IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022

ISSN: 2320-1363

Detecting Fake News in forms of Text and Image by using the Logistic Regression Model

[1]M. Yamini Lahari, [2]Ch. Ramadevi

[1]M.Tech Student, Dept. of CSE, Sir C R Reddy College of Engineering College, Eluru.

[2]Assistant Professor, Dept. of CSE, Sir C R Reddy College of Engineering College, Eluru.

Abstract: The cyberspace news consumption is increasing day by day all over the world. The main reason for cyber space news consumption is due to its rapid spread of information and its easy access which lead people to consume news rapidly without the knowledge of whether the news is false or true. Thus, it leads to the wide spread of false news which leads to the negative impacts on society. Therefore, false news prediction on cyberspace is attracting a tremendous attention. The issue of fake-news prediction on cyberspace is both challenging and relevant as spreading of fake news occurs in various streams like text, audio, video, images etc. This model works on processing the text and images together by providing an interactive Application Interface (API), i.e. text by applying the model Logistic regression classifier and image by applying self-consistency algorithm. The natural language tool kit (NLTK) model is used for these implementation through python. Once the news is predicted fake, a report is redirected to the authorized website (cybercrime department) to take the immediate necessary actions required to stop this news from spreading.

Key Words: Logistic regression classifier, self-consistency algorithm, Cyberspace, fake-news, text and image, report, redirect.

I. Introduction: Present days' people spend a lot of time in Internet and consume news. The main reason for rapid spread of news in cyberspace is due to its low cost, easy access and easy sharing facility. This made people to consume news from cyberspace rather than fetching it from television or newspaper. The widespread of fake-news

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022

ISSN: 2320-1363

will have a serious negative impact on society and individuals. Fake-news detection on cyberspace has led to tremendous research all over the world to predict with the exact accuracy as the content of false-news is diverse in topics. People consuming news from cyberspace produce data which is diverse and difficult to predict This model is a solution to all these problems of fake news in cyberspaces that is fast growing. In particular, the datasets which are trained by various machine learning techniques like data pre-processing, feature selection, self-consistency etc. and all these are implemented by natural language processing in python. Here we detect both forms of fake news, i.e., both text and image streams. Once the prediction is false the report is generated and it is immediately redirected to the authorized page (cybercrime department) insisting the seriousness of the news for which the actions will be taken accordingly. Through this we try to bring a safe and trustable cyberspace experience to people who rely on this. They can now verify news before they are believing or forwarding them to others.

II. Comparative Study: Our Study on both text and image classification and prediction of news, we present a detailed view of work on both text and image streams.

A. Text Features of News Verification: The research paper nowadays presents the prediction of fake news using URL [1] and tweet based text features [2]. These text content semantic features are extracted from tweet to find out the sentiment scores and opinion words [3] rather than statistical features. The above researches of prediction of text are not effective in today's increasing traffic of cyberspace. The extraction of semantic features from text is not that easy since it is dependent on text mining [1]. Whereas, in this [4] research paper using N-Gram Analysis and Machine Learning Techniques, the major disadvantage of this research is that it completely lacks to run publicly available datasets. It experiments only on one technique and predict the output analysis.

B. Image Features of News Verification: Nowadays images in cyberspace are popular in forms of posts, memes etc. along with the description of text. Image features of news prediction still exists in basic level of

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022 ISSN: 2320-1363

research. It is important to find out the multimedia attachment [5] in tweet if it is either text or image. The fake-news prediction of image mostly consists of outdated images which is published earlier [6]. Therefore, in order to automatically predict the multimedia content in tweet, this research [7] has not met the complete accuracy of image prediction. Image features of news prediction basically involves two features i.e. visual content and statistics [1]. These features are proposed to characterize the distinctiveness in images [1]. It is clear that the above work of both text and image does not meet the complete accuracy of prediction of the news. Therefore, to visually describe both the text and image, we use the implementation of the most accurate classifier of text prediction and self-consistent method of image prediction which meets clarity score, coherence score and diversity score.

III. Existing System: The research project nowadays presents the prediction of fake news using URL and tweet based text features. These text content semantic features are extracted from tweet to find out the sentiment scores and opinion words rather than statistical features. The above researches of prediction of text are not effective in today's increasing traffic of cyberspace. The extraction of semantic features from text is not that easy since it is dependent on text mining. Nowadays images in cyberspace are popular in forms of posts, memes etc. along with the description of text. Image features of news prediction still exist in basic level of research.

Disadvantages

- There is not much research done of this topic.
- Manual detection is done, which is very time consuming.

IV. Proposed system: Therefore, false news prediction on cyberspace is attracting a tremendous attention. The issue of fake-news prediction on cyberspace is both challenging and relevant as spreading of fake news occurs in various streams like text, audio, video, images etc. This model works on processing the text and images together by providing an interactive Application Interface (API), i.e. text by applying the model Logistic regression classifier and image by applying self-consistency algorithm. The natural language

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022 ISSN: 2320-1363

tool kit (NLTK) model is used for these implementation through python. Once the news is predicted fake, a report is redirected to the authorized website to take the immediate necessary actions required to stop this news from spreading.

Advantages:

- It will process both text and images for detecting fake news.
- Report will generate once as news is predicted fake.

V. Modules:

Pandas: pandas are an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Numpy: NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

MatPlotLib: matplotlib.pyplot is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

Tensor flow: Tensor Flow is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. Tensor flow is a symbolic math library based on dataflow and differentiable programming.

Keras : Keras is a neural networks library written in Python that is high-level in nature – which makes it extremely simple and intuitive to use. It works as a wrapper to low-level libraries like Tensor Flow or Theano high-level neural networks library, written in Python that works as a wrapper to Tensor Flow or Theano.

VI. Algorithms:

Classification Process: the accurate model, we carry out all the five classification process namely Naïve-Bayes classifier, Logistic regression classifier, Linear Support Vector Machines (LSVM), Stochastic Gradient Descent (SGD), Random Forest classifier. The implementation of these classifiers are done using python Natural Language Toolkit (NLTK). These five classifiers are validated using K-fold cross validation [8] and the

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022

ISSN: 2320-1363

accuracy of the output are found for all these models. Out of all these the best two performing model is taken which is known as candidate models. The candidate models chosen are Logistic regression and random forest classifiers in terms of precision and recall. Further, logistic regression and random forest classifier are performed to find out the best parameter with Grid Search method. By knowing the accuracy of these models, the best model is saved and it is used further for our classification.

The logistic classifier is also known as the sigmoid function in which it takes the real number and classify it between 0 and 1.

It can be defined as:

1 / (1 + e^-value)

Where e is the base of the natural logarithm and value is the numerical value in the dataset. In below graph of logistic regression model is performed with training examples and their equivalent score, by doing so the corresponding score is predicted for the training examples, in which the orange and blue line implies the false and the true news respectively.

Random forest classifier is another method that is classified by using decision trees. The final output is based on the accuracy of the trees and is classified by the random forest by reducing the training time. The most important advantage of random forest is that it is efficient for large amount of database. The graph of random forest is performed in a similar way to the logistic regression classifier.

VII. System Architecture: There are various models involved in implementation of each module starting from data pre-processing to classification of news. The system flow is shown in Fig.

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022

ISSN: 2320-1363

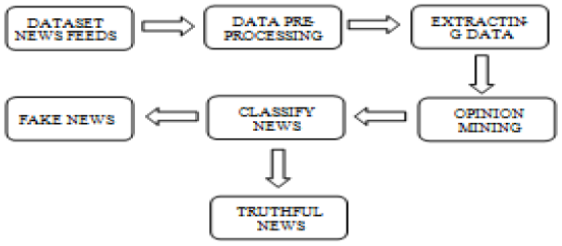

Fig. System Architecture

VIII. Conclusion: The consumption of news is increasing day by day in cyberspace than the traditional media. Due to its increasing popularity and user friendly access it leaves a huge impact on individuals and society. Therefore, in this model we have found a way to detect such fake news in both the forms of text and image by using the Logistic regression model. By redirecting the fake news to the authorized website (cybercrime department), we hereby frame a high social impact and thus it reduces the spreading of false news distinctly.

IX. Future Enhancement: This model can be further discussed for the future improvement in fake news detection which can be in audio, video streams and commercialize the field to other applications.

References:

[1] K.Sakthidasan, G.Srinithya, V.Nagarajan (FEB 2014), "Enhanced Edge Preserving Restoration for 3D Images Using Histogram Equalization Technique", International Journal of Electronic Communications Engineering Advanced Research, Vol.2, SP-1, Feb.2014, pp. 40-44

[2] S. Kwon, M. Cha, K. Jung, W. Chen and Y. Wang, "Prominent features of rumor propagation in online social media," IEEE Int. Conf. Data Mining, pp. 1103–1108, 2013.

[3] Hadeer Ahmed, Issa Traore and Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," Springer, pp. 127–138, 2017.

[4] K. Wu, S. Yang, and K. Q, "False rumors detection on sina weibo by propagation structures," IEEE Int. Conf. Data Engineering, 2015.

[5] S. Sun, H. Liu, J. He, and X. Du, "Detecting event rumors on sina weibo automatically," Web Technologies and Applications, Springer, pp. 120–131, 2013.

[6] Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian Fellow, "Novel Visual and Statistical Image Features for

Microblogs News Verification," IEEE Trans. Inf. Multimedia, pp. 1520-9210, 2016.

[7] Sanjay Yadav and Sanyam Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," IEEE Int. Conf. Advanced Computing, 2016.

[8] Yuanfang Guo, Xiaochun Cao, Wei Zhang and Rui Wang, "Fake Colorized Image Detection," IEEE Trans. Inf. Information forensics and security, pp. 1556-6013, 2018.

[9] Himank Gupta, Mohd. Saalim Jamal, Sreekanth Madisetty and Maunendra Sankar Desarkar, "A framework for real time spam detection in Twitter," IEEE Int. Conf. Communication Systems and networks, pp.2155-2509, 2018.

About Authors:

M Yamini Lahari is currently pursuing her M.Tech (CST) in CSE, Sir C R Reddy College of Engineering, West Godavari, A.P.

CH. Ramadevi is currently working as an Assistant Professor in Dept. of CSE Sir C R Reddy College of Engineering, West Godavari, A.P.