LOCATION PREDICTION ON TWITTER USING MACHINE LEARNING TECHNIQUES

M.Swathi[1], P. Rohini[2]

[1]Assistant Professor, Dept of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, Tirupati (Dt), AP, India.

[2]PG Scholar, Dept of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, Tirupati (Dt), AP, India.

ABSTRACT_ Nowadays, there is a lot of study being done on location prediction of people via online social media. Decades of research have been dedicated to the automatic recognition of locations associated to or referenced in documents. Twitter is a prominent player in the online social network space, with a large user base that regularly sends millions of tweets. These days, location prediction on Twitter has garnered significant attention due to its global user base and constant stream of posts. Researchers face numerous obstacles in their field of study when it comes to tweets, which are brief, loud, rich messages. A general overview of location prediction using tweets is examined in the suggested framework. In particular, tweet location is predicted from tweet contents. By outlining tweet content and contexts, it is fundamentally featured that how the issues rely upon these text inputs. In this work, we predict the location of user from the tweet text exploiting machine learning techniques.

1.INTRODUCTION

The location of a user's location can be included explicitly in the tweet text they post, or it can be included implicitly in certain situations by including relevant criteria. Tweets are not a specifically language, in which clients might post easygoing with feeling pictures. Tweet texts are noisy because of their condensed form, misspellings, and extra characters for emotional words. For the purpose of analyzing tweets, the methods used for standard documents are insufficient. If the context of the tweet is not investigated, the 140-character limit may make the tweet difficult to comprehend. For Wikipedia and page documents, the topic of location prediction, also known as geolocation prediction, is examined. For many years, research has focused on entity recognition from these formal documents. Various kinds of content and setting taking care of on these archives are likewise concentrated widely. However, tweet content heavily influences the location prediction problem from Twitter. Users who live in specific

regions or locations can look up nearby attractions, landmarks, and buildings, as well as events related to those attractions. Place of Birth: Home location is considered to be the user's residential address or the location provided at account creation. Recommendation systems, location-based ads, health monitoring, polling, and other applications can all benefit from home location prediction. Co-ordinates, administrative location, or geographical location can be used to specify the home location. Where is the Tweet? The region from which a user posts a tweet is referred to as the tweet's location. One can determine a person's mobility by interpreting their tweet location. Normally home area gathered from client profile, while tweet area can be shown up from client's geo tag. POIs are widely accepted as representations of tweet regions due to the initial perspectives on tweet location. Location Mentioned: While creating tweets, client might make reference to the names of a couple of areas in tweet texts. Referred to area forecast might empower better comprehension of tweet content, and benefit applications like proposal frameworks, area based notices, wellbeing checking, and surveying and so forth. We include two sub-modules of the mentioned location in this study: The first is recognizing the location that is mentioned in the text of a tweet. To do

this, extract text from a tweet that uses geography names as references. The second method involves solving the tweet text for entries in a geographical database to determine the location.

2.LITERATURE SURVEY

Many existing techniques have been studied by the researchers on location prediction problem from tweet content and social media content, few of them are discussed below. In [1], the author refers to the problem of finding location from social media content. The author from [1] and [2] motivated by Term frequency (TF) and inverse document frequency (IDF), they arrived Inverse City Frequency (ICF) and Inverse Location Frequency (ILF) respectively. They raked the features by using these frequency values and TF then by TF values. From this they arrived that local words spread in document in few places and have high ICF and ILF values. Han et al [3] in their work, they approached model for identifying local words indicative or used in certainlocations only. They aimed to identify automatically by ranking the local words by their location, and they find their degree of association of location words associated to particular location or cities. Li et al. [4] proposed multiple locations profiling (MLP) model to arrive user location accurately by finding the

probability based on Bernoulli distribution. Their work represents that users home location can be predicted accurately using this model. The author used multinomial distribution to estimate probability of tweet versus the venue name from each location. Mahmud et al. proposed classification model for predicting location, they improved the accuracy of prediction by first predicting regions and then city. They registered the movement of users using classifier models, if the user travels for a certain period, then they are registered to improve the accuracy of prediction. The authors considered the person is travelling when the location distance for two tweets is more than 100 miles. Most of the techniques used in existing works are machine learning, whereas few works in deep learning also proposed. Miura et al. [6] on his work used neural network is implemented for twitter location prediction. The author classified tweet or user using neural networks and they integrated metadata with tweet texts and trained the model. Their model achieved around 41 percentage of accuracy on predictions

3.PROPOSED SYSTEM

Authentication keys are used to gather Twitter data in real-time streams as a dataset. The suggested system's goal is to forecast a user's location based on tweet content, home address, and tweet location. The "twitter.json" file contains the live tweet stream that is gathered from Twitter for the term "apple." By registering a consumer key, consumer secret, access token, and access token secret for authentication, as well as by gathering a live stream of tweets, live Twitter data can be gathered. More than a thousand tweets with specific keywords, such "Chennai, Mumbai, Kerala," have been gathered by us. Tweetid, name, screen name, tweet content, HomeLocation, TweetLocation, MentionedLocation, and Lvalue are among the data that was taken from live. The initial examination involved a fundamental processing of the tweets' content.

3.1 IMPLEMENTATION

1. **Train the Model**: Create and encode a dataset of sentences and cities, generate sentence embeddings using Sentence-BERT, and train a logistic regression model.

2. **Create the Streamlit App**: Load the pre-trained Sentence-BERT model, the trained location prediction model, and the label encoder in a Streamlit app. Create an interface to accept user input, generate embeddings, predict the city, and display the result.

3. **Run the App**: Use Streamlit to run the app, providing an interactive interface for users to input sentences and get city predictions.
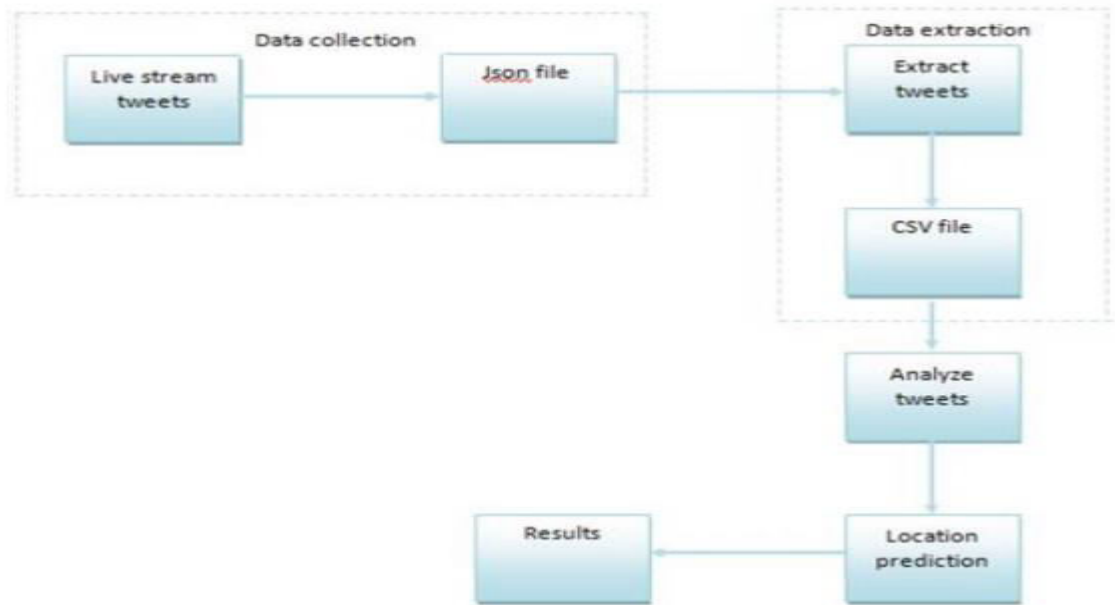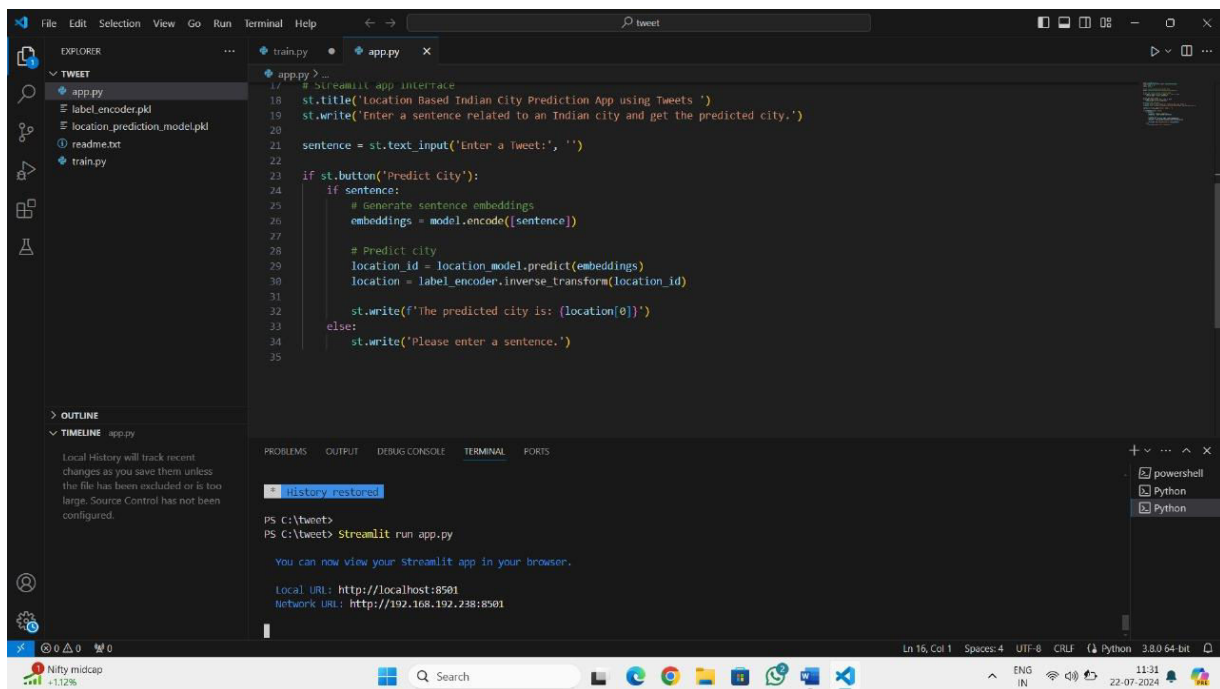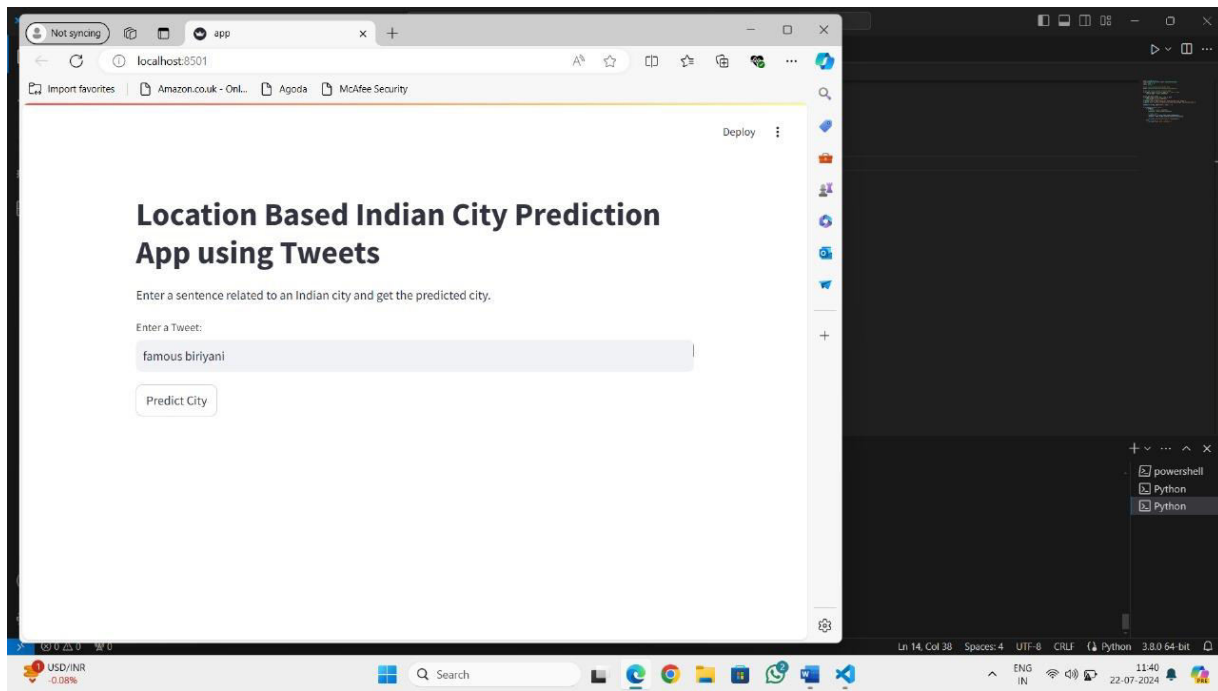

Fig 1:ARCHITECTURE

4.RESULTS AND DISCUSSION

5.CONCLUSION

Three locations—the home location, the location of the mention, and the location of the tweet—are taken into account using Twitter data. The challenge of geolocation prediction becomes difficult when taking into account the data from Twitter. The limited character count and textual format of tweets make them difficult to interpret and evaluate. In this work, we have used machine learning techniques to predict the user's geolocation based on the text of their tweets. To demonstrate the best-performing method that is appropriate for the geolocation prediction problem, we have constructed three different algorithms. We found from our experiment analysis that decision trees work well for location prediction problems and tweet text processing.

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. DOI: http://dx.doi.org/10.1145/2528548

[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).

[7] A. Schulz, A. Hadjakos, H. Paulheim, J. Nachtwey, and M. M¨ uhlh¨auser, "A multi-indicator approach for geolocalization of tweets," in Proc. 7th Int. Conf. on Weblogs and Social Media, 2013.

[8] R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang, "Towards social user profiling: unified and discriminative influence model for inferring home locations," in Proc. 18th ACM Int. Conf. on Knowledge Discovery and Data Mining, 2012, pp. 1023–1031.

[9] B. Han, P. Cook, and T. Baldwin, "A stacking-based approach to twitter user geolocation prediction," in Proc. 51st Annual Meeting of the Association for Computational Linguistics System Demonstrations, 2013, pp. 7–12.

[10] D. Flatow, M. Naaman, K. E. Xie, Y. Volkovich, and Y. Kanza, "On the accuracy of hyper-local geotagging of social media content," in Proc. 8th ACM Int. Conf. on Web Search and Data Mining, 2015, pp. 127–136.

[11] O. V. Laere, J. A. Quinn, S. Schockaert, and B. Dhoedt, "Spatially aware term selection for geotagging," IEEE Trans. Knowl. Data Eng., vol. 26, no. 1, pp. 221–234, 2014.

[12] J. Mahmud, J. Nichols, and C. Drews, "Where is this tweet from? inferring home locations of twitter users," in Proc. 6th Int. Conf. on Weblogs and Social Media, 2012.

Author Profiles

MRS. M .SWATHI has received her MCA degree from JNTUA in 2013.she is dedicated to teaching field from last 10 years .she has guided 5 P.G and 10 U.G students. currently she is working as Assistant professor in Audisankara college of engineering and technology, gudur.

P. ROHINI is pursuing MCA from Audisankara College of Engineering & Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.