iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.34939

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 VI Jun 2021- Available at www.ijraset.com

Location Based Fake News Detection using Machine Learning

Prof. Rohit Nikam[1], Ms. Renuka Bhokare[2], Mr. Santosh Chavan[3], Mr. Roshan Sonawane[4], Ms. Dipali Adhav[5]
[1, 2, 3, 4, 5]*Dept. of Information Technology Sanjivani collage of Engineering,* Kopargaon, India

Abstract: ***Now a days lots of crime news incremented. In this system we can easy find which type of crime happened in particular city by using pin code. The easy access and exponential growth of the information available on social media network has made it detect to news or information fake or not. The easy dissemination of shared information has added to exponential growth of its falsification. On social media spreading lots of fake news. Thus it has become research challenge to automatically check detected news or information fake or real. Machine learning plays important role to classify the information in different categories. This paper reviews finding different types of crime news in particular city and detected news fake or real. One more feature is predict the future of crimes.***

I. INTRODUCTION

Luxurious life has become quite suitable and the people of the world have to thank the lots of contribution of the internet technology for information sharing. There is no doubt that internet has made our lives easier and access to surplus information viable. When people go unknown place they don't know which type crimes happen in particular city or area. This web app helps them to showing all crime news. This is an evolution in human history, but at the same time it unfocused the line between true media. Badly fake news collection a great deal of attention over the on social media. This kind of news fade but not without doing the misuse it intended to cause. The given social networking sites that play a major role in supplying various news include Facebook, Twitter, Instagram, Whatsapp etc. Many scientists believe that fake news issue may be point out by means of machine learning and artificial intelligence. This is because recently artificial intelligence algorithms have begun to improve work on lots of classification problems i.e. image recognition, voice detection. because hardware is cheaper and largest datasets are available. Various techniques are used to provide an accuracy range of 60-75 percent Which comprises of Native Bayes classifier Linguistic features based, SVM etc. Crimes are a social cost our society dearly in various ways. The Indian Government has taken steps to develop applications for the use of Cities and Central Police in relation with the National Crime Records Bureau(NCRB). About 10 percent of the criminals execute about 50 percent of the crimes. People who study criminology will be able to find the criminals based on the traces, characteristics and methods of different crime which can be collected from the crime sense. The large amount of crime datasets and the relationships between these type of information have made criminology an appropriate field for applying data mining, machine learning techniques. Identifying crime characteristics is the first step for proceeding with any further analysis and future crime rate prediction.

II. LITERATURE SURVEY

This literature review, shows a study of the factors that involved in the outspread of fake news. Fake news that is intentionally set up to delude and to cause harm to the publicis referred to as digital misrepresentation[2].Misrepresentation has the unrealized to cause obstacle, within minutes, for millions of people[3].Misrepresentation has been known to obstruct election processes, create harassed , argue and op- position among the public[2].In this review, the main reason of the spreading of fake news are located to reduce the inspirit of such false information. To clash the spreading of fake news on social media, the reasons behind the spreading of fake news must first be recognize. Thus, this literature review trying to recognized the number of feasible causes beyond the propagate of fake news. The purpose of this literature review is to identify why discrete incline to share false information and to possibly help in detecting fake news before it circulate[1].In [5] the authors actually develop two systems for fraud detection en- couraged support vector machines and Naive Bayes classifier respectively. They gather the info by means of asking people to directly provide true or false information on various topics like abortion, execution and fellowship. The truthfulness of the detection accomplish by the system is around 70 This text describes an easy fake news detection method supported one among the synthetic intelligence algorithms like native Bayes classifier, Random Forest and Logistic Regression. The aim of the research is to look at how these particular methods work for this specific problem given a manually labelled news dataset and to support (or not) the thought of using AI for fake news encounter.

©IJRASET: All Rights are Reserved

Also, the developed system was proved on a analogously new data set, which presented achance to estimate its performance on a current data. MykhailoGranik et. al. in their paper [4]expose naive Bayes classifier proceed towards as a software system and check against adata set news or posts. They were collected data from datasets according to different location to find out the particular post is real or fake. They accomplish classification perfection for fake new. Himank Gupta et. al. [6] present a framework based on different machine learning approach that distribute with various problems including accuracy shortage, time lag and high processing time to control thousands of news in 1 sec.Marco L. Della Vedova et. al. [7] first proposed a innovative ML fake news detection method which, by merge news content and social context features, exceed existing methodsin the literature, increasing its accuracy up to 78.8 Second, they implemented their method within a Facebook Messenger Chabot and validate it with a real-world application, procure afake news detection accuracy of 81.7Their goal was to allocatea news item as reliable or fake; they first portray the datasets they used for their test, then presented the content-based approach they implemented and the method they proposed to merge it with a social-based approach present in the literature.

III. FAKE NEWS TYPES

The various types of fake news by Authors of paper [8] , in their recent paper is summarize below.

1) *Visual-based:* These fake news posts use graphics a lot more in as content, which may include morphed images,doctored video, or combination of both.
2) *User-based:* This type of fabricated news is generatedby fake accounts and is targeted specific audience whichmay represent certain age groups, gender, culture, political affiliations .
3) *Knowledge-based:* these types posts give scientific (so called) explanation to the some unresolved issues and make users to believe it is authentic. For example naturalremedies of increased sugar level in human body .
4) *Style-based:* Posts are written by pseudo journal lists whopretend and copy style of some accredited journalists
5) *Stance-based:* It actually is representation of truthfulstatements in such a way which changes its meaningand purpose.

IV. METHODOLOGY

A. Datasets

The datasets we used in this study are open source and freely available online. The data includes both fake and real news articles from multiple platforms. The True/Real news articles published contain true description of real world events, while the fake news websites contain claims that are not aligned with facts. The conformity of claims from the politics domain for many of those articles can be manually checked with fact checking websites such as indianexpress.com and indiapress.com.

B. Performance Metrics

Confusion matrix is used to described the performance of classification model. Most of them are based on the confusion matrix. Confusion matrix is a tabular representation of a classification model performance on the test set, which consistsof four parameters: true positive, false positive, true negative, and false negative (see Table 1).

	Predicted true	Predicted false
Actual true	True Positive (TP)	False Negative (FN)
Actual false	False Positive (FP)	True Negative (TN)

TABLE I
DATASET

Fig. 1. Confusion Matrix

1) *Accuracy:* Accuracy is frequently used metric rep- resenting the percentage of correctly predicted observations, either true or false. To calculate the accuracy of a modelperformance, the following equation can be used:

$$Accuracy = \frac{TP + TN}{TP + TN + PF + FN}$$

In most cases, high accuracy value represents a good model,but considering the fact that we are training a classification model in our case, an article that was predicted as true whileit was actually false (false positive) can have negative effect; similarly, if an article was predicted as false while it containedwrong data, this can create trust issues. Therefore, we have used three other metrics that take into account the incorrectly classified observation, i.e., precision, recall, and F1-score.

©IJRASET: All Rights are Reserved

2) *Recall:* Recall is the ratio of True positive to Predicted Result i.e. Addition of True Positive and False Negative. In ourcase, it represents the number of news articles predicted as trueout of the total number of true news articles.

$$Recall = \frac{TP}{TP + FN}$$

3) *Precision:* precision score represents the ratio of true positives to all events predicted as true. In our case, precision shows the number of articles that are marked as true out ofall the positively predicted (true) articles:

$$Precision = \frac{TP}{TP + FP}$$

4) *FI Score:* F1-score represents the harmonic mean between precision and recall. It calculates the relation betweeneach of the two. Thus, it takes both the false positive andthe false negative observations into account. F1-score can be calculated using the following formula:

$$FIScore = 2\frac{Precision * Recall}{Precision + Recall}$$

V. ALGORITHAMS

A. TF (Term Frequency)

The number of times a word appears in a paragraph is its Term Frequency. A higher value means a term appears more often than others, and so, the document is a good match when the term is part of the search terms.

B. IDF (Inverse Document Frequency)

Words that occur many times a document, but also occur many times in many others, may be irrelevant. IDF is a measure of how significant a term is in the entire corpus.
The TfidfVectorizer converts a collection of raw documentsinto a matrix of TF-IDF features.

C. NLP (Natural Language Processing)

Natural Language Processing is used for data mining. data isdivided into many parts. i.e. Tokenization, stemming, Lemmatization, POS Tags, Name Entity Recognition, Chunking. Tokenization is a way of separating a piece of text into smaller units called tokens. Here, tokens can be either words, characters, or subwords. Hence, tokenization can be broadly classified into 3 types – word, character, and sub-word (n- gram characters) tokenization Stemming is is the removes the past future tenses from root word i.e. waited, waits, waitingit gives wait which is root word. Lemmatization is gives the Proper root word POS tags is Part Of Speech Tags this step will removes grammer. Name Entity Recognition in this stepall type of classification will store i.e. location, name, map etc. Chunking is the last step in whole NLP in this step all divided and sorted words is combine into specific sentences.

D. Logistic Regression

Logistic regression is a type of supervised machine learningused to predict the probability of a target variable. It is used to estimate the relationship between a dependent (target) variable and one or more independent variables. The output of the dependent variable is represented in discrete valuessuch as 0(False)and 1(True). The Sigmoid function (logistic

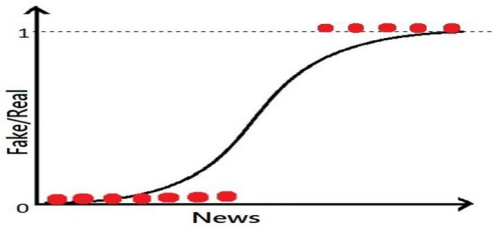


Fig. 2. Logistic Regression

©IJRASET: All Rights are Reserved

Regression model) is used to map the predicted predictions to probabilities. The Sigmoid function represents an 'S' shaped curve when plotted on a map. The graph plots the predicted values between 0 and 1. The values are then plotted towards the margins at the top and the bottom of the Y-axis, with the labels as 0 and 1. Based on these values, the target variablecan be classified in either of the classes. The equation for the Sigmoid function is given as:

$$y = 1/(1 + ex)$$

where, ex= the exponential constant with a value of 2.718. This equation gives the value of y(predicted value) close to zero if x is a considerable negative value. Similarly, if the value of x is a large positive value, the value of y is predicted close to one. A decision boundary can be set to predict the class to which the data belongs. Based on the set value, the estimated values can be classified into classes. For instance,let us take the example of classifying emails as spam or not.If the predicted value(p) is less than 0.5, then the email is classified spam and vice versa.

E. Support Vector Machine

Support vector machines (SVMs) are supervised machine learning algorithms which are used both for classification and regression An SVM model is basically a representation of different classes in a hyperplane in multidimensional space.In our paper data is consist of two categories i.e. Fake or real.

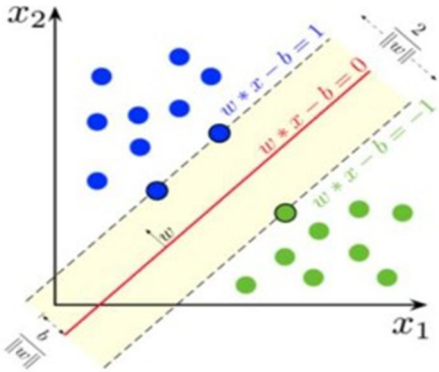


Fig. 3. SVM graph.

Based on two classes we are train the dataset. There are n numbers of records arranged in (x1+y2) format.

F. Native bayes

Naive bayes is the set of Algorithms which we used for implementation in our paper i.e. Natural Language Processing, Support Vector Machine etc. Bayes algorithm is improves the performance of the system.

Laplace smoothing is a smoothing technique that helps get the problem of zero probability in the Native Bayes machine learning algorithm

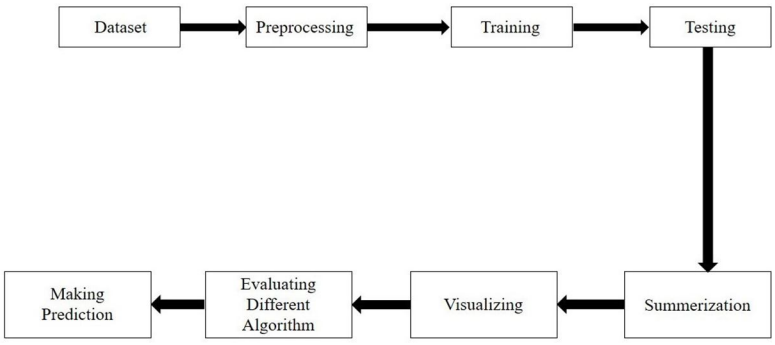


Fig. 4. Pipeline Representation

©IJRASET: All Rights are Reserved

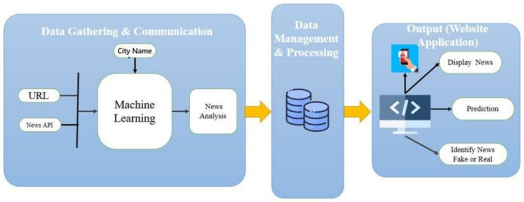


Fig. 5. System Architecture.

VI. SYSTEM ARCHITECTURE

As mention in Fig 5 ,System is divided into 3 parts i.e. 1.Data Gathering and communication 2.Data management and Processing 3.Output (Web Application)

1) *Data Gathering:* Data is fetched from different types of urls of Newspaper and there is news API cloud we are fetching all type of news from many sources.
2) *Data Management and Processing:* We are used MySQL database to store the whole data And for processing all Algorithms which is mention in paper.
3) *Output:* As our title we are fetching Location based news as well as by entering city name you can find the news.

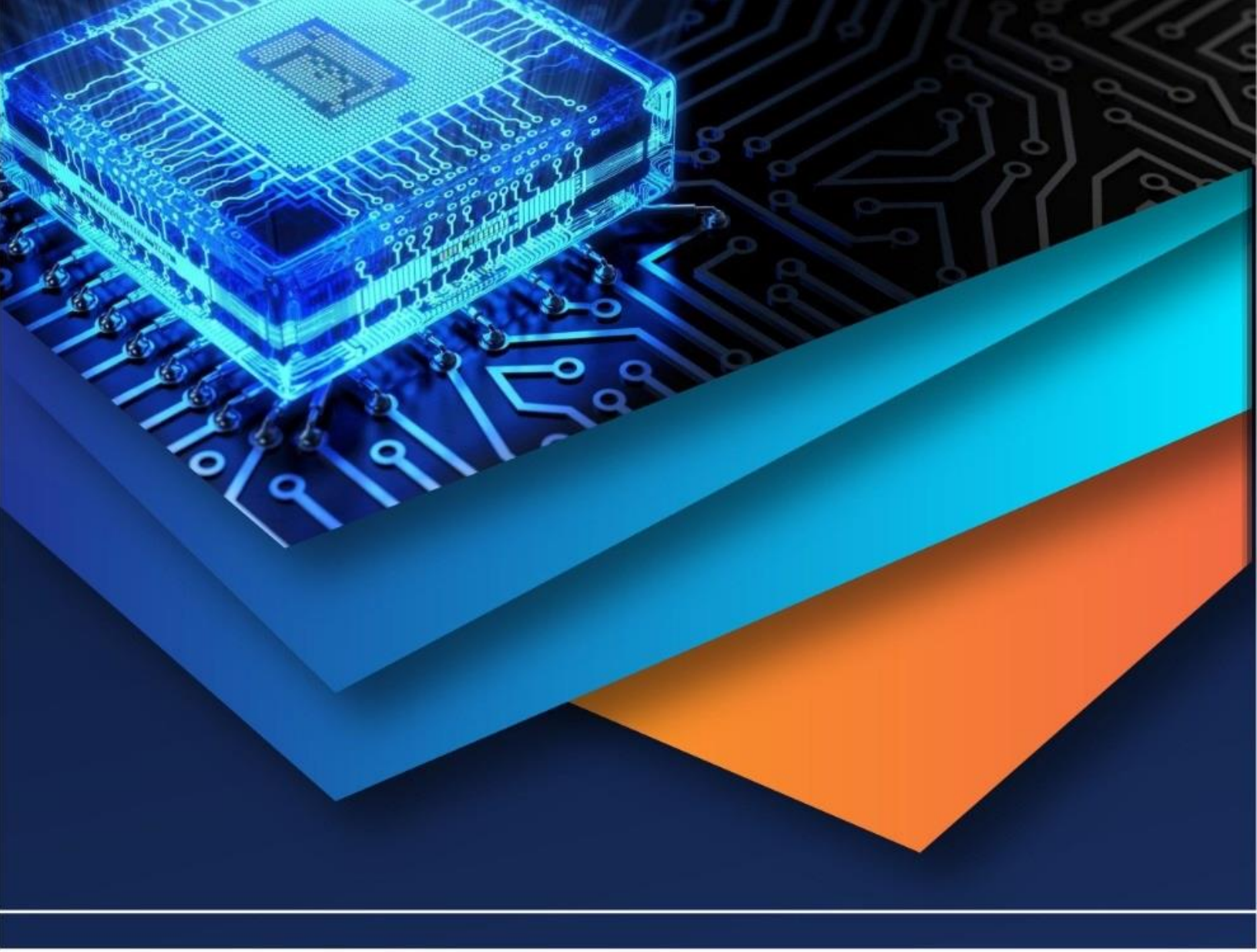
VII. CONCLUSION

There is a success to find the crime news for particular location. After entering pin code display all crime news which is belonged to particular city. Predict the future of crime news and shown on graph. Crime will be increment or decrement in upcoming days. There is evident success in detection news fake or real by using various machine learning approaches. By using Logistic regression, naive bayes and Support vector ma- chine will find news is fake or real.by using linear regression will predict the future rate of crimes.

REFERENCES

[1] Marlie Celliers and Marie Hattingh Department of In- formatics, University of Pretoria, Pretoria, South Africa.
[2] IEC South Africa: Real411. Keeping it real in digital media. Disinformation Destroys Democracy (2019)
[3] Figueira, Á., Oliveira, L.: The current state of fake news: challenges and opportunities. Proc. Computer. Sci. 121, 817–825(2017)
[4] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
[5] Rada Mihalcea , Carlo Strapparava, The lie detector: explorations in the automatic recognition of deceptive language, Proceedings of the ACL-IJCNLP
[6] H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar,"A framework for real-time spam detection in Twitter," 2018 10th International Conference on Communication Systems andNetworks (COMSNETS), Bengaluru, 2018, pp. 380-383
[7] M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M.DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," 2018 22ndConference of Open Innovations Association (FRUCT), Jy- vaskyla, 2018, pp. 272-279.
[8] Conroy, N. J., Rubin, V. L., Chen, Y. (2015,November). Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIST Annual Meeting: Information Science with Impact: Research in and for the Community.American Society for Information Science.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)