iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34599

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

Prototype of an Optimistic Fake News Detection Method

Swethasubrahmanian P[1], Sruthy MR[2], Biji KP[3]
[1]*Department of Computer Science and Engineering, APJ Abdul Kalam Technological University*
[2, 3]*Assistant Professor Department of Computer Science and Engineering, Nehru collage of Engineering and Research Centre Pampady*

Abstract: ***The influence of social media in the present word is more than anything, at the same time there is a lot of real news and fake news are spreading through it. But we cannot ignore online as well as offline social media in the present world. So that we need to analyse and extract the real side of the news that spread through media. The prototype of an optimistic fake news detection is try to convey a simple and fast processing method with best accuracy. I focus on an automatic identification of fake contents in the social media. First analyse dataset for the processing of the fake news detection and implemented data mining approaches in data set. The classification method used is the logistic regression along with data mining and machine learning.***
Keywords: Machine learning, Fake news, Logistic regression, classification algorithms.

I. INTRODUCTION

The term fake news is a common word nowadays. It is the process of false content that spread through various media and lead to misunderstanding the information. It may purpose fully or not. The main cause of fake news spreading the rapid growth of online media, they update news very instantly. Due to the fast growing network people can easily and also fast access of news are possible. But the thing is that news that shared by those media are real? We need to analyse it. “Basic introduction to fake news detection” [6]explained about what is fake news, type of fake news, a small introduction to fake news method, and finally about the algorithms used for fake news detection and classification. The aim of this project is to build prototype of an optimistic method fake news detection. “Comparison of machine learning algorithm in fake news detection perspective” compare various classification algorithm in accordance with fake news detection. This paper analyses the topic in four sections. The first section explain about the survey papers. Almost five papers are covered and each one focus on different methods and theories. The next section explain about the proposed system and finally about conclusion.

II. LITERATURE REVIEW

In [1] the author has analysed the fake news detection in machine learning and the deep learning perspective. They are used twitter data for their research. They are used twitter data for their research. This paper has put forward a model for finding forged news from the twitter. Establish a flow for how the fake news detection can be possible in machine learning and deep learning algorithms. They provide an overview of fake news recognition and it shows the NLP study on the topic. They use the countvectorizer and word embedding method. Likewise, They put forward more advanced and somewhat complex method . The work mainly focuses on the twitter strings and data. In[2]the author focus on machine learning algorithms for the prediction. They use probabilistic context free grammar for extraction of feature. In this paper they put forward a pipeline representation of their project . The best feature extraction method can be extracted through this paper. In[3] is an introduction to natural language processing and it explore NLP, NLP methods and their scope as well. They try to easily covey how python language and NLP are connected and how to use open source library natural language tool kit work.
In[4] they explore data mining and its method and also their applications.first analysed the knowledge recovery process and goes through data ming techniques. In[5] the research on information retrieval model was implemented, and they are analysed an information retrieval model with natural language processing.

III.PROPOSED SYSTEM AND METHODOLOGY

This project aims to use the natural language processing method along with machine learning approaches and data mining in fake news detection. The proposed system contains mainly four modules. First module contains data analysis and data pre processing, second module is machine model training third section prediction section, and finally model deployment.

A. Data Analysis

This section discusses data set and data processing method. The data set used for this project are from public domain. The data set contain column and rows and it stores information like author, title, text.

1) *Data pre-processing:* There are lot of data preprocessing methods available. For data preprocessing here we use regex, tokenization, stop word removal, lemmatization and finally stemming are applied. The tokenization means breaking the senescence in to individual tokens for the fast processing. Regex is nothing but regular expression here used to remove punctuation based on cortex free grammar. Stop words are the most commonly occurring words in any language. After undergoing data preprocessing the data s free from unwanted columns, removed all the missing record values, it breaks all the additional information and numeric values.

B. The Feature Extraction Method.

In the real time scenario we are dealing with a huge amount of data. In order to analyse and understands this data we need some process. It is not possible to handle in manually. The feature extraction help to avoid these kinds of problems. The feature extraction is a kind of dimensionality reduction method. Here bag of words /count vectorizer and TFIDF is compared.

C. The Model Building and Prediction.

Almost five machine learning algorithms[6] are analyzed for building the best classifiers for predicting the news state. Each of the filtered feature applied in these classifiers. After analyzed the accuracy, the best one is opted. Compared to remaining classifies logistic regression can use for prediction.

D. Model Deployment.

For model accessing an interface with web and implemented through python flask.

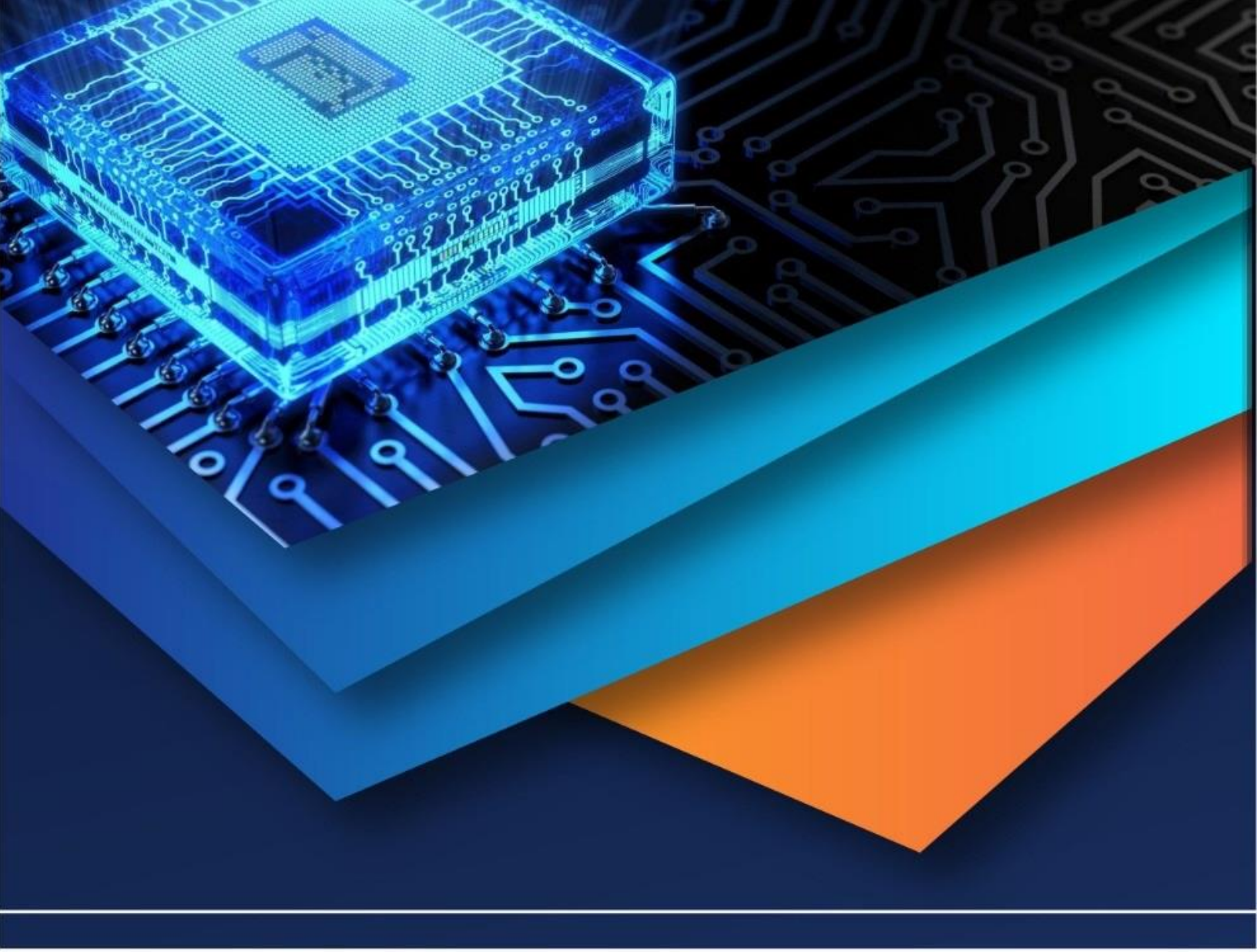
This paper is an outcome of the analysis of various classification algorithms. In the final round of comparison we choose logistic regression, naive bias, and stochastic gradient. We required dataset (both test and train) Text preprocessing methods, NLP methods and for deployment use python flask method.

IV. CONCLUSIONS

The prototype of an optimistic method fake news detection has analyzed various machine learning algorithms in the light of machine learning and NLP method. The effectiveness and efficiency are compared, from the analysis logistic regression better for fake news detection method with the best accuracy.

REFERENCES

[1] “Detecting Fake News using Machine Learning and Deep Learning Algorithms “Abdullah-All-Tanvir Ehesas Mia Mahir, Saima Akhter ,Mohammad Rezwanul Huq Dept. of Computer Science and Engineering
[2] “Evaluating Machine Learning Algorithms for Fake News Detection” Shlok Gilda Department of Computer Engineering
[3] “Natural Language Processing with Python “Steven Bird, Ewan Klein, and Edward Loper
[4] “Data mining and its applications” Bharati M. Ramageri
[5] “Research on Information Retrieval Algorithm Based on TextRank” Chenchen Xu School of computer science, Wuhan University of Technology, Wuhan, China
[6] “Basic Introduction to Fake News Detection.” Swethasubrahmanian P, Sruthy MR , Biji KP.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY