*International Journal for Modern Trends in Science and Technology*, 8(09): 01-06, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0809001
Available online at: http://www.ijmtst.com/vol8issue09.html

# Location Tracking Web Application Using Machine Learning

**Priti Mishra | Poonam Bhogale**

Department of Computer Engineering, KJ Somaiya College of Engineering, Mumbai, Maharashtra, India
Email: mishrap4122@gmail.com

**To Cite this Article**
Priti Mishra and Poonam Bhogale. Location Tracking Web Application Using Machine Learning. International Journal for Modern Trends in Science and Technology 2022, 8(09), pp. 01-06. https://doi.org/10.46501/IJMTST0809001

**Article Info**
Received: 30 July 2022; Accepted: 26 August 2022; Published: 28 August 2022.

## ABSTRACT

*Many applications in intelligent transportation systems are demanding an accurate web application-based location prediction. In this study, we satisfy this demand by designing an automated mobile user location prediction system based on the well-known traditional Auto-Regressive Integrated Moving Average (ARIMA). To increase the proposed model accuracy, make it dynamic, and reduce its execution time, the traditional ARIMA model has been modified extensively by using different combinations of design options of the model. To perform user location prediction, the proposed model depends the previous recorded user locations to predict the user future locations. To make the proposed model dynamic, it is designed to regenerate all its parameters periodically. To deal with such dynamic environment, only a specified window of the historical data is used. To reduce the regeneration of the model execution time, the model selection process is enhanced and several model selection approaches are proposed.*

*The proposed model and the different design options are evaluated using a realistic user location dataset trace that are recorded using a WIFI embedded, as well as, using traces from a previous study called the Kaggle Dataset. To deal with any imperfection in the data used in generating the model in this study. The results show that the proposed framework can generate ARIMA models that can predict the future user locations of a user accurately and with a reasonable execution time. The results also show that the proposed model can predict the user's location for several future steps with an acceptable accuracy.*

***KEYWORDS:****Machine Learning, Geo-location, Arima Model, Accuracy.*

## 1. INTRODUCTION

It is almost obvious that people have certain routines in their daily life and these routines repeat daily, weekly and every month (e.g., a bank employee has 9–5 job so he is at the bank from Monday to Friday.) and sure, there are some exceptions like sickness and holidays. Still, user activities on Foursquare show that people demonstrate different patterns at the weekend compared to the rest of the week [1]. Based on researchers' studies [2], most people have low-entropy lives, which means their lives have long term regularity. This long-term regularity makes it available to understand and extract meaningful patterns out of their life.

Entropy Lives, which means their lives have long term regularity. This long-term regularity makes it available to understand and extract meaningful patterns out of their lives. Typically, the positioning technology for smartphones can be divided into GPS, Cellular, and Wi-Fi positioning [4]. There is also a lot of information, in where these certain people go; this information can be extracted and used as well in various ways like providing more applicable advertisements based on their personalities. Consequently, a plethora of publicly available data is generated from platforms like Foursquare and Twitter, and can potentially reflect the behavior of millions of citizens at a remarkable level of detail [5]. It is vital to discover mobility patterns to predict the next location, and it is impossible to reach, without having the user's spatial background since pattern discovery and prediction are closely related tasks and repetitive. When it comes to predicting, two types of approaches are usually got considered. The first one is the Markov model, and the second one is the decision tree, both of which are appropriate; however, there have been few pieces of research that used ANNs for this purpose.

Alharbi and de Doncker [8] used a Convolutional Neural Network to analyze user behaviors on Twitter. Monreale et al. [9] presented the Where Next method, which predicted the next location based on a decision tree. Oh [10] also used a decision tree method for predicting user's locations based on their past movement patterns in an indoor environment with a data from only four users. Ying et al. [11] prediction model was based on a cluster-based prediction strategy. Their model evaluates the next location of a mobile user based on the frequent behaviors of similar users in the same cluster which determined by analyzing users' common behavior in semantic trajectories. A method called DestPD was proposed by Yang et al. [12] to overcome problems such as heavy computation, and data sparsity. This method was consisted on two phases of offline training and online prediction. They also proposed two data constructs: Efficient Transition Probability (ETP) and Transition Probabilities with Detours (TPD) to improve the efficiency of matrix multiplication.

Ashbrook and Sterner [13] presented a system to cluster GPS data automatically into meaningful locations to get used into a Markov model to predict user next location. Similarly, Do and Gattica-Perez [14] used various personalized Markov models, and developed a framework to predict the user's future location and the applications for the next 10 min. Likewise, Lu et al. [15] used a hidden Markov model with one huge difference which was considering traffic times with a few other factors. Moreover, Chen et al. [16] used three basic Markov models for three tasks. A Global Markov Model that uses allavailable trajectories to discover global behaviors, the Personal Markov Model that focuses on mining the individual patterns of each moving object, and the Regional Markov Model that clusters the trajectories to mine the similar movement patterns. He et al. [17] used a transition graph to support efficient sub-trajectories concatenation to tackle the sparsity issue. They also developed a novel similarity metric to measure the similarity between two sets of trajectories to validate whether the reconstructed trajectory set can well represent the original traces. Du et al. [18] proposed a Continuous Time Series Markov Model (CTS-MM) to predict the next location of users instantly. Herder et al. [19] also used Markov Model with five different approaches (top visited places, last visited places, hours spent, closest locations, and simple Markov Model) to predict the next visited location. Additionally, Gambs et al. [20] created a mobility model called Mobility Markov Chain (MMC) to incorporate the n previous visited locations and developed an algorithm for next location prediction based on their model, and their best result was with n = 2.

Li et al. [21] presented a framework, referred to as hierarchical-graph-based similarity measurement (HGSM), for Geographic Information Systems (GIS) to consistently model each individual's location history, and effectively measure the similarity among users. With the same purpose, Xiao et al. [22] used user's GPS trajectories with a Semantic Location History (SLH) to determine user's interests, and understand the similarity among users beyond geographic positions. Chen et al. [23] proposed a new model called Human Mobility Representation Model (HMRM) to predict the users' next location, and simultaneously produce the vector representations of the data which was consisted by user ID, location ID, and timestamp and activity type.

## 2. Related Work

In this project, the training and testing of the ARIMA model is the next phase after finding the optimum p, d and q values of ARIMA model. The training and testing data is split in a ratio of 70:30, where in the 70% of data is trained and remaining 30% of the data is used for testing in the model. The model is fitted and a model prediction object is created for further forecasting process. After the implementation of ARIMA model, with the optimum values, the prediction values are depicted using a visualization plot, with the Actual location and Predicted location of the dataset. The input dataset taken in our model is the real time dataset. The database is a large database of user location with different locations which is commonly used for training in system.From dataset we have taken 1000 training locations and 700 testing locations.

The locations in the dataset are present in form of an array consisting of Lat Long values representing locations along with their labels.

In this proposed system to an approach to predicting the movement of a mobile user based on historic data on their prior movements. In mobility management there are two key operations: location updating and paging. Location update is to collect mobile users' current location while paging is used to locate a mobile user. For high accuracy for models depends on the training dataset. Once a user's location has been identified the system identifies the users next location. The 'Future location prediction' method is used so that the system can take action before the user arrives at the destination. Location History keeps record of previous calculated locations. There are 2 reasons for keeping this type of context. Live Data of a number of people is collected and it is divided for training and test cases. Training data set is used for learning the system which is then further used for predicting the future location of the user. Test data set is used to check the accuracy of the system similar way as that of the training data set which is then checked using the stats.The input dataset taken in our model is user dataset. The database is a large database of user's information of regions which consist of which is commonly used for training in various image processing systems. The database is also widely used for training and testing in the field of machine learning.The quality of service for the location prediction affects the user's daily activity. location prediction is one of the solutions to provide this type of service. It is essential to provide better system performance in an intelligent transport system. However, the prediction models that have been done so far by different researchers are not adequate to provide a robust next location prediction. A user's next location prediction plays a vital role in location-based services, recommender systems, and network resource optimization. location prediction needs comprehensive investigations to enable users to use these plenty of applications. It is useful for proactive actions taken to assist a person in a constantly changing environment.Themovement regularities can be temporal, periodic, or sequential.

## 3. Block Diagram

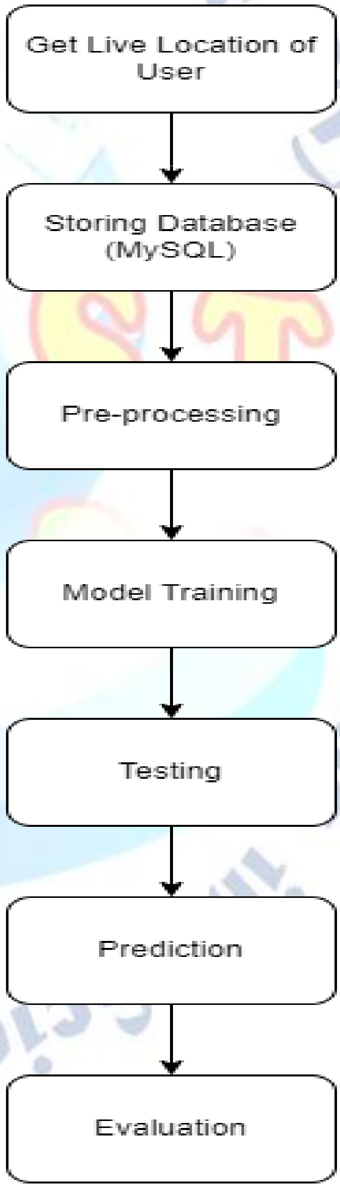


*Fig. 1. Architecture of the model*

## 4. Implementation

**1.Dataset: -**

This data set is focused on WLAN fingerprint positioning technologies and methodologies (also known as WIFI Fingerprinting). It was the official database used in the IPIN2015 competition. The database covers three buildings of university with 4 or more floors and almost 110.000m2. It can be used for classification. It was created by means of more than 20 different users and 25 Android devices. The database consists of 19937 training/reference records

The 529 attributes contain the WIFI fingerprint, the coordinates where it was taken, and other useful information. Then the coordinates (latitude, longitude, floor) and Building ID are provided as the attributes to be predicted. The particular space (offices, labs, etc.) and the relative position (inside/outside the space) where the capture was taken have been recorded. Outside means that the capture was taken in front of the door of the space.

**2.Preprocessing: -**

After the collection training data, the next step will be to proceed to data preparation, loading the data to the appropriate location, and then preparing for training. The first part used to train the model will be most of the dataset, and the second part will be used to evaluate the performance of the trained model. The data cannot be fed directly into the model so we need to perform some operations and process the data to make it ready for our network. The model will require one more dimension or channel for preprocessing.

**3.Feature Extraction: -**

Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). These new reduced set of features should then be able to summarize most of the information contained in the original set of features. In this way, a summarized version of the original features can be created from a combination of the original set. This involves applying a suite of common or commonly useful data preparation techniques to the raw data, then aggregating all features together to create one large dataset, then fit and evaluate a model on this data.

**4.Training and Testing Process: -**

The most critical factor affecting the success of machine learning is the training and testing process. An effective training process improves the quality of the developed system. Researchers divide datasets into two parts for training and testing. However, the separation process is done according to specific rules. These are described in detail insection "Sampling Methods." The amount of training and test is the most critical factor in the success rate. If there is a high correlation between the features and the label, the Training-Test set is divided by 70%–30%. This means that 50% of all the data will be used for training and 30% for the test. However, if there is a fear of success falling, the rate of training can be increased. The training-testing ratio used in the literature varies according to the data structure. Less than 50% of the training data is not preferred because the test results will be negatively affected

After the machine learning model is trained according to the training data, it is also tested using the training data. The purpose of this is to determine how much data is learned. Performance evaluation procedures are performed according to specific criteria. These criteria vary according to the structure of the data. Section "Performance Evaluation Criteria" presents the performance evaluation criteria in detail.

Once the training process is completed, the machine learning model tested with test data has never been seen before. The researcher evaluates the test performance according to the performance evaluation criteria (section "Performance Evaluation Criteria"). The research can be repeated by changing the training and test data in the training and testing process to avoid the situation of unstable data. In this case, the researcher uses the average of performance values.

**5. Accuracy: -**

Accuracy is defined as the ratio of efficiently classified samples to overall samples. Accuracy is a suitable metric whilst the dataset is balanced. In actual network environments; however, everyday samples are far extra considerable than are unusual samples; hence, accuracy may not be an appropriate metric.

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

## 5. Results

In the location prediction of mobile users, multi-step prediction ability is also one of the keys to reflect the prediction performance of the model, that is, predicting the region of interest where the user is after n steps. In order to verify the multi-step prediction ability of the model, the prediction accuracy of the ARIMA model.

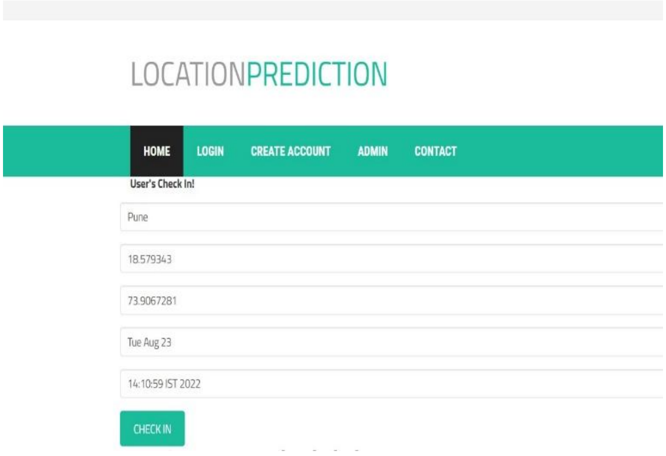


*Fig 1.User's Checkin Details*

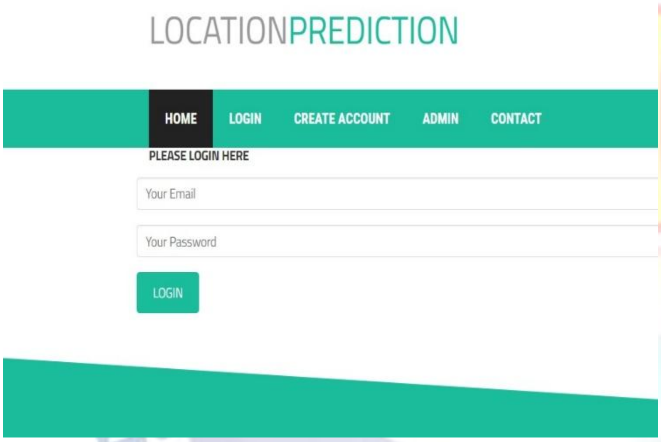


*Fig 2.Login Page*

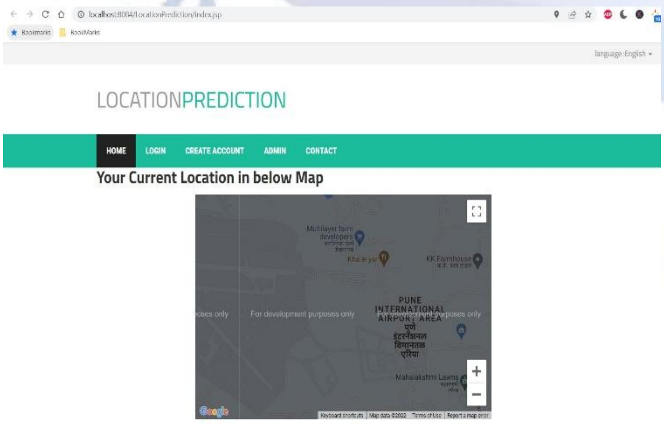


*Fig 3 Predicted Location of the user*

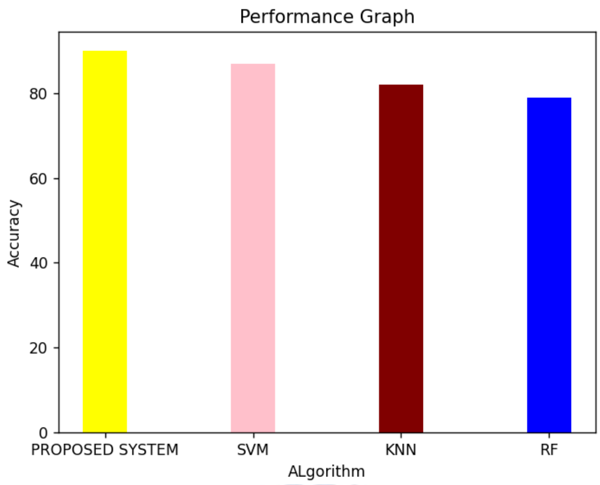


*Fig. 4. Performance Graph*

| Algorithm | Dataset | Accuracy (%) |
|---|---|---|
| SVM | Kaggle | 87 |
| KNN | Kaggle | 82 |
| RF | Kaggle | 79 |
| Proposed System (ARIMA) | Kaggle | 90 |

*Table 1. Accuracy Table for existing and proposed Algorithm*

## 6. Conclusion and Future work

This paper studies a location prediction method based on the proposed model for user trajectory data. Experiments on the real user live location data prove that the proposed method overcomes the low prediction accuracy of ML model and the high sparse rate of high-order proposed model to some extent, and makes full use of the user's prefix trajectory information. The next step will take into account factors such as weather, time (working days and rest days) and user-related social data, in order to further improve the prediction accuracy of the model.

## Conflict of interest statement

Authors declare that they do not have any conflict of interest.

## References

[1] Noulas, A., et al. (2011). Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. The Social Mobile Web, 11(2), 1–4.

[2] Eagle, N., & Pentland, A. (2006). Reality mining: Sensing complex social systems. Personal Ubiquitous Computing, 10(4), 255–268.

[3] Patrick, C. (2005). Life patterns: Structure from wearable sensors.

[4] Li, F., et al. (2019). Exploiting location-related behaviors without the GPS data on smartphones. Information Sciences, 527, 444–459.

[5] Roick, O., & Heuser, S. (2013). Location based social networks—definition, current state of the art and research agenda. Transactions in GIS, 17(5), 763–784.

[6] Gonzalez, M. C., Hidalgo, C. A., &Barabasi, A.-L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779.

[7] Clauset, A., & Eagle, N. (2007). Persistence and periodicity in a dynamic proximity network. In Proc. DIMACS, 2007.

[8] Alharbi, A. S. M., & de Doncker, E. (2019). Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. Cognitive Systems Research, 54, 50–61.

[9] Monreale, A., et al. (2009). WhereNext: A location predictor on trajectory pattern mining. In Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. Paris: ACM (pp. 637–646).

[10] Oh, S.-C. (2012). Using an adaptive search tree to predict user location. Journal of Information Processing Systems, 8, 437–444.

[11] Ying, J. J.-C., et al. (2011). Semantic trajectory mining for location prediction. In Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems. Chicago: ACM (pp. 34–43).

[12] Yang, Z., et al. (2018). An efficient destination prediction approach based on future trajectory prediction and transition matrix optimization. In IEEE transactions on knowledge and data engineering (pp. 1–1).

[13] Ashbrook, D., &Starner, T. (2003). Using GPS to learn significant locations and predict movement across multiple users. Personal Ubiquitous Comput., 7(5), 275–286.

[14] Do, T. M. T., &Gatica-Perez, D. (2014). Where and what: Using smartphones to predict next locations and applications in daily life. Pervasive and Mobile Computing, 12, 79–91.

[15] Lu, K., Hsu, C., & Yang, D. (2010). A novel approach for efficient and effective mining of mobile user behaviors. In 2010 4th international conference on multimedia and ubiquitous engineering.

[16] Chen, M., Yu, X., & Liu, Y. (2015). Mining moving patterns for predicting next location. Information Systems, 54, 156–168.

[17] He, D., et al. (2019). Efficient and robust data augmentation for trajectory analytics: A similarity-based approach. World Wide Web, 23, 361–387.

[18] Du, Y., et al. (2018). A geographical location prediction method based on continuous time series Markov model. PLoS ONE, 13(11), e0207063.

[19] Herder, E., Siehndel, P., & Kawase, R. (2014). Predicting user locations and trajectories (vol. 8538, pp. 86–97).

[20] Gambs, S., Killijian, M.-O., & Cortez, M. N. D. P. (2012). Next place prediction using mobility Markov chains. In Proceedings of the first workshop on measurement, privacy, and mobility. Bern:2012, Association for Computing Machinery (p. Article 3).

[21] Li, Q., et al. (2008). Mining user similarity based on location history (p. 34).

[22] Xiao, X., et al. (2010). Finding similar users using category-based location history (pp. 442–445).

[23] Chen, M., et al. (2020). Modeling spatial trajectories with attribute representation learning. In IEEE transactions on knowledge and data engineering (pp. 1–1).

[24] Zheng, J., & Ni, L. (2012). An unsupervised framework for sensing individual and clustebehavior patterns from human mobile data (pp. 153–162).

[25] Zheng, Y., et al. (2009). Mining interesting locations and travel sequences from GPS trajectories (pp. 791–800).

[26] Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks, 6(4), 525–533.

[27] Jafarian, H., &Behzadi, S. (2020). Evaluation of PM2.5 emissions in Tehran by means of remote sensing and regression models. Pollution, 6(3), 521–529.

[28] Adankon, M. M., &Cheriet, M. (2009). Support vector machine. In S. Z. Li & A. Jain (Eds.), Encyclopedia of biometrics (pp. 1303–1308). Boston: Springer.

[29] Adankon, M. M., &Cheriet, M. (2009). Support vector machine. In S. Z. Li & A. Jain (Eds.), Encyclopedia of biometrics (pp. 1303–1308). Boston: Springer.

[30] Black, J., Hashimzade, N., & Myles, G. D. (2009). A dictionary of economics. Oxford: Oxford University Press.

[31] Troncoso, A., et al. (2007). Electricity market price forecasting based on weighted nearest neighbors techniques. IEEE Transactions on Power Systems, 22, 1294–1301.