Detecting Fake News Using Python

Anand Gupta, Nisarg Savalia, and M. Manickam, *SRM Institute of Science and Technology, Kattankulathuran9079@srmist.edu.in,nj6019@srmist.edu.in,Manickam@srmist.edu.in*

***Abstract*—The consumption of news from online sources especially, the social media is on fire in today's world. Users are benefitted because of ease of access coupled with its cheap cost. Contamination of the quality of information comes into play, people tend to get infected with fake news pretty fast, with misleading data being fed and reposted without necessary fact-finding and validation. In such a scenario where fake news is producing adverse effects on its consumers, comes the call for an authentic system to check the reliability of the information. Machine learning algorithms play a great role when devising a system. Different classifiers are brought into use for this purpose. In this paper, we have implemented the Passive-Aggressive Classifier, which fulfills the task with an accuracy of over 90%. The best performance is achieved when using Passive-Aggressive Classifier in comparison to others.**

I. INTRODUCTION

With the rise in popularity of the social media, more people turn towards consuming news from social media rather than the traditional news media. Fake news is a matter of serious concern as it tends to become a great threat to democracy and the freedom of expression. Public trust in the government is distorted. Fake news spreads like wildfire, its extensive spread poses a great threat of extreme negative effects on its consumers as well as on the society as a whole. A kind of yellow journalism, in which pieces of news, probably hoaxes, is being spread through the online media, especially the social media. This derives its purpose as to impose certain ideas and often times has political agendas. These type of news items have a great chance of containing false and/or exaggeration filled claims, which in turn misleads the consumer creating a false imagery.

II. LITERATURE SURVEY

A. *Opinion spam detection*

The reviews are broadly categorized into 3 types: fake reviews, reviews defaming an individual brand, and reviews promoting a product. Standard supervised machine learning classifiers such as logistic regression (LR) and Naïve Bayes (NB) are used to identify the last two types from the three categories mentioned. The accuracy for type 1 was 78% using all features and around 60% with the text features only. For type 2 and 3, it could identify with an accuracy of over 90%.

B. *Content based detection*

The n-gram term frequency is used to detect fake opinions. Using SVM classifier, 84% accuracy is achieved, separating the positive and negative into groups helped increase the accuracy to 86%. When humans were employed to judge the reviews, the highest score was 65%. Other models make use of semantic similarity metrics.

C. *Reviewer behavior-based detection models*

Assuming that a spammer may target certain brand or a product for its better growth, hence rating given would be higher than the average rating. Fake reviews will be written in a short period. The evaluation results proved the model to be effective and was seen to outperform other baseline methods.

III. PROPOSED WORK

To be able to easily differentiate between fake and real news and help people through this fake news detection (including fake posts and tweets). The kind of information source becomes a determining factor be it textual content/profile/descriptions, identifying fake news from online sources, especially the social media is our target. We achieve it by developing a structured framework for the thorough study of fake news.

A. *Abbreviations and Acronyms*

TF-IDF stands for Term Frequency- Inverse Document Frequency. It refers to the weightage of words in the document, it takes in account the number of times a word appear in the document.

B. *Systematic Approach*

- *Data Preprocessing*

  Input, in the form of a dataset, is fed to the system, which is handled by the model build, making use of the tfidf vectorizer.

- *Stop Word Removal*

  after removing the stop words, common words, like “the”, ”when” and “there”, taking in account only those ones that appear at least a specific number of times.

- *Feature Extraction*

  Extracting the most optimal features for the tfidf vectorizer is done by using a n-number of the words which appear most of the times,

- *Term Frequency*

  Utilizing the counts or words which appear in the data to find out the similarity between different documents.

- *Classification*

  Text classification is the basis of this problem, a combination of Naïve Bayes classifier along with the passive aggressive algorithm’s implementation shall prove to be the best and optimal solution as this acts as a standard for text-based processing.

- *Confusion matrix*

  In the end, we calculate the accuracy and print the true and false negatives and positives based on the confusion matrix, a table used to determine the performance of classification model on the data-set.

- *Accuracy*

  Accuracy is determined by how well does the classifier work on the given dataset. It is calculated by taking in account the false and true positives and negatives.

C. *Equations*

Accuracy is calculated using:-

Accuracy = (TP+TN)/(TP+FP+TN+FN)

IV. IMPLEMENTATION

A. *COUNT VECTORIZER*

A powerful method of tokenizing a collection of text data to build a set of known words, which is later used to encode new data.

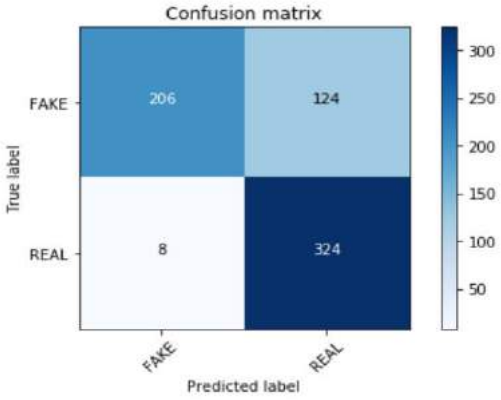
B. *TFIDF VECTORIZER*

Converts the text to feature vectors which are used as input to the vocabulary, it is a dictionary which changes every token (word) to feature index in the form of a matrix, and every unique token is assigned a feature index.

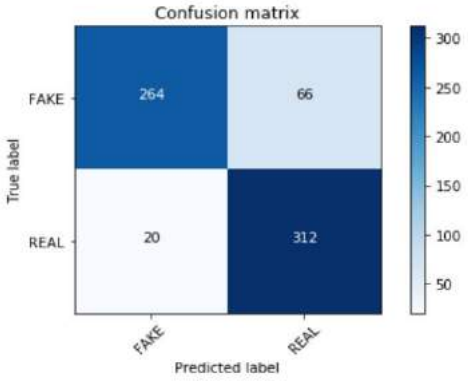
C. *NAÏVE BAYES CLASSIFIER*

A machine learning model generally used for classification purposes. It works on the principle of the Bayes Theorem.

- *USING TFIDF VECTORIZER*



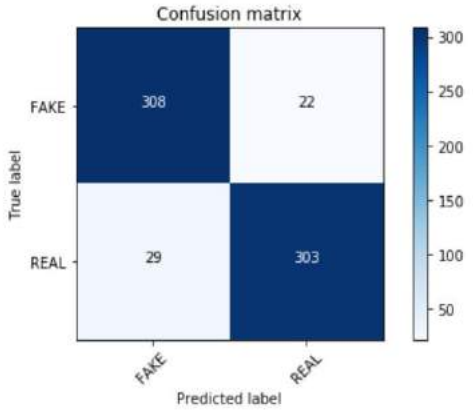
- *USING COUNT VECTORIZER*



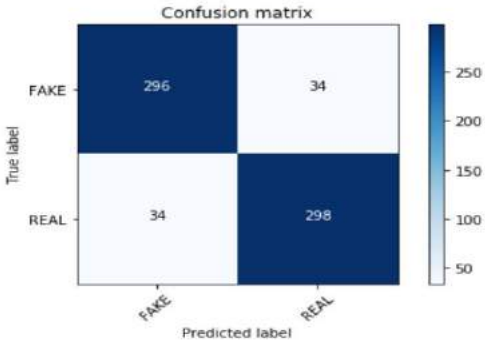
D. *PASSIVE AGGRESSIVE CLASSIFIER*

It is a member of the large-scale learning algorithms’ family. Somewhat identical to the Perceptron but do not need rate of learning, instead a regularization parameter is implemented.

- *USING TFIDF VECTORIZER*



• *USING COUNT VECTORIZER*



V. RESULTS DISCUSSION

The implementation of a machine learning model for detection of the authenticity of an article and its classification as fake or fact has been successfully done. 298 articles were correctly identified as real and 296 articles were correctly identified as fake. In such a classification, it is important to limit the number of false positives as facts may be classified as fake.

VI. CONCLUSION

Online media, specifically the social media has been drastically used to manipulate information and mislead people by sharing fake news. It has impacted the individuals and the society in a strong negative manner. This project focuses on structured review, summarization as well as evaluation of the existing research on fake news. Through this paper, it concludes that by means of passive aggressive classifier and TF-IDF vectorizer, the accuracy achieved is over 90%.

REFERENCES

[1] Kushal Agargwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema –"Fake News Detection on Machine Learning and Natural Language Process".

[2] Arulprakash, M., Kamal, A., Manisha, A., "QR-Code scanner based vehicle sharing "ARPN Journal of Engineering and Applied Sciences, 2018, 13(10), pp. 3441-3448.

[3] Arulprakash, M. "Dynamic evolutionary information diffusion over mobile social networks " ARPN Journal of Engineering and Applied Sciences, 2016, 11(19), pp. 11457-11464R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] Kai Shu , Amy Sliva , Suhang Wang , Jiliang Tang and Huan Liu- "Fake News Detection using Data Mining Perspective".

[5] Bhowmik D, Zargari S, Ajao O (2018) Fake news identification on twitter with hybrid CNN and RNN models

[6] Wang Y, Ma F, Jin Z, Yuan Y, Xun G, Jha K, Su L, Gao J (2019) EANN: event adversarial neural networks for multi-modal fake news detection. In: 24th ACM SIGKDD international conference on knowledge discovery & data mining, London.

[7] Markines B, Cattuto C, Menczer F (2009) Social spam detection. In: 5th international workshop on adversarial information retrieval on the web, 2009.

[8] Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y (2006) Online passive-aggressive algorithms. J Mach Learn Res 7:551–585.

[9] Study-on-machine-learning-methods - Junaed Younus Khan , Md. Tawkat Islam Khondaker , Anindya Iqbal and Sadia Afroz – "A Benchmark