LOCATION PREDICTION ON TWITTER USING MACHINE LEARNING TECHNIQUES

Mr. A. Saibabu[1], A. Meghana[2], N. Bhavya Shree[3], U. Vidhya nandini[4]

Article History: Received: 07.05.2023 **Revised:** 19.06.2023 **Accepted:** 14.07.2023

Abstract:

There is a lot of interest in figuring out how to forecast online activity. Research on the possibility of automatic recognition of locations mentioned or mentioned in documents dates back decades. Twitter has attracted a massive audience, and every day, millions of people use it to broadcast their thoughts online. Twitter's tweets has given much attention as of late. Research into both loud and nature, presents number of difficulties. picture investigated instance, the location of a tweet may be predicted based on its content. Issues' reliance on text inputs is highlighted by providing an overview of tweet content and circumstances. In this study, we use methods Machine, and Decision Tree to estimate a based on a content.

Keywords: Tweet content, decision tree, online.

---

[1]Assistant Professor, Department of CSE, Sridevi Women's Engineering college, Hyderabad, Telangana, India.
[2,3,4]UG Student, Department of CSE, Sridevi Women's Engineering college, Hyderabad, Telangana, India.

Email: [1]swecsaibabuyadav@gmail.com, [2]meghanaalapati31@gmail.com, [3]nelluribhavyasree1217@gmail.com, [4]vidhyanandini300@gmail.com

DOI: 10.31838/ecb/2023.12.s3.692

1. SCOPE

In certain implicitly accessible relevant parameters, while in others, users may declare their location directly on the tweet text they submit. Users may submit informal, emotion-laden photos in their tweets since it is not a firmly written language. Tweets are distracting because of their abbreviated style, spelling errors, and emotive phrases that add unnecessary characters. Normal document analysis methods are not applicable to tweets.

2. INTRODUCTION

If the tweet's context is not carefully examined, the character constraint of about 140 characters may make the tweet difficult to interpret. For web page and Wikipedia content, we investigate a location prediction-related problem called geolocation prediction. There has been much study on the topic of identifying entities using legal records. forms of management on these papers. Nonetheless, tweet content plays a significant role in the location prediction issue from Twitter. Local users may research attractions, landmarks, and events happening in their immediate area. Your home location is your residence address or the place you specified when you created your account. systems, ads, health monitoring, polling, and so on may all benefit from accurate predictions of where people are at home. Administrative location, geographical location, or coordinates may all be used to describe a person's abode. The location of a tweet is the geographic area from where it was sent. One's movement may be inferred from tweet location. Although a user's residence may be gleaned from their profile, a tweet's geotag can lead to their exact location. The first viewpoints on tweet location ensure that POIs are universally understood as symbols for tweet areas. Places Mentioned in Tweets: Users may include the names of certain places in their tweets. Predicting where a tweet was sent from might improve recommendation systems, location-based ads, health monitoring, polling, and other services that rely on a thorough familiarity with tweet content. Two components of the aforementioned locale are included in the present investigation: The first is to identify the location from the tweet's text, which may be done by parsing the tweet for geographical terms. The second method involves extracting location information from tweets by matching keywords to coordinates stored in a database.

3. RELATED WORK

"Geolocation Prediction in Social Media Data by Finding Location Indicative Words"
Localized lookup and local event detection are only two examples of how geolocation prediction is useful in the geospatial realm. Whole text data, which includes frequently used terms with no geographical dimension (like "today") and noisy strings (like "tmrw"), is often used as the basis for social media geolocation models, which may reduce prediction accuracy and cause models to be slower and more memory-intensive. In this study, we investigate whether a smaller number of features might improve geolocation accuracy by focusing on the problem of identifying location indicating words (LIWs). Our findings reveal that an information gain ratio-based approach excels at LIW selection, with mean and median prediction error distances being reduced by 45kilometres 209kilometres, respectively, on a publicly available dataset, outperforming government geolocation prediction models by 10.6% in accuracy. In addition, we develop ideas of prediction confidence and show that performance is even better when our algorithm is more confident, finding a balance between precision and breadth. In conclusion, the discovered LIWs provide light on linguistic diversity at the regional level, which might be valuable for future lexicographers.

"Geo-locating Twitter Users Based on Tweets and Social Networks"
In this work, we look at the benefits of constructing models for forecasting where individuals settle down based on their profiles, which provide information about cities and towns. We propose two local word filters, Inverse Location Frequency (o and Remote Words (RW) filter, to detect local terms in tweets content based on the people that willingly provide their locations in their accounts. Using the (Named Entity Recognition)NER software, we also extract the location referenced in tweets, which is then filtered using the city distance. Friends of the individual and 2-hop followers are taken into account. Finally, we integrate these two dimensions in our experiment to estimate user

location, achieving Accuracy of 56.6% within 100 miles in city-level and 45.2% within 25 miles in town-level, both of which are improvements over the single-dimensional prediction and the baseline. **"Text-Based Twitter User Geolocation Prediction."**

Instant search and incident detection are only two examples of geospatial applications that rely heavily on location data. Here, we go into the challenge of predicting Twitter users' whereabouts based on their tweets and provide several solutions. Geographical references (such as gazetteer terms or dialectal phrases) in a text have been used in previous research on this issue on the assumption that they reveal something about the author's place. It is not easy to unearth these references since they are often buried in unstructured, illogical, and multilingual data. We provide a unified geolocation represent a potential and delve into the elements that affect prediction quality. Prior to obtaining "location indicating words," we assess a variety of feature selection techniques. Next, we assess how language and user-declared information, in addition to the lack of geotags in tweets, affect geolocation prediction. We also address the varying degrees of geolocatability across users and assess the effect of temporal variation on model generalisation. In addition to providing the deepest study of the text-based Twitter user localization challenge to date, we obtain state-of-the-art results. Our research results are helpful for developing more reliable and usable text-based geolocation prediction systems.

4. METHODOLOGY

Data from Twitter is gathered To achieve this goal, the suggested system takes into account home of the tweet, and the content of tweet to make a prediction about the user's position. In order to tackle this, we used three distinct machine learning strategies for making predictions and selecting the most effective model. The graphic below depicts the proposed system's architecture in its entirety, including the many components that make up the aforementioned technique.

The file 'twitter.json' contains a collection of real-time tweets from Twitter with the term 'apple. You may gather real-time data from Twitter by signing up for Around thousand including terms like "Chennai," "Mumbai," and "Kerala" compiled. content, L-value all directly tweets' text was processed first for primary analysis. To do this, all of a user's tweets were combined into a single "document" and analysed at once.

5. RESULTS

This study examines the use of SVM, Naive Bayes, and Decision Tree, three machine learning methods, to estimate a user's location on Twitter. We've retrieved several tweets from Twitter, including tweet messages and tweet locations, to use for location prediction. To forecast user locations, we will train the aforementioned machine learning algorithms on a dataset of tweets. We have developed the following components to carry out this project. The 'Dataset' folder contains the datasets used to train the aforementioned ML algorithms.

The first row of a dataset usually consists of the column headings, while the subsequent rows include the data itself; for example, in the dataset cited above, the "text and location" column contains the tweets themselves, while the "location" column contains their respective locations. In the meanwhile, we have a test dataset consisting simply of tweets, and ML is predicting their locations.
We must first upload the dataset before making any predictions.

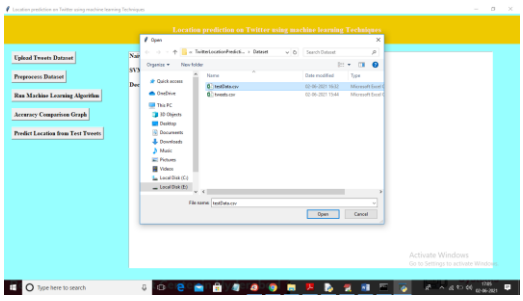

Figure 1 Upload Test data

With the aid of a decision algorithm, we can predict the position with great precision.

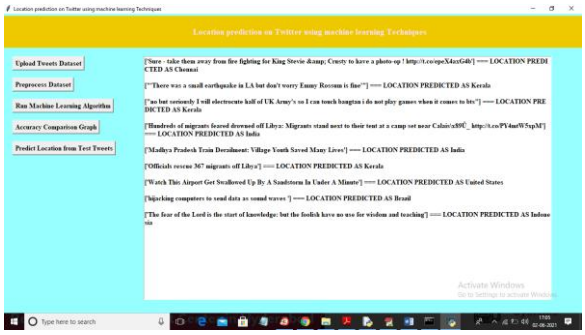

Figure 2 Location predicted

6. CONCLUSION

Data from Twitter is analysed in terms of three locations: home, mentioned, and tweet. Using Twitter's data complicates an already difficult challenge of geolocation prediction. informal tone tweets them to algorithms determine location tweet content. We've constructed three different algorithms to demonstrate which one is most suited to solving the geolocation prediction challenge. Based on our examination of the data we gathered from our experiments, we found that decision trees are a good fit for the text analysis and location prediction problems in tweets.

7. REFERENCES

1. Han, Bo & Cook, Paul & Baldwin, Timothy. (2012). Geolocation Prediction in Social Media Data by Finding Location 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: Hou Y., 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.
4. Li, Rui & Wang, Shengjie & Chen-Chuan Chang, Kevin. (2012). Multiple Location Profiling for Users and Relationships from Social Network and Content. Proceedings of the VLDB Endowment. 5. 10.14778/2350229.2350273.
5. Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. 2014. Home Location Identification of Twitter Users. ACM Trans. Intell. Syst. Technol. 5, 3, Article 47 (July 2014), 21 pages.
6. Miura, Yasuhide, Motoki Taniguchi, Tomoki Taniguchi and Tomoko Ohkuma. “A Simple Scalable Neural Networks based Model for Geolocation Prediction in Twitter.” NUT@COLING (2016).