social media platforms [5] are given below:

- Worldwide trust on social media has dropped about eight percent from 2020 to 2021.
- In Quater three, there were 1.8 billion fake news engagements in facebook.
- 52 percentage of Americans encounter fake news on a daily basis

To tackle the problem above, while various social media platforms are spending a high amount of time and research to tackle the issue of fake news. The proposed algorithm can help them tackle the problem easily, as it generalises over different types of fake news (News articles, Tweets) across multiple subjects.

III. Contributions

- The proposed system is able to generalise without much data related to a specific subject.
- The proposed system can also be generalised over different types of data (News articles and Tweets).

IV. Dataset

The proposed approach uses dataset complied from various resources [6]–[12]. The dataset consists of about 63k tweets and news articles which are distributed over two classes fake and real. Example of Fake news and Real News from the dataset are given below:

- *A post claims compulsory vaccination violates the principles of bioethics, that coronavirus doesnâ€™t exist, that the PCR test returns many false positives, and that influenza vaccine is related to COVID-19*.
- *Hays County in Central Texas has seen an "incredible 845% increase in (COVID-19) cases since June 7*

More information about the dataset is given below:

- The dataset consists of four columns:
  - Text: This column contains the tweets/ news articles headlines.
  - Class: This column contains labels for the text. Real (1) or Fake (0).
  - Subject: This column contains details about the context/topic of the text. (COVID-19, General, Politics, SyriaWar).
  - Type: This column contains if the data is a tweet or a News article headline.

The next section discusses about the Exploratory data analysis.

V. Exploratory Data Analysis

The different type of exploratory data analysis performed are discussed below:

A. Class Distribution

This analysis discuss about the class distribution in the entire dataset and across different types. Table I shows the percentage of class distribution.

TABLE I: Class Distribution in the dataset

Dataset	% Fake Data	% Real Data
Entire	60	40
Tweets	55	45
News Articles	58	42

Based on the information in Table I, the dataset is slightly skewed towards class 0 (Fake News).

B. Readability Score

This analysis identifies the level of difficulty in understanding natural language sentence [13]. To measure the level of difficulty, we have employed Flesch Readability Score metric [14]. The score takes into consideration two important factor:

- AWL: Average world length in a sentence.
- ASL: Average word length in syllables (number of syllables divided by number of words. For example, One: Life, love; Two:About, Tuesday).

Further the two factor are combined using an equation as shown below:

$$Readability\!-\!Score = 206.835 - (1.015AWL) - (84.6ASW) \quad (1)$$

The equation given an score from 0 to 100, where 0 refers to most difficult and 100 refers to the easiest. The distribution of the level of difficulty among different class of data are shown in Table II

TABLE II: Readability Score Distribution over the class

Class	% Difficult	% Easy
Fake	40	60
Real	70	30

The information given in Table II shows that comparatively most of the fake news is easier to read where as Real news is much difficult to understand.

C. Click-bait

The fake news pieces are created intentionally to influence a large population for political or financial gain. Hence, they are crafted using inflammatory, opinionated language to report objective claims know as clickbait [15]. Clickbaits are used to encourage user to click on a specific link or follow and read a specific content. The distribution of clickbait percentage over the different class is given in Table III

TABLE III: Percentage of clickbait data over class

	% Real Data	% Fake Data
Clickbait	35	75

The information give in Table III shows that most of the fake data are clickbaits, whereas very less percentage of real data can be considered as clickbait.

D. Average number of words

This analysis tells us about the average number of words over different class, as shown in Table IV

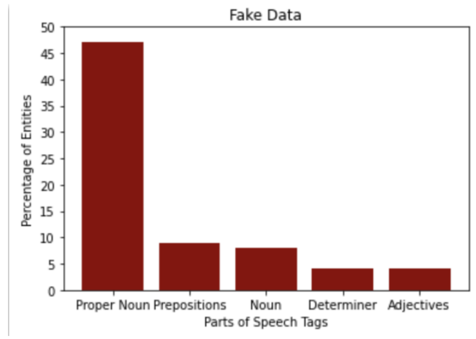
TABLE IV: Average number of words over class

	Real Data	Fake Data
Average Word Length	12	18

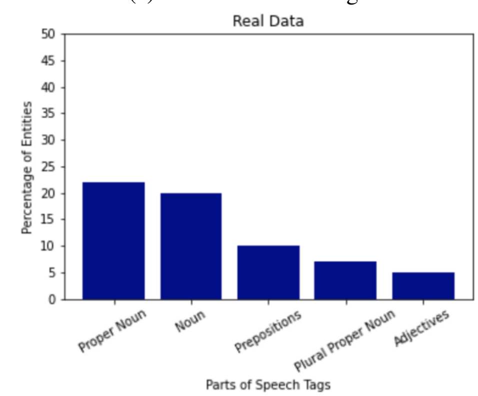
The information given in Table IV shows that fake news tends to have a longer sentence compared to that of real news

E. Parts-of-Speech Tags

The method of assigning a word in a corpus to a corresponding part of a speech tag depending on its meaning and description is known as POS tagging. This is a difficult task since a word may have a different POS depending on the context in which it is used. It can be used to understand the meaning of the sentence [16] and further to create knowledge graph using Natural language sentences. The distribution of five most frequent POS tags in both the classes are shown in the Figure 1.



(a) Fake Data POS Tags



(b) Real Data POS Tags

Fig. 1: POS Tags distribution over different class

The Figure 1 shows that, Fake news data majorly uses Proper Nouns, Prepositions, i.e using important entities entities to gain the trust of the user on the same time keeping it simple enough to be understand by most of the population. Whereas for real data we see Proper noun and noun, i.e, using all the important entities, also making data difficult for population to understand.

VI. Technical Approach

The proposed approach employs the usage of Siamese Neural Network [17] to solve the issue of handling data from an unseen subject.

A. Siamese Neural Network

To obtain a high performing data, a large number of data points for each class is required. After which, the model can perform well on those particular class of data. Whereas on the other hand, in a one shot learning setting can even train model for a class with one data point. Siamese neural network employs one shot learning [18]. In layman terms, Siamese Neural Network is used to classify data into positive and negative data points.

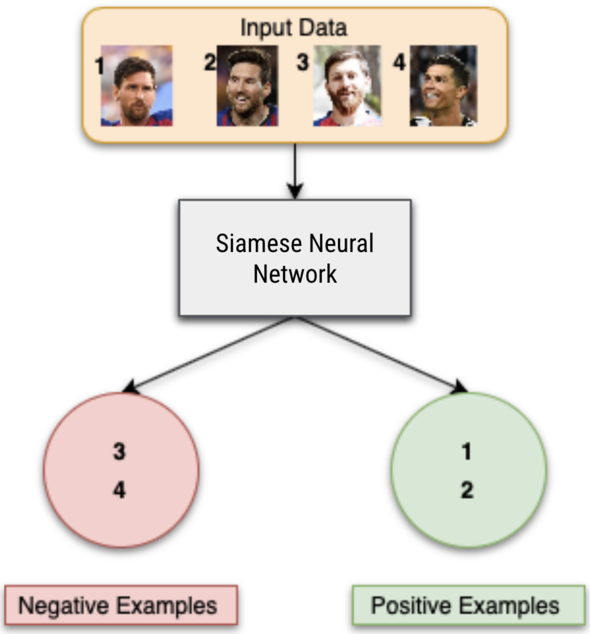


Fig. 2: Siamese Neural Network

For example, Figure 2 shows four data points given as input to the model where 1,2 are pictures of the same man and 3,4 are different pictures. The idea of using Siamese neural network is that, it will classify 1,2 as positive and 3,4 as negative. The Siamese Neural Network is trained with a loss function, known as Triplet Loss function. The triplet loss function, takes into account an anchor (A) data point, a positive (P) data point (one with a similar class as Anchor), and a negative (N) data point (one with a different class as anchor). Triplet loss function converges by decreasing the distance d(A,P) and increasing the distance d(A,N), with the help of a margin variable (alpha). Consider the two equations below,

$$d(A,P) - d(A,N) + (alpha) <= 0 \tag{2}$$

$$d(A,P) - d(A,N) <= 0 \tag{3}$$

The Equation (2,3) shown above are trying to decrease the distance d(A,P) and increase the distance d(A,N). But Equation (3) can also be satisfied by giving the values of A,P,N as NULL. In that case, the network won't learn anything, even after the loss function converges. Hence, a margin variable is introduced, which handles the issue of trivial solution satisfying the equation and ensures the learning of the network.

B. Approach

In this given approach (as shown in Figure 3), we use Siamese Neural Network, which takes input in batch of three (Anchor, Positive, Negative). It is used to learn the representations of the text [19]. Along with the representation obtained from the trained network, outcomes of EDA is also added to the feature set of the model. The features obtained from EDA are as follows:

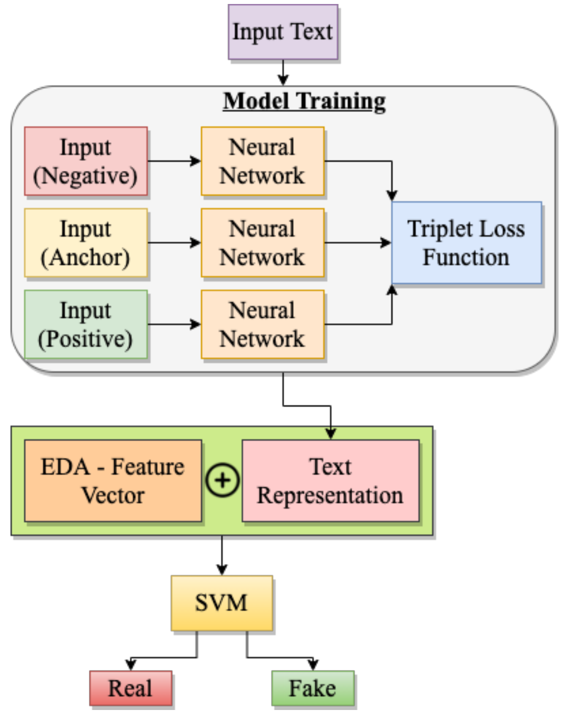


Fig. 3: Proposed Approach

- Readability Score: The score is normalized between 0 to 1, where 0 refers to most difficult and 1 refers to the easiest sentence.
- ClickBait: This contains the information if a sentence is clickbait or not.
- POS Tags: One-hot encoding of the presence of six most frequent tags from both the classes are obtained. (Proper Noun, Prepositions, Identifiers, Noun, Adjective, Plural Proper Noun).

Once the complete feature set is obtained, it is further used to train a Support Vector Machine Classifier. The next section discusses about the experimental setup and results.

VII. Experimental Setup and Results

There were two experimental setups, which are discussed in the next two sub sections.

A. Generalising over type

In this experiment, the train and test datasets were divided based on the type of data. Here, tweets was train dataset and news articles was Test dataset. The model tries to generalises based on the its learning over train dataset, i.e tweets. An ablation study is also performed by remove the EDA feature set from the input representation. The results of this experiment are shown in Table V

The information in the Table V shows that, using the EDA feature set aids in generalising over the type of data.

TABLE V: Generalising over type

Feature Set	Accuracy	Precision	Recall
Embedding	73.87	73.87	75.22
Embedding + EDA Features	77.32	77.54	79.33

B. Generalising over subjects

In this experiment, four-cross fold validation is performed with test dataset is taken as the entire single subject and the train is taken as the rest of the dataset. The model tries to generalise to the new subject (unseen) based on its learning over the train dataset. An ablation study is also performed by remove the EDA feature set from the input representation. The results of this experiment are shown in Table VI

TABLE VI: Generalising over subject

Feature Set	Accuracy	Precision	Recall
Embedding	78.38	78.38	79.12
Embedding + EDA Features	82.44	81.98	83.63

The information in the Table VI shows that, using the EDA feature set aids in generalising over the unseen subject.

VIII. Conclusion and Future Works

A generalised fake news classifier is implemented, which can be used to handle the situation even for a new subject over tweets as well as news articles. The classifier employs usage of linguistic characteristics of the sentence to further aid during the classification process.

The future scope of this work can employ usage of more linguistic features [20] which can be used to model the language styling of different classes. The approach can also be implemented using various state of the art text classification algorithms.

References

[1] 2021. [Online]. Available: https://www.statista.com/statistics/617136/digital-population-worldwide/

[2] 2021. [Online]. Available: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

[3] J. Gottfried and E. Shearer, "News use across social media platforms 2016," 2016.

[4] B. Lee, "Fake covid-19 vaccination record cards are a growing problem, says fbi," 2021. [Online]. Available: https://www.forbes.com/sites/brucelee/2021/05/01/fake-covid-19-vaccination-record-cards-are-a-growing-problem-says-fbi/?sh=74f4c4fc2d6d

[5] 2021. [Online]. Available: https://letter.ly/fake-news-statistics/

[6] F. K. A. Salem, R. Al Feel, S. Elbassuoni, M. Jaber, and M. Farah, "Fa-kes: A fake news dataset around the syrian war," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, 2019, pp. 573–582.

[7] P. Patwa, S. Sharma, S. PYKL, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, "Fighting an infodemic: Covid-19 fake news dataset," *arXiv preprint arXiv:2011.03327*, 2020.

[8] 2021. [Online]. Available: https://www.kaggle.com/mrisdal/fake-news

[9] E. Source, "Impact - isot fake news dataset," 2021. [Online]. Available: https://www.impactcybertrust.org/dataset_view?idDataset=952

[10] 2021. [Online]. Available: https://www.kaggle.com/mrisdal/fact-checking-facebook-politics-pages

[11] 2021. [Online]. Available: https://figshare.com/articles/dataset/Twitter_Death_Hoaxes_dataset/5688811

[12] 2021. [Online]. Available: https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset

[13] R. Santos, G. Pedro, S. Leal, O. Vale, T. Pardo, K. Bontcheva, and C. Scarton, "Measuring the impact of readability features in fake news detection," in *Proceedings of The 12th Language Resources and Evaluation Conference*, 2020, pp. 1404–1413.

[14] 2021. [Online]. Available: https://en.wikipedia.org/wiki/Flesch\%E2\%80\%93Kincaid_readability_tests

[15] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD explorations newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[16] O. Das and R. C. Balabantaray, "Sentiment analysis of movie reviews using pos tags and term frequencies," *International Journal of Computer Applications*, vol. 96, no. 25, 2014.

[17] M. Bryan and J. N. Philipp, "Unsupervised pretraining for text classification using siamese transfer learning." in *CLEF (Working Notes)*, 2019.

[18] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2. Lille, 2015.

[19] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 378–383.

[20] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Systems with Applications*, vol. 169, p. 114171, 2021.