iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45239

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: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Position Forecast on Twitter Using Machine Learning Techniques

T. Nithisha[1], T. Deepthi Reddy[2], M. Sangeetha[3]
[1, 2, 3]*Department of Computer Science And Engineering, Sridevi Women's Engineering College,Vattinagulapally,Gandipet, R.R.DIST -500075, India*

Abstract: ***These days, a lot of study is being done on the location predictions of users from online social media. It has been studied for decades how to automatically identify locations that is connected to or mentioned in documents. As a leader in the online social network industry, Twitter has attracted a sizable user base that regularly sends millions of tweets. Location prediction on Twitter has gained significant attention lately as a result of the global reach of its users and the constant stream of posts. Researchers face several difficulties when attempting to do study using tweets, which are brief, noisy, and rich in content. A broad overview of location prediction using tweets is examined in the suggested framework. In particular, tweet contents are used to forecast tweet location. By describing the content and context of tweets, it is fundamentally highlighted how these text inputs play a role in the problems. In this study, we employ naive bayes, support vector machines, and decision trees as machine learning approaches to predict the user's location from the text of tweets.***

I. INTRODUCTION

Users can directly state their location in the text of a tweet, however in some circumstances the location may be implied by the inclusion of certain pertinent characteristics. Users may share informal visuals with emotions in tweets because they are not a strongly typed language. Tweet texts are noisy because of abbreviations, typos, and extra characters added to words with strong emotions. The methods used to analyse documents other than tweets are not appropriate. If the context of the tweet is not understood, the 140-character character limit of tweets may make them difficult to understand.

We included two sub-modules of the area specified in this study: The first is identifying the location that is mentioned in a tweet by extracting text information from a tweet that uses geography names. The second involves matching tweet content to entries in a geographical database to determine the location.

II. OBJECTIVE

The main objective of this project is to predict the person's location using their tweets and hashtags.

1) *Home Location:* The user's residential address or the location they specified when opening their account are both regarded as their homes. Home location prediction can be used for a variety of purposes, including polling, location-based advertising, recommendation systems, and health monitoring. You can specify your home address using administrative, geographic, or coordinate locations.
2) *Tweet Location:* The area from which a user posts a tweet is referred to as the tweet location. One can determine a tweeter's mobility by interpreting their tweet location. Home location is typically gathered from a user's profile, although a tweet's location can come from a user's geotag. POIs are widely accepted as representations of tweet regions as a result of the initial perspectives on tweet location.
3) *Mentioned Location:* When writing tweets, users may mention a few specific locations by name. Referenced location prediction could lead to improved comprehension of tweet content.

III. METHODOLOGY AND DATABASE USED

The location prediction problem using tweet content and social media material has been researched by researchers using a variety of existing methodologies, some of which are detailed below. The author discusses the difficulty of deriving location from social media posts in [1]. The authors of [1] and [2] arrived at inverse city frequency (ICF) and inverse location frequency (ILF), respectively, driven by term frequency (TF) and inverse document frequency (IDF). These frequency values were used to rake the features, followed by TF values. This led them to the conclusion that local words are hardly used in documents and have high ICF and ILF values.

In their research, Han et al. [3] utilised a modelto find local words that were specific to or onlyused in certain locations. They sought to identify automatically by classifying the local words according to where they were used, and they determined the degree to which location words were linked to particular locations orcities. Twitter data is collected in real-time as adataset utilising authentication keys. The suggested system's goal is to determine a user'slocation from tweet content by taking into account the user's home address as well as tweetlocation and content. To deal with this, we applied three machine learning techniques, selecting the most effective model among them.Fig. 1 shows the proposed system's general design as well as various technique modules.

The "twitter.json" file contains the live tweet stream that Twitter has collected for the term "apple." By registering a consumer key, consumer secret, access token, and access tokensecret for authentication and gathering a live stream of tweets, live twitter data can be gathered. More than 1000 tweets containing specific keywords, such as "Chennai, Mumbai, Kerala," have been gathered by us. Tweetid, name, screen name, tweet text, home location, tweet location, mentioned location, and Lvalue are among the data taken from live.

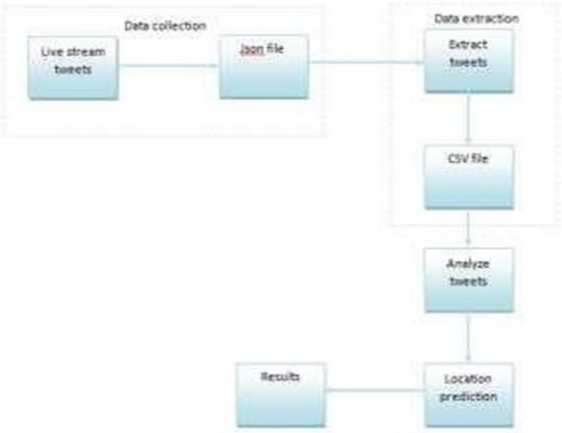


Fig.1: System Architecture

IV. ALGORITHMS

The algorithms used in this work are Naïve Based Classification, Support Vector Machine, and Decision Tree algorithm.

A. Naive Bayes Classification

The most well-known andstraightforward classification model is the naiveBayes model. The word distribution in the document is used by this model to determine theposterior probability. The Bag Of Words (BOW) feature extraction approach used by theNaive Bayes classifier does not take the positionof the word within the document into account. In order to predict a certain label from the provided feature set, this model applied the Bayes Theorem. A trainset and a testset arecreated from the dataset. NB model is used to find the position prediction on the test set.

B. Support Vector Machine

One of the most popular supervised learning methods is the support vector machine,which is frequently applied to bothclassification and regression issues. Each piece of data is plotted as a point in n-dimensional space according to how the algorithm isdesigned, with the feature values acting as the values of each co-ordinate.

C. Decision Tree

The learning model that makes use of the classifications problem is the decision tree. The dataset is divided into at least two sets for the decision tree module to operate. Internal nodes of a decision tree represent a test of the characteristics, branches show the results, and leaves represent decisions reached once atraining procedure has succeeded.

This is how a decision tree functions.

1) The root node of the decision tree is connectedto all training cases in the beginning.
2) The dataset is divided into a train set and a testset.
3) It chooses attributes to label each node based on information it gathers. Information with a similar feature property is present in subsets.

4) The aforementioned procedure is continued ineach subgroup until leaves are produced in a tree.

No root to leaf node path in the tree contains the same property more than once.Every sub tree on the training instances, which is classified through the path in the tree, isconstructed by repeating this action. The class label prediction issue begins at the tree's root foreach record in the dataset. For the specified record, the root attributes are examined before moving on to the subsequent record's attributes.The operation keeps going until the value next node is removed.

V. RESULT

We used the pre-processed dataset for the machine learning process and used the NaiveBayes, SVM, and Decision Tree algorithms to the data. We used the dataset, of which 80% wasused for training and 20% for testing, to predictthe position and compare accuracy. The performance evaluation of three machine learning algorithms, Naive Bayes, Support Vector Machine (SVM), and Decision Tree, is shown in the table below. Accuracy of prediction is one of the metrics for evaluation shown in the table. The table shows unequivocally that decision trees perform betterthan the other algorithms in terms of accuracy and efficiency.

VI. CONCLUSION

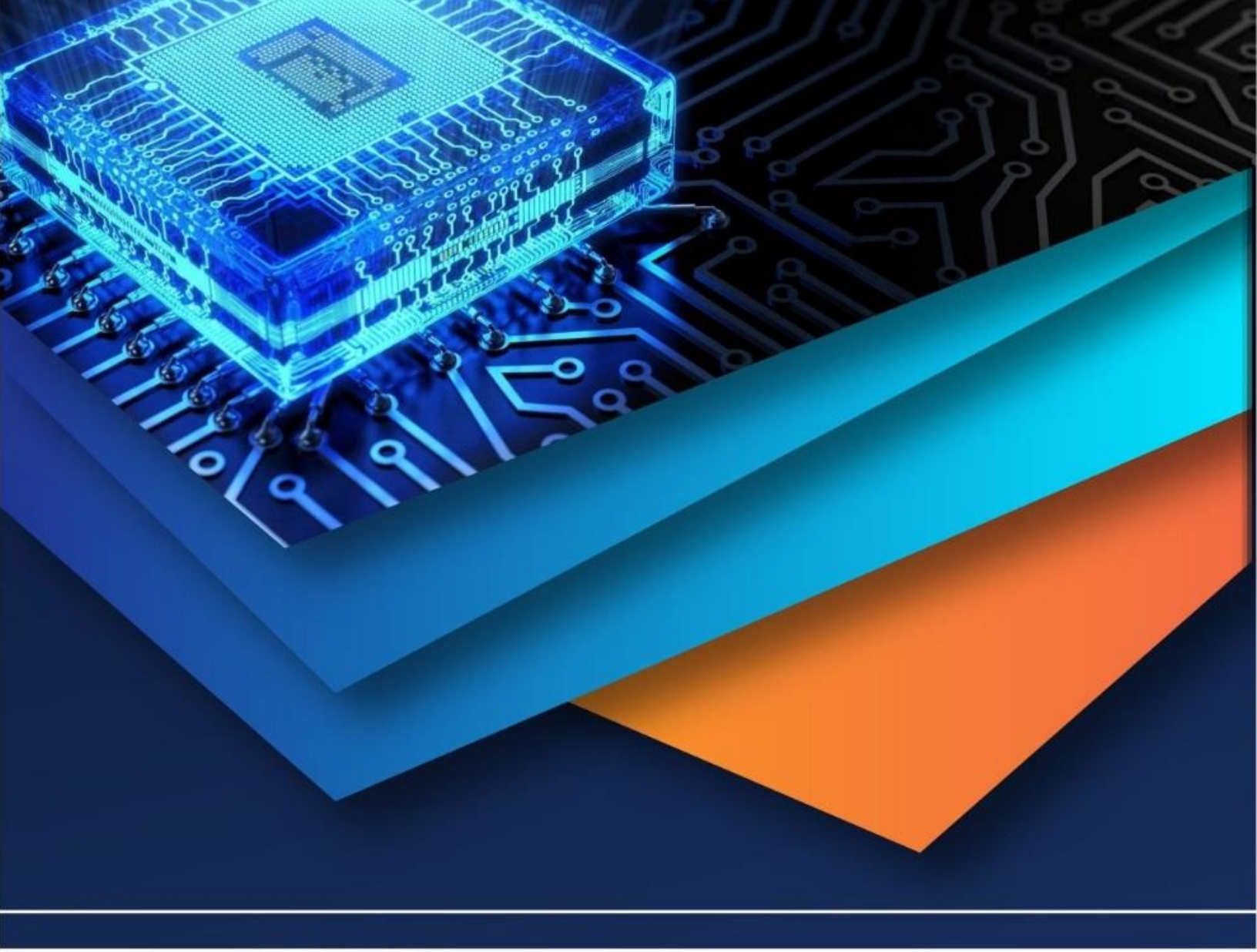
From the information provided byTwitter, three locations—home location, mentioned place, and tweet location—are taken into account. Geolocation prediction becomes adifficult challenge when the twitter data is takeninto account. The limited character count and nature of tweet language make it challenging tocomprehend and evaluate. In this study, we usedmachine learning techniques to predict a user's geolocation based on the text of their tweets. Three algorithms have been put into practise in order to demonstrate the best one that works best for the problem of geolocation prediction. According to the results of our experiment analysis, decision trees are appropriate for problems involving location prediction and tweet text processing.

REFERENCES

[1] Han, Bo & Cook, Paul & Baldwin, Timothy. (2012). GeolocationPrediction in Social Media Data by FindingLocation Indicative Words. 24th International Conference on Computational Linguistics - Proceedings of COLING 2012: Technical Papers. 1045-1062.

[2] Ren K., Zhang S., Lin H. (2012) Where Are You Settling Down: Geo- locating Twitter Users Based on Tweets and Social Networks. In: HouY., Nie JY., Sun L., Wang B., Zhang P. (eds) Information Retrieval Technology. AIRS 2012.Lecture Notes in Computer Science, vol 7675.Springer, Berlin, Heidelberg.

[3] Han, Bo & Cook, Paul & Baldwin, Timothy.(2014). Text-Based Twitter User Geolocation Prediction. The Journal of Artificial Intelligence Research (JAIR).49. 10.1613/jair.4200.



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)