Fake News Detection Using NLP

RESULTS

'Fig.2' shows thee number of fake news and real news in the dataset. We have used word clouds to check which are the words which appear frequently in the fake and real news. 'Fig.3' shows the Word Cloud for real news and Fig. 5.4 shows the Word Cloud for fake news.

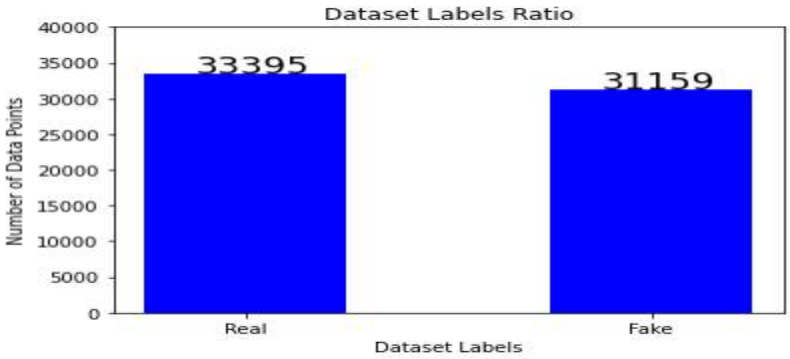

Fig.2. Dataset Labels Ratio.

Fig.3. Word Cloud for Real News.

Fig.4. Word Cloud for Fake News.

After the models were trained we calculated the performance metrics accuracy, precision and recall. 'Table. I' shows the performance metrics of the models.

TABLE I

Model	Accuracy	Precision	Recall
MODEL 1 [Fed with text vectors of 'Title' obtained by GloVe]	89.71%	91.84%	86.61%
MODEL 2 [Fed with text vectors of 'Text' obtained by GloVe]	92.8%	92.3%	92.9%
MODEL 3 [Fed with text vectors of 'Title' obtained by Word2Vec]	86.49%	87.14%	84.04%
MODEL 4 [Fed with text vectors of 'Text' obtained by Word2Vec]	96.26%	95.40%	96.86%
MODEL 5 [Fed with text vectors of 'Title' obtained by TF-IDF]	80%	70%	70%
MODEL 6 [Fed with text vectors of 'Text' obtained by TF-IDF]	81%	71%	76%

The performance is measured using the accuracy, precision and recall.

Accuracy: It shows the overall accuracy of the instances which are correctly classified to the total number of the instances. It is calculated by the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Where, TP = true positive, TN = true negative, FP = false positive, FN = false negative.

Precision: It represents the percentage of relevant sarcastic headlines. That is, it measures the amount of headlines categorized as sarcastic against the total number of headlines classified as sarcastic. It is calculated by the following formula:

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

Recall: It represents the percentage of relevant sarcastic headlines that have been searched. That is, against the total number of sarcastic headlines, measured the number of headlines that are normally classified as sarcastic. It is calculated by the following formula:

$$Recall = \frac{TP}{TP+FN} \qquad (3)$$

"Table 1" shows the results obtained by the models. From the results obtained we can observe that the model trained using the content of the news gives better output than the other models. Also, we can see that the models which have used Glove and WordVec method work better than the models using TF-IDF.

VI. CONCLUSION

Fake news have increased in recent years and it has caused a lot of harm to the society. This research project aimed to develop a model using the techniques of NLP and ML to detect if a news article/headline is fake or not and identify which methods give better output. In this paper, we have presented six LSTM models and three different methods were used for feature extraction. We have used different attributes like the title and text of the news to perform fake news detection. For future work we can work on larger dataset and also future research can be done on images, videos which can help in improving the models.