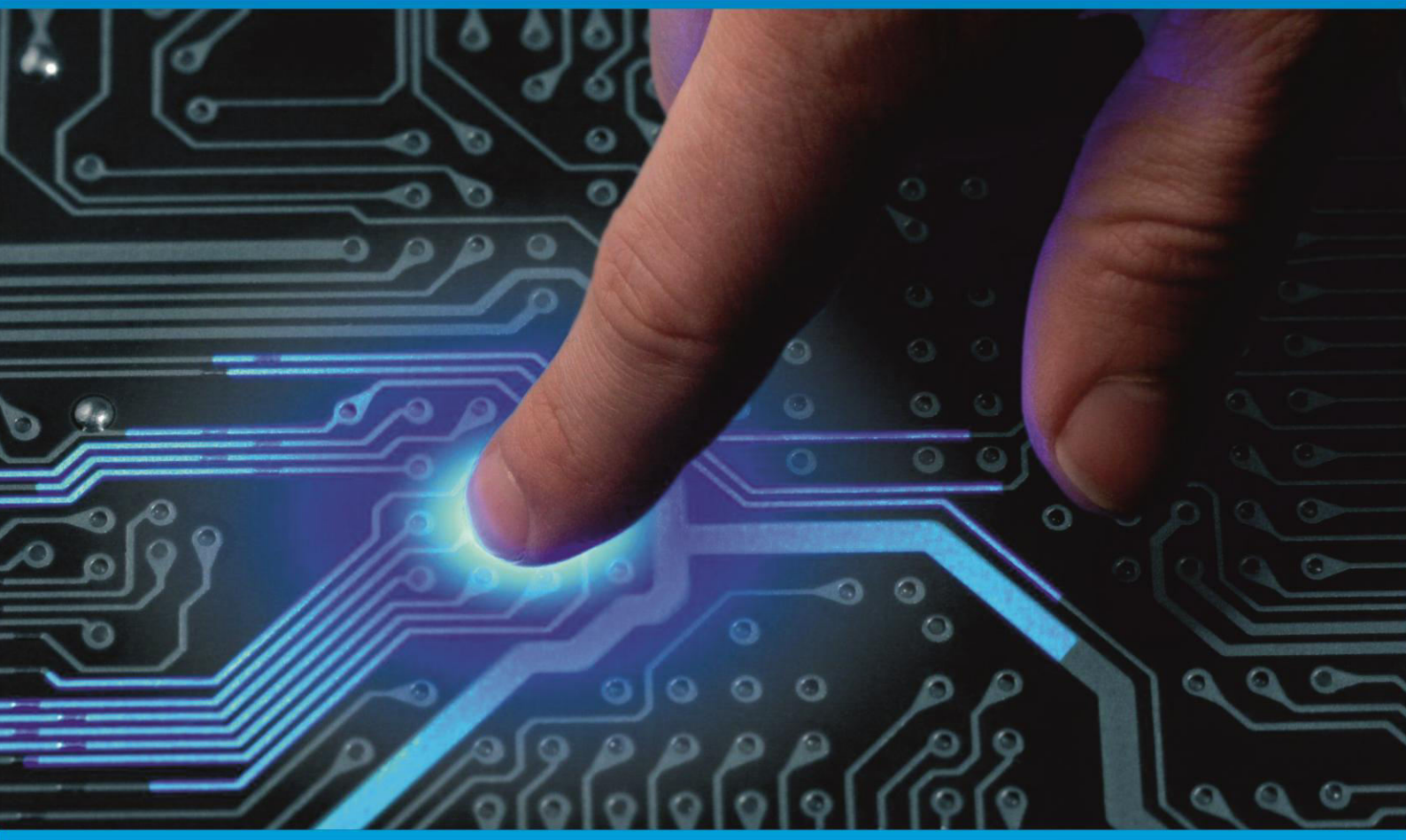
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

Mining GPS Data for User Movement Analysis

Shivakant Gupta[1], Mr.Shivam Shukla[2]

M. Tech, Goel Institute of Technology and Management, Lucknow, Dr. APJ Abdul Kalam Technical University, Lucknow, India [1]

Asst. Professor, Goel Institute of Technology and Management, Lucknow, Dr. APJ Abdul Kalam Technical University, Lucknow, India [2]

ABSTRACT: It is possible to obtain fine grained location information fairly easily using Global Positioning System (GPS) enabled devices. It becomes easy to track an individual's location and trace once trajectory using such devices. By aggregating this data and analyzing multiple users' trajectory a lot of useful information can be extracted. In our work, we aim to analyze aggregate GPS information of multiple users to mine a list of interesting locations and rank them. By interesting locations we mean the geographical locations visited by several users. It can be an office, university, historical place, a good restaurant, a shopping complex, a stadium, etc.

KEYWORDS: Data Mining, Geographical region, GPS logs.

I. INTRODUCTION

Recent advances in wireless communication and positioning devices like Global Positioning Systems (GPS) have generated significant interest in the field of analyzing and mining various patterns present in GPS data. The pervasiveness of location-acquisition technologies (GPS, GSM networks, etc.) has enabled convenient logging of location and movement histories of individuals. The increasing availability of large amounts of GPS data pertaining to the movement of users has given rise to a variety of applications and also the opportunity to discover interesting locations and travel patterns. Managing and understanding the collected location data are two important issues for these applications

A large amount of data of many users is required to mine interesting locations. A user visits limited locations in his daily travel routine. Interesting locations visited by any user during once daily schedule depends upon interest and locations situated around once residential and work place. Generally a user visits a nearby restaurant, a shopping mall, a worship place, a park, sports complex, etc. located in once own geographical region. So ideally we need data from different types of users (students, company employees, housewives) to obtain results that give us interesting insights about the city and the travel patterns of its citizens.

Large-scale geo-data analysis has become an important field of study due to the increasing volume of geo-information obtained from mobiles and GPS devices. One example of an application of geo-data analysis is to understand individuals' behaviour and movement patterns in cities. However, the massive amount of data to be analysed has lead researchers to develop computational tools and data mining techniques combined with machine learning algorithmsto enable a better management and understanding of geographic information [4]. The importance of the data mining studies for the modern society is to help individuals and companies to extract useful information from large data sets [5][11], which is manually impracticable; hence, these techniques can be likewise applied to analyse spatial data in a large scale.

II. RELATED WORK

Toncee has been a lot of prior work on using GPS data to track movement history. Toncee has also been work around integrating multiple users' information to learn patterns and understand a geographical region. Understanding a geographical region means "What are the interesting locations in that region." Toncee are many proposed methods for analyzingbehavior history using GPS data.

GeoLife is a location-based social-networking service on Microsoft Virtual Earth. GeoLife enables users to share travel experiences using GPS trajectories [12][13][14][17]. This is done by mining multiple users' location histories.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

Geolife finds the top most interesting locations, classical travel sequences in a given geospatial region. GeoLife also measures the similarity between users by understanding individual location history. In Geolife to find out interesting locations in a given geospatial region, a HITS model-Hypertext induced topic model search is introduced to find user's travel experiences and the relative interest of a location. The HITS-based model is based upon tree-based hierarchical graph (TBHG). In TBHG, clusters of stay points are created in hierarchical form. The hierarchy of the TBHG denotes different geospatial scales like a city, a district and a community. Otonce systems like Geowhiz [4] and CityVoyager [10] are designed to recommend shops and restaurants by analyzing multiple users' real-world location history. Mobile tourist guide systems [2][7][10][17] typically recommend locations and sometimes provide navigation information based on a user's real-time location. In contrast our approach determines interesting locations by applying simple relational algebra set operations combined with statistics.

Otonce methods include mining, indexing and querying historical GPS Data [9] that work under the assumption that the objects follow the same routes (approximately) over regular time intervals. This framework retrieves hidden maximum periodic patterns in GPS data. Among interactive systems mPATH: An Interactive Visualization Framework for Behavior History [15] presents a programmable analysis and visualization framework for behavior histories. In this framework users can create their own analysis method for behavior histories using data sources of behavior histories, analysis filters, and viewers. Otonce systems use inference from Partial Trajectories [5] that use the history of a driver's destinations, along with data about driving behaviors, to predict woncee a driver is going as a trip progresses. Multiperson Location Survey is used to test the destination prediction algorithm. In this multiple user data mining is used.

Otonce systems infer high-level behavior from low-level sensors [8] learning a Bayesian model of a traveler moving through an urban environment. This technique learns a traveler's current mode of transportation and his most likely route, in an unsupervised manner. This model is implemented using particle filters and it learns using Expectation-Maximization. Otonce work includes detecting user behavior based on individual location history represented by GPS trajectories [1][3][5][6]. This includes detecting significant locations of a user predicting the user's movement among these locations and recognizing user-specific activities at each location [8]. In contrast to these approaches our approach determines interesting locations by mining multiple users GPS data.

III. EXISTING SYSTEM

As shown Figure 1, in GeoLife, users can upload their GPS logs as well as associated multimedia data to the system for personal and/or public use depending on their own will. After parsing the received files, we tag multimedia content with corresponding GPS coordinates woncee they are taken. Then, based on user behavior of uploading GPS trajectories, we build spatial-temporal index over the parsed GPS data for fast retrieving GPS tracks over maps. I.e. given a spatial range over maps and/or temporal interval that a user is interested in, our system will retrieve all the GPS tracks across the spatial range and/or temporal interval. Indexing Recommendation Spatial-Temporal Index Route Search Data Pre-Process GPS Data Mining Public Knowledge Personal Knowledge GPS/ Multimedia Data GPS Log File Visualization Personal Archive Video Multimedia Files Image Audio Upload Users Web UI Mobile UI GPS Phone.

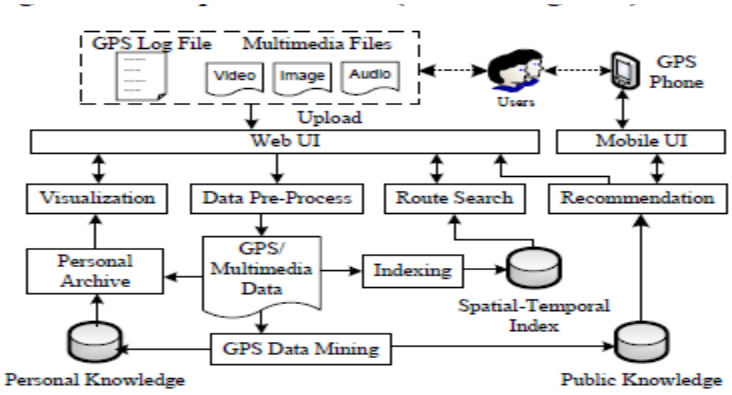


Fig 1: Architecture of GeoLife

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

IV. PROBLEM STATEMENT

For personal use, we help each user archive his/once own historical data from which we can mine a lot of knowledge such as personally transportation routines, significant places, life pattern etc. Furtonce, the knowledge is leveraged to help users summarize their own travel/sports experience and acquire healthy habit in daily life. From the public data, we learn the classic sports routes, popular travel routes, hot places and traffic condition of different routes in different time etc. The mined knowledge can be recommended to users via Web or mobile user interface (UI) when they need suggestions.

To achieve this various Classification Prediction Modelling approach are applied on the GPS trajectory data of multiple users. The end result is a ranked list of interesting locations. We show the results of applying our methods on a large real life GPS dataset of users collected over a period of two years.

V. RESEARCH OBJECTIVE

In our work presents methods to mine interesting locations from GPS data. The goal of this our work is to aggregate individual GPS traces to gain interesting insights in terms of user movement analysis. These interesting locations can be obtained by processing the data from GPS enabled devices of users living in a particular geographical region. By interesting locations we mean places that are frequented by a large group of people. Such locations could be offices, universities, historical places, museums, restaurants, parks, shopping malls, etc. Such information can be useful for city planning, traffic planning, advertising and even for suggesting interesting locations to visit, say for a tourist. In this work Classification modeling techniques to mine the GPS Data has been used.

VI. RESEARCH METHODOLOGY

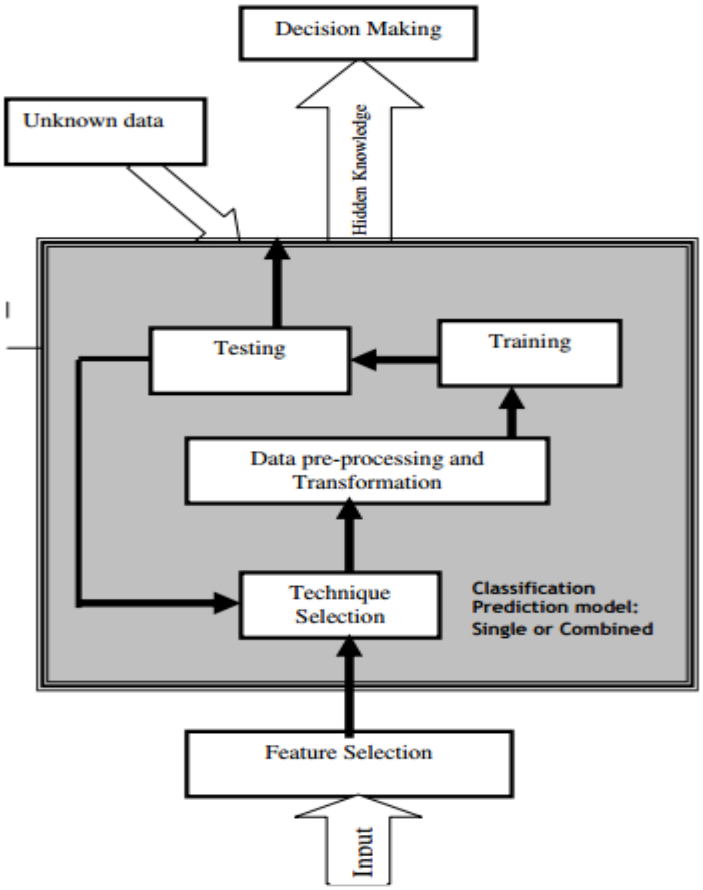


Fig 2: A simplified model showing the iterative data mining process with emphasis on the classifier building.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

This section presents the proposed technique to analyze GPS data for the development of a model using Classification approach based on

1. Naive Bayes classifier, a machine learning model that is used to discriminate different objects based on certain features. A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem.
2. Decision Tree Classifier, a simple representation for classifying examples. It is a Supervised Machine Learning where the data is continuously split according to a certain parameter.
3. Rule Induction which learns a pruned set of rules with respect to the information gain from the given ExampleSet.

6.1 Selected prediction Techniques

It deals with the concepts and principles of the techniques well-known in carrying out classification prediction.

1. Decision Tree

A decision tree is a tree like collection of nodes intended to create a decision on values affiliation to a class or an estimate of a numerical target value. Each node represents a splitting rule for one specific Attribute. For classification this rule separates values belonging to different classes, for regression it separates them in order to reduce the error in an optimal way for the selected parameter criterion.

The building of new nodes is repeated until the stopping criteria are met. A prediction for the class label Attribute is determined depending on the majority of Examples which reached this leaf during generation, while an estimation for a numerical value is obtained by averaging the values in a leaf.

This Operator can process ExampleSets containing both nominal and numerical Attributes. The label Attribute must be nominal for classification and numerical for regression.

After generation, the decision tree model can be applied to new Examples using the Apply Model Operator. Each Example follows the branches of the tree in accordance to the splitting rule until a leaf is reached.

2. Naïve Bayes (NB)

Naïve Bayes follows Bayesian theorem. This techniques deals with the analysis of attribute and class for instance so as to be able to arrive at a conditional probability in order to find the relationship between class and attribute values. The principle behind Naïve bayes for classification is a fairly simple process. During training, the likelihood of each class is calculated by counting how many times it occurs in the training dataset. This is called the "prior probability" P(C=c). This algorithm also calculates the likelihood of instance x for a given c with the hypothesis of independent attributes. This probability becomes the product of the probabilities of each single attribute. The probabilities can then be anticipated from the number of occurences of the instances in the training dataset. These numeric attributes have infinite number of values and thus it is difficult to find out the probability from these frequency distribution which ultimately leads to performance reduction of Naïve Bayes. [11].

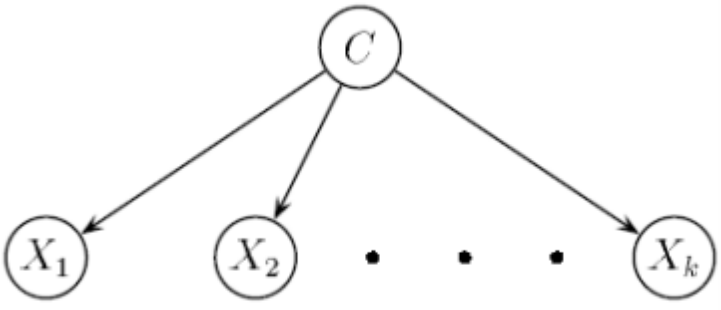


Fig 3: Naïve Bayes algorithm

Fig 3 is a NB algorithm showing how the predictive attributes X1, X2 ... Xk are conditionally independent given the class attribute C. NB is popularly known to be extremely efficient and it learns in a linear factions using ensembles principle (discussed later in this chapter). However, when attributes are redundant and not normally distributed this tends toaffect the prediction accuracy [11]. Little work has been done in terms of application of Naïve Bayes to solving data mining problems. It has however been used in conjunction with other techniques to solve classification and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

prediction tasks. However, numerous studies have compared Naïve Bayes with other machine learning (data mining) techniques. Naïve Bayes keenly competed with other techniques such as decision tree and neural networks. This may probably be due to the fact that researchers have not paid attention to the capabilities of Naïve Bayes in data mining. Nonetheless, this does not make Naïve Bayes less useful since it outperformed (in most cases) other techniques when they were compared.

3. Rule Induction

This operator learns a pruned set of rules with respect to the information gain from the given ExampleSet.

The Rule Induction operator works similar to the propositional rule learner named 'Repeated Incremental Pruning to Produce Error Reduction' (RIPPER, Cohen 1995). Starting with the less prevalent classes, the algorithm iteratively grows and prunes rules until there are no positive examples left or the error rate is greater than 50%.

In the growing phase, for each rule greedily conditions are added to the rule until it is perfect (i.e. 100% accurate). The procedure tries every possible value of each attribute and selects the condition with highest information gain.

In the prune phase, for each rule any final sequences of the antecedents is pruned with the pruning metric p/(p+n).

Rule Set learners are often compared to Decision Tree learners. Rule Sets have the advantage that they are easy to understand, representable in first order logic (easy to implement in languages like Prolog) and prior knowledge can be added to them easily. The major disadvantages of Rule Sets were that they scaled poorly with training set size and had problems with noisy data. The RIPPER algorithm (which this operator implements) pretty much overcomes these disadvantages. The major problem with Decision Trees is overfitting i.e. the model works very well on the training set but does not perform well on the validation set. Reduced Error Pruning (REP) is a technique that tries to overcome overfitting. After various improvements and enhancements over the period of time REP changed to IREP and RIPPER.

Pruning in decision trees is a technique in which leaf nodes that do not add to the discriminative power of the decision tree are removed. This is done to convert an over-specific or over-fitted tree to a more general form in order to enhance its predictive power on unseen datasets. A similar concept of pruning implies on Rule Sets.

In the first stage a pre-processing model is proposed to optimize the dataset. In the second stage experiments are performed using the machine learning classification methods to obtain the performance vector for various software fault prediction models.

VII. RESULTS

The multivariate dataset consists of 15 real attributes with 163 instances with the methodology being applied is classification.

The main attributes on whose the classification regression analysis has been carried out includes latitudes, longitudes, mode of transport, track id, speed of transport and distance covered.

The dataset consists of one includes a linguistic attribute (CAR/BUS) to indicate travel mode. If an attribute is labeled as CAR and is classified as CAR it is counted as true positive else if it is classified as no it is counted false negative. Similarly, if a label is labeled no and is classified as no it is counted as true no else if it is classified as yes it is counted as false no. Based on these outcomes a two by two confusion matrix can be drawn for a given test set. This is shown in Figure 5.8 below for model. This confusion matrix in figure 5.8 forms the basis for the calculation of the following metrics.

Accuracy = (tp+tn)/ (P+N)

Further, the datasets have been subjected to decision tree and rule induction learning processes. The screenshots of the output of the model are given below.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

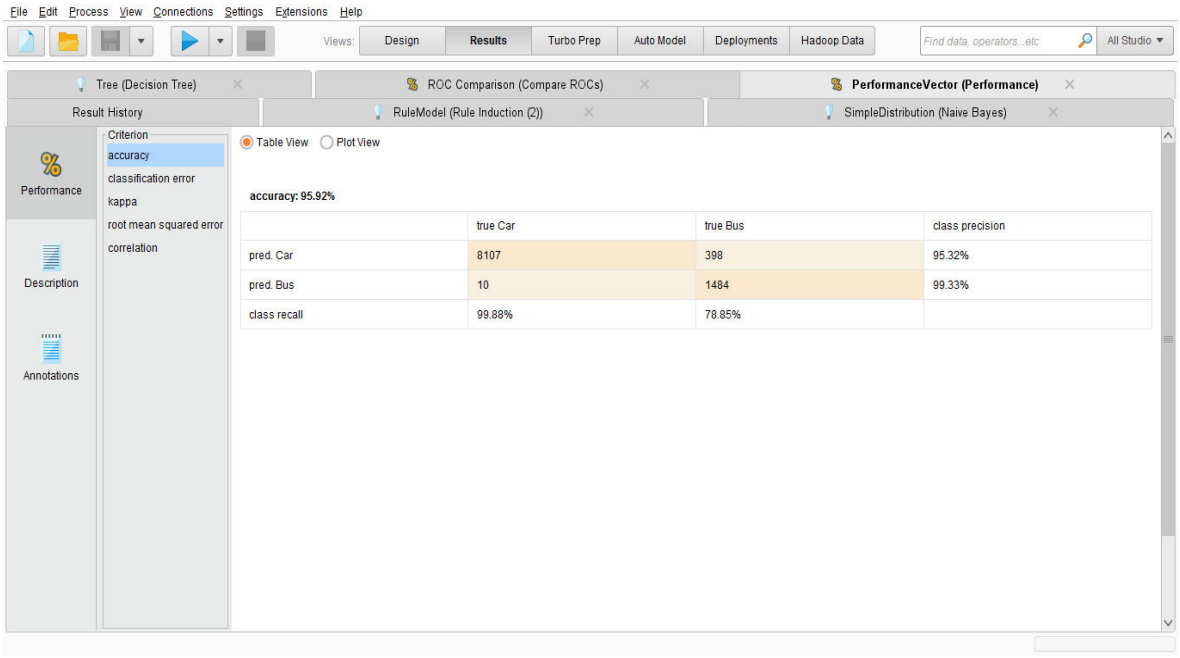


Fig.4: Confusion Matrix for MODEL with Accuracy parameter

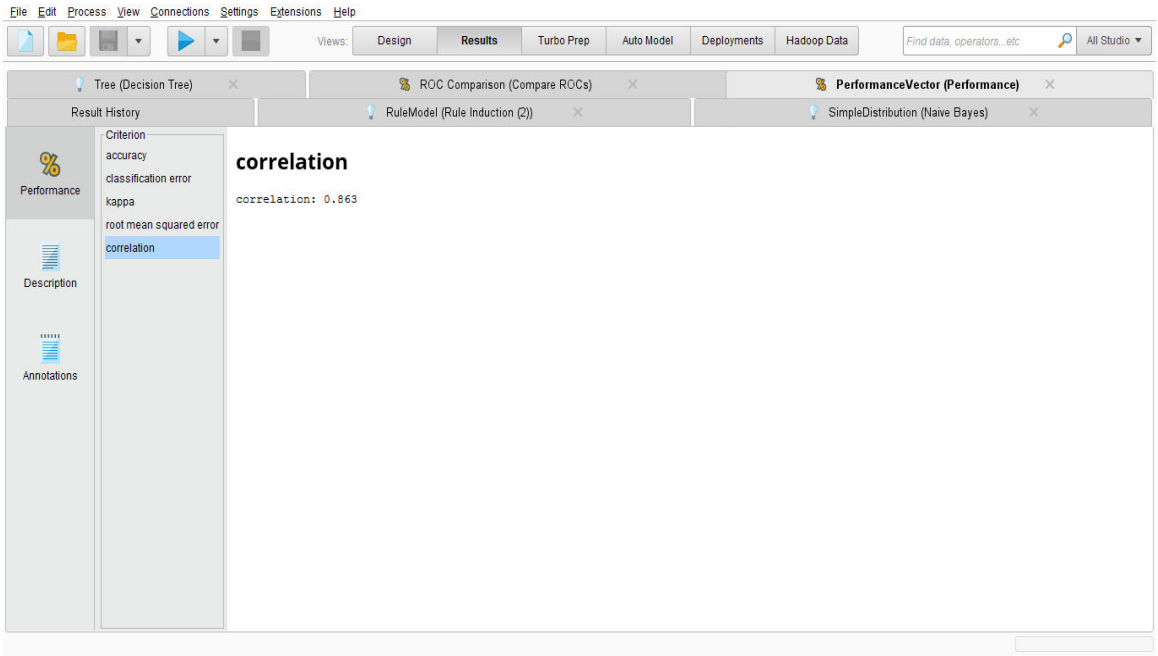


Fig. 5: Preformance correlation for MODEL

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

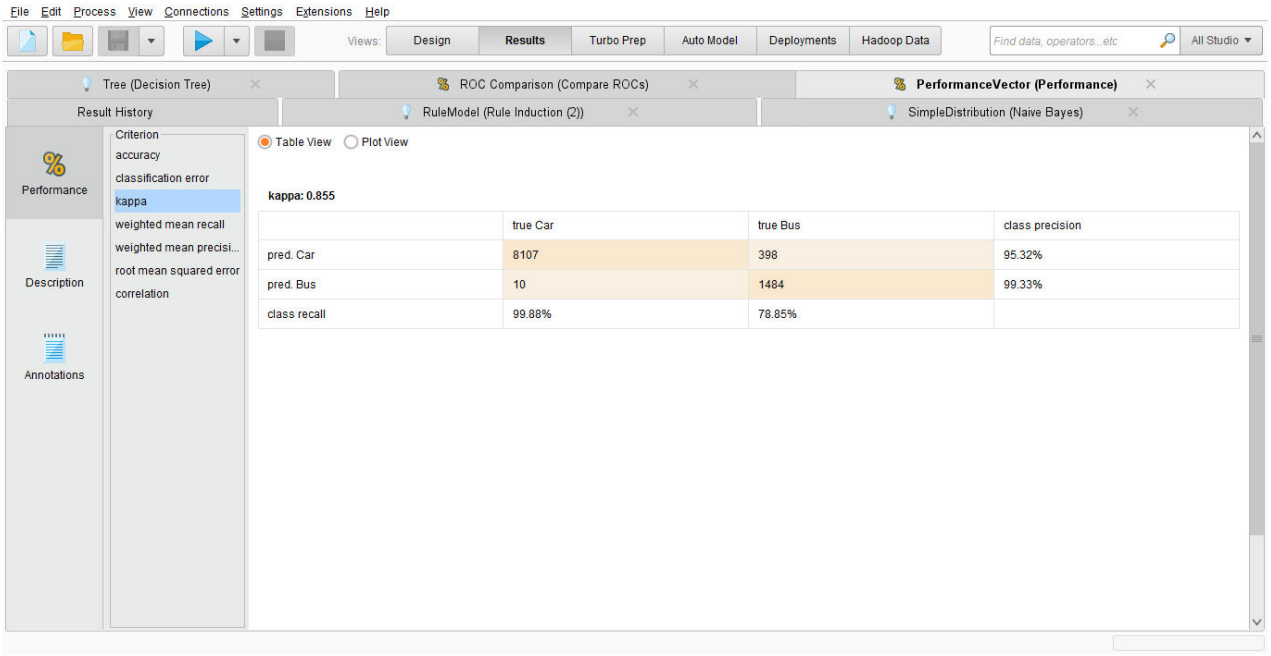


Fig. 6: Preformance Kappa for MODEL

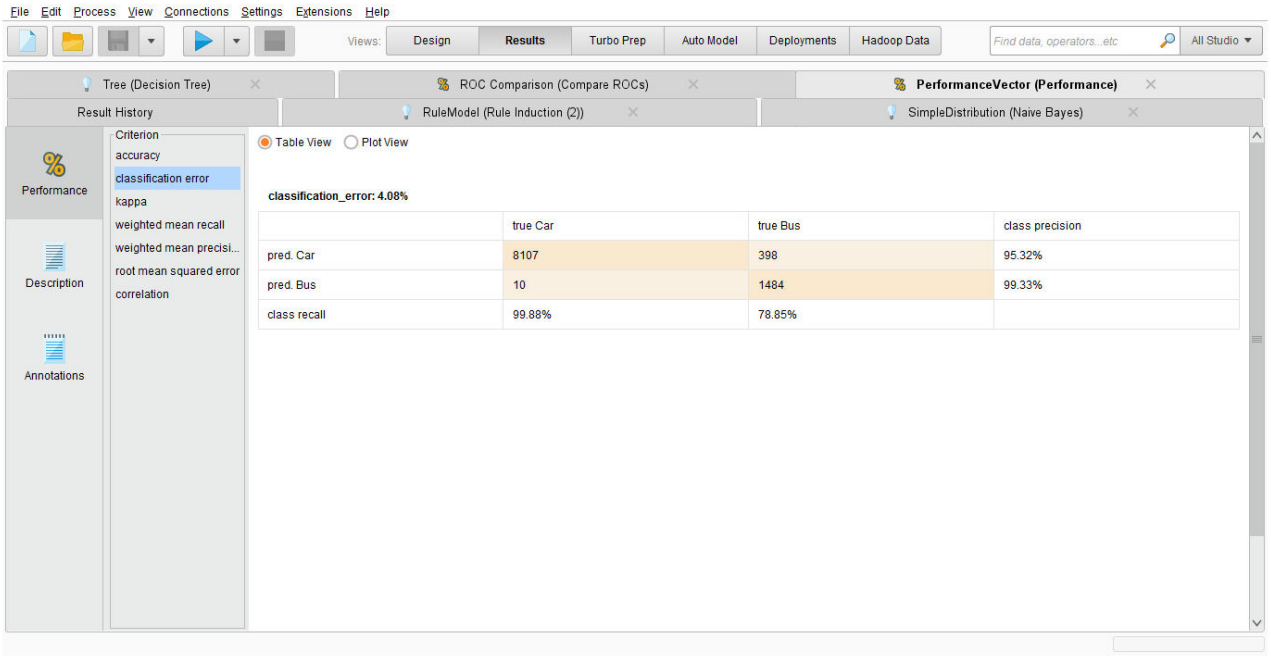


Fig. 7: Preformance Classification Error for MODEL

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

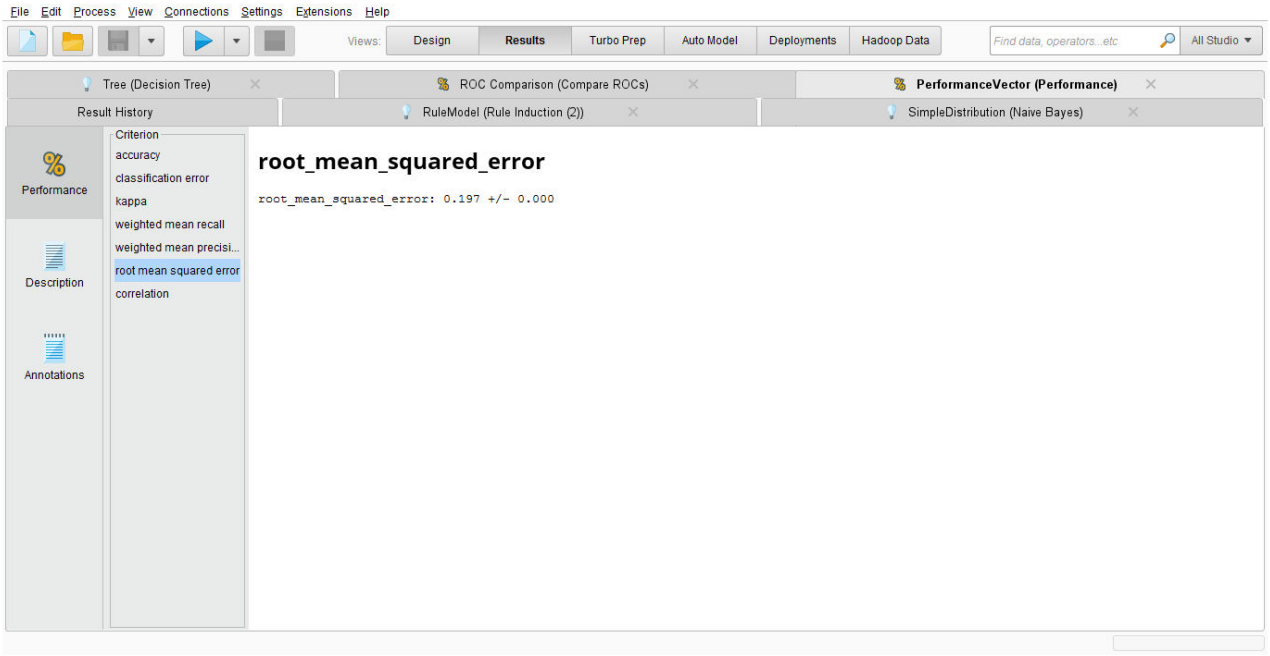


Fig. 8: Preformance RMSE for MODEL

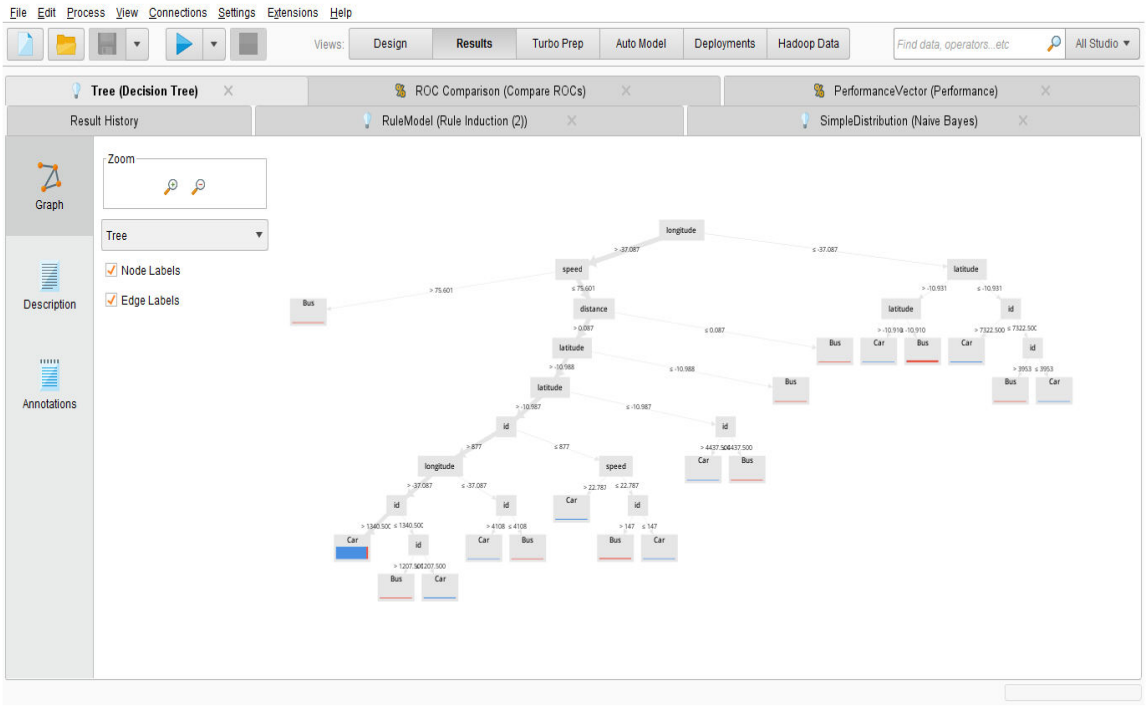


Fig. 9: Decision Tree Output for Model

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

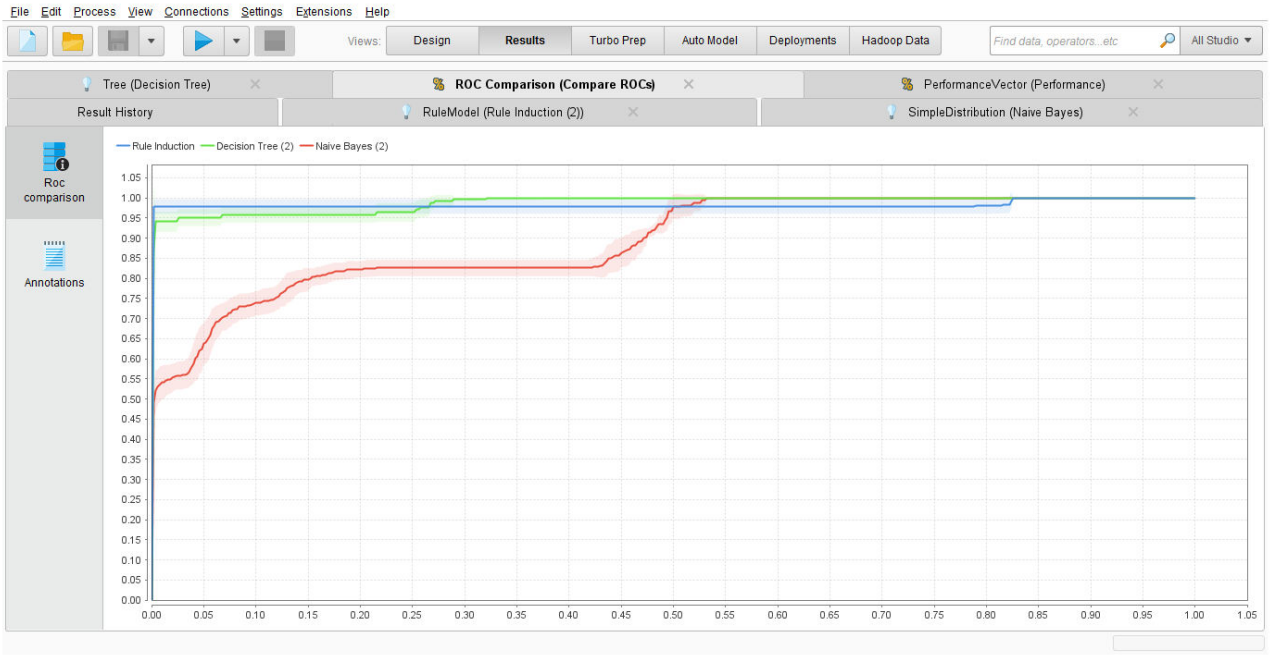


Fig 10: ROC Curve for various Learning Algorithm

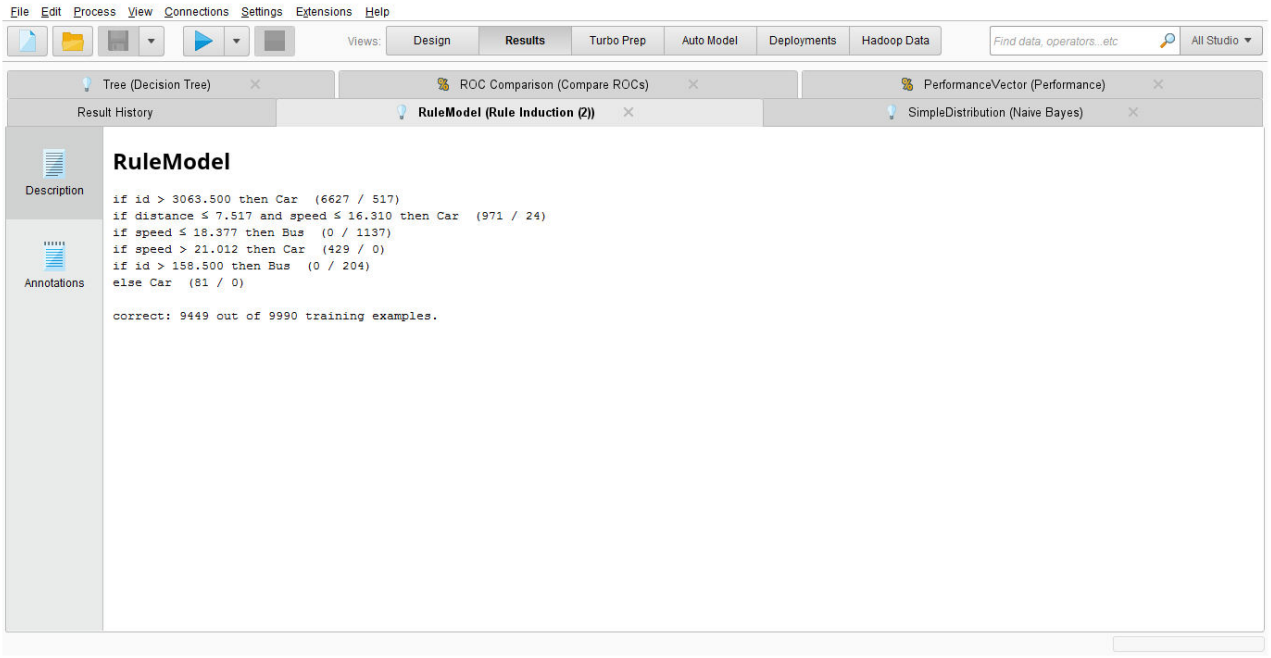


Fig. 11: Rule Induction Output for Model

VIII. CONCLUSION

Data mining through different technique turn raw data in to meaningful information. In this research data mining methods have been used to mine trajectory dataset taken from UCI Machine Learning Repository, Centre for Machine Learning and Intelligent System, which is a social network service that incorporates users, locations and GPS trajectories. The final goal of this research is prediction the mode of transportation users use based on geographic location they are. To achieve this goal considerable effort has been put to prepare the proper data set in preprocessing level. Different sizes of trajectories have been selected and each of them categorized in to stop and move data set.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905142|

Prediction the state of transportation is achieved by applying classification algorithms on move data set .Decision tree, naïve Bayesian and Rule Induction were used as classifiers and their efficiency were evaluated by precision, Correlation, Accuracy, RMSE and Kappa. Decision Tree and Rule Induction achieved more or less similar performance and Naïve Bayes was in the second place.

REFERENCES

1. Ashbrook, D., Starner, T., Using GPS to learn significant locations and predict movement across multiple users. In Proceedings of Personal and Ubiquitous Computing 7(5): 275- 286, 2013.
2. Beeharee, A., Steed, A., Exploiting real world knowledge in ubiquitous applications. In proceedings of Personal and Ubiquitous Computing 11(6): 429-437.
3. Hariharan, R., Toyama, K., Project Lachesis: Parsing and Modeling Location Histories. In Proceedings of GIScience, 2014.
4. Horozov, T., Narasimhan, N., Vasudevan, V., Using Location for Personalized POI Recommendations in Mobile Environments. In Proceedings of SAINT, 2016.
5. Krumm, J., Horvitz, E., Predestination: Inferring Destinations from Partial Trajectories. In Proceedings of the Ubicomp'103, 2013.
6. Liao, L., Patterson, D., Fox, D., Kautz, H., Building Personal Maps from GPS Data. In proceedings of IJCAI MOO05, 2015.
7. Park, M., Hong, J., Cho, S., Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices. In Proceeding of UIC, 2017.
8. Patterson, D., Liao, L., Fox, D., Kautz, H., Inferring High-Level Behavior from Low-Level Sensors. In Proceeding of Ubicomp'18, 2018.
9. Mamoulis, N., Cao, H., Kollios, G., Hadjieleftheriou, M., Tao, Y., Cheung, D., Mining, Indexing and Querying Historical Spatiotemporal Data.In Proceedings of KDD, August 2014.
10. Simon, R., Frohlich, P., A Mobile Application Framework for the Geospatial Web. In Proceedings of WWW, 2017.
11. Takeuchi, Y., Sugimoto, M., CityVoyager: An Outdoor Recommendation System Based on User Location History. In Proceedings of UIC, 2016.
12. Zheng, Y., Zhang, L., Xie, X., Wei-Ying Ma , Mining Correlation between Location Using Human Location . Inproceedings of ACM GIS '09, November 4-6, 2019.
13. Zheng, Y., Zhang, L. , Xie, X., Wei-Ying Ma., GeoLife: Managing and understanding your past life over maps. In Proceedings of MDM, 2019.
14. Zheng, Y., Zhang, L., Xie, X., Wei-Ying Ma., Understanding mobility based on GPS data. In Proceedings of Ubicomp, 2018.
15. Ito, M., Nakazawa, J., Tokuda, H., mPath: An Interactive Visualizatio Framework for Behavior History. In Proceedings of 19thInternational conference AINA, 2015.
16. GeoLife GPS Trajectories user guide, http://research.microsoft.com/en-us/downloads/b16d359d-d164- 469e-9fd4-daa38f2b2e13/default.aspx, February 15, 2017.
17. Zheng, Y., Zhang, L. , Xie, X., Wei-Ying Ma, Mining Interesting Locations and Travel Sequences from GPS Trajectories. In proceedings of WWW 2009, April 20-24, 2019, Madrid, Spain.
18. Douglas Alves Peixoto, Mining Trajectory Data, u5312727, November 1, 2013

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details