Web based Framework for Prediction of Malicious Driving Route

Gurudatta verma[a] Dr. J.P. Patra[b]

Asst. Professor, Professor

Department of Computer Science and Engineering

Shri Shankaracharya Institute of Professional Management and Technology

gurudatta.verma@gmail.com [a]; patra.jyotiprakash@gmail.com [b];

Web based Framework for Tagging of Malicious Driving Route and Machine Learning approach for Malicious Driving Route Prediction

A 32-year-old woman from Kolkata was murdered by an Ola cab driver in an attempt to rob her while she was on her way to Kempegowda International Airport (KIA) on July 31, this incident motivated us towards this research. Currently, many drivers use different kinds of navigation software to acquire better driving routes. The main function of vehicle route recommendation in the software is to find several routes between given origins and destinations by combing some path algorithms with historical traffic data, for example, Google Map and Baidu Map and then a driver could select one of those recommendation routes according to personal preference, driving distance, and current road congestion information. People usually would like to choose routes with more smooth roads. However, the above methods for driving route recommendation have some problems. In this paper we have prepared web framework to get malicious route geolocation from user, further algorithm will identify user geotagged location is authentic or not, we have applied several machine learning classifiers to predict malicious driving route, ensemble classifier produces higher accuracy.

Keywords: TPR, FNR, F-Score, Accuracy, SVM, DTC, NB, HITL

Introduction

Knowledge of a driver's destination is an important parameter for delivering useful information during the drive. However, prediction of driving route, traffic jam deification etc. are issues but one incident happed in Bengaluru, India that A 32-year-old woman from Kolkata was murdered by an Ola cab driver in an attempt to rob her while she was on her way to Kempegowda International Airport (KIA) on July 31. The suspect was traced by Bengaluru police through a meticulous investigation. Police learnt that Pooja had booked an Ola cab online on July 30 to reach her hotel. After reaching the hotel she called up the same driver and asked him to pick her up early next morning to reach the airport as there would be less cab availability during the wee hours. She then boarded the cab from Hosur Road at 4.15 am on July 31. The cabbie who drove her that day was HM

Nagesh (22), a resident of Hegganahalli Cross in Bengaluru, this incident motivated us towards us towards this research, in which we will gather data through our web framework and will apply machine learning classifier for prediction of malicious route, once malicious route identified user will be notified for the same.

Apply machine learning over input data depend on the nature of data. Gathered Data has following two categories depicted in fig.-1.

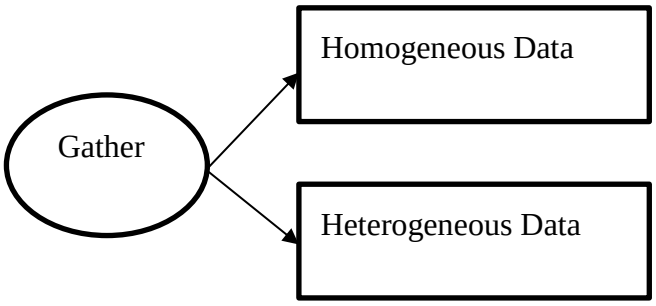

Figure-1 Data Category

Various learning algorithm comes under machine learning algorithm, Machine learning is the subset of AI (Artificial Intelligence), fig-2 depicts the same.

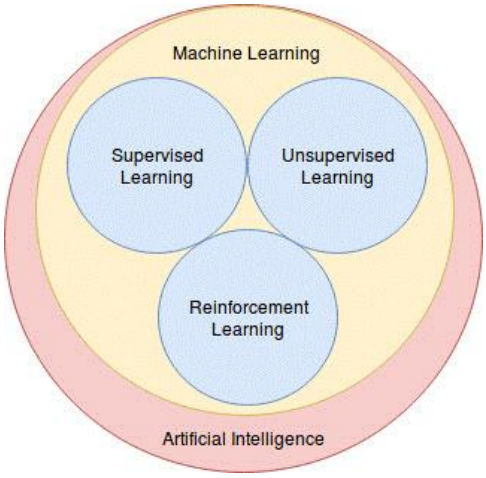

Figure-2 Learning Methods and AI

Before applying machine learning over gathered data need to know about some basic terminologies like Training data, Test data, Predictor, Response etc.. Training data is the historical data or gathered data which contains matrix formatted data, over which

machine learning algorithm like Naïve Bayes, Decision Tree etc. applied, test data is the input data for which we need to predict the result, training data is the collection of variables, as we can see in fig.-3 blue colored column is response or dependent variable i.e. variable which is meant for prediction and gray colored columns are independent variable or predictor variable. Applying machine learning over data set consist of 3 steps.

1. Pre-processing

2. Attribute Selection

3. Machine Learning Classification

As we saw gather data is in two different nature as depicted in fig.-1, Homogeneous Data mean, if the values of dependent variable can categories for example, wanted to predict route as malicious / Not malicious so here value of dependent variable varies in two classes.

Table-1 Dataset Fragment

id	place	longitude	latitude	class
1	jaistambh chowk raipur	21.2376284	81.5962147	1
2	sejbahar	21.1948	81.6497	1
3	Datrenga	21.1583	81.6416	1
4	kolar	21.1583	81.6616	1
5	khorpa	21.116	81.6672	1
6	khilora	21.1581	81.6635	1
7	chachanpari	21.1159	81.6668	1

There are some basic aspects of machine learning algorithm that need to identify how dependent variable changes according to independent variable, let us recall the geometry portion again equation of line:

$y=mx+c$ ……..(1)

Here m and c are the parameters and y is the dependent variable and x is the independent variable. We can easily understand above by following fig.-3.

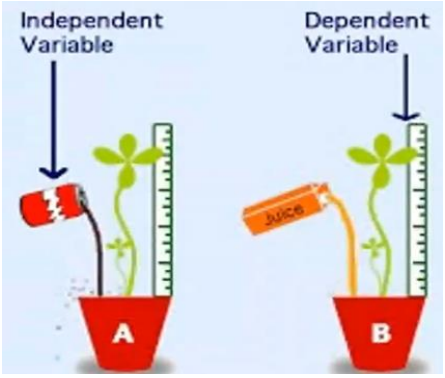

Figure-3 Analogy of y=mx+c

Literature Survey

Jon Froehlich 2018 , showed how the regularity of a driver's traveling behavior could be exploited to predict the end to-end route for their current trip. Author made three primary contributions. First, we provided a methodology for automatically extracting routes from raw GPS data without knowledge of the underlying road structure.

Daniel Ashbrook et. al. on 2004 said that Wearable computers have the potential to act as intelligent agents in everyday life and to assist the user in a variety of tasks, using context to determine how to act. Location is the most common form of context used by these agents to determine the user's task. However, another potential use of location context is the creation of a predictive model of the user's future movements. We present a system that automatically clusters GPS data taken over an extended period of time into meaningful locations at multiple scales. These locations are then incorporated into a Markov model that can be consulted for use with a variety of applications in both single-user and collaborative scenarios.

ognitive Assistance at. al. on 2004 presented an automated transportation routing system, called "Opportunity Knocks," whose goal is to improve the efficiency, safety and independence of individuals with mild cognitive disabilities. Our system is implemented on a combination of a Bluetooth sensor beacon that broadcasts GPS data, a GPRS-enabled

cell-phone, and remote activity inference software. The system uses a novel inference engine that does not require users to explicitly provide information about the start or ending points of their journeys; instead this information is learned from users' past behavior. Futhermore, author demonstrated how route errors can be detected and how the system helps to correct the errors with real-time transit information. In addition author present a novel solution to the problem of labeling positions with place name

Methodology

It is necessary to train the prediction model from drivers' past history. In particular, the larger the size of training examples is, the more accurate the prediction model for path predictions is. In view of the limitation of given training examples, the training set cannot contain all of routes that drivers will take in the future. So the we have proposed to use Geotagging by users, in our proposed prediction model we will use ensemble as Machine Learning Classifier.

Ensemble Machine Learning Model: In this machine learning paradigm number of classification model (Weal Learners) trained to solve the problem. Hypothesis behind ensemble model is when weal learners are correctly integrated then we can attain accurate prediction. There are major three kinds of approaches to combine weak learners as follows:

1. Bagging
2. Boosting
3. Stacking

In our proposed prediction model we have used Boosting for combining weak learners. So let us describe how Boosting works. Boosting uses weighted average to make weak learner to strong learner, boosting can be easily described bu fig.-3.

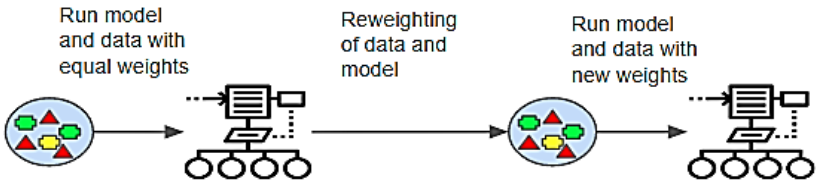

Figure-3 Boosting

Ensemble

Step-1. Take any learner as base learner that keeps all the distributions and allocate identical weight or attention to every observation.

Step-2. If prediction error occurs by 1st base learning algorithm, then put higher weight to sample having prediction error. Then, we put on the next base learning algorithm.

Step-3. Repeat Step 2 till the limit of base learning algorithm is reached or higher accuracy is attained.

There are different types of boosting algorithms, opted based on nature of response variable of dataset, if the response variable having binary class then we have to go for following boosting algorithm as:

- Ada BoostM1
- Logit Boost
- Gentle Boost
- Robust Boost
- LP Boost
- RUS Boost
- Total Boost

For more than 2 class:

- Ada BoostM2

- LP Boost
- RUS Boost
- Total Boost

In our dataset there is two classes hence we have applied binary classification boosting algorithm. 100 classifier combined to make strong learner. In this paper we have used hold out method to validate the prediction model. In this method we partition the dataset to training set and test dataset as per the given input. Fig.-4 shows the how data portioned into test set and training set.

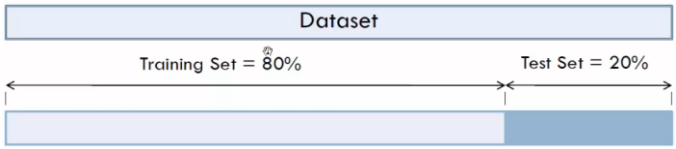

Figure-4. Data partition in Hold out

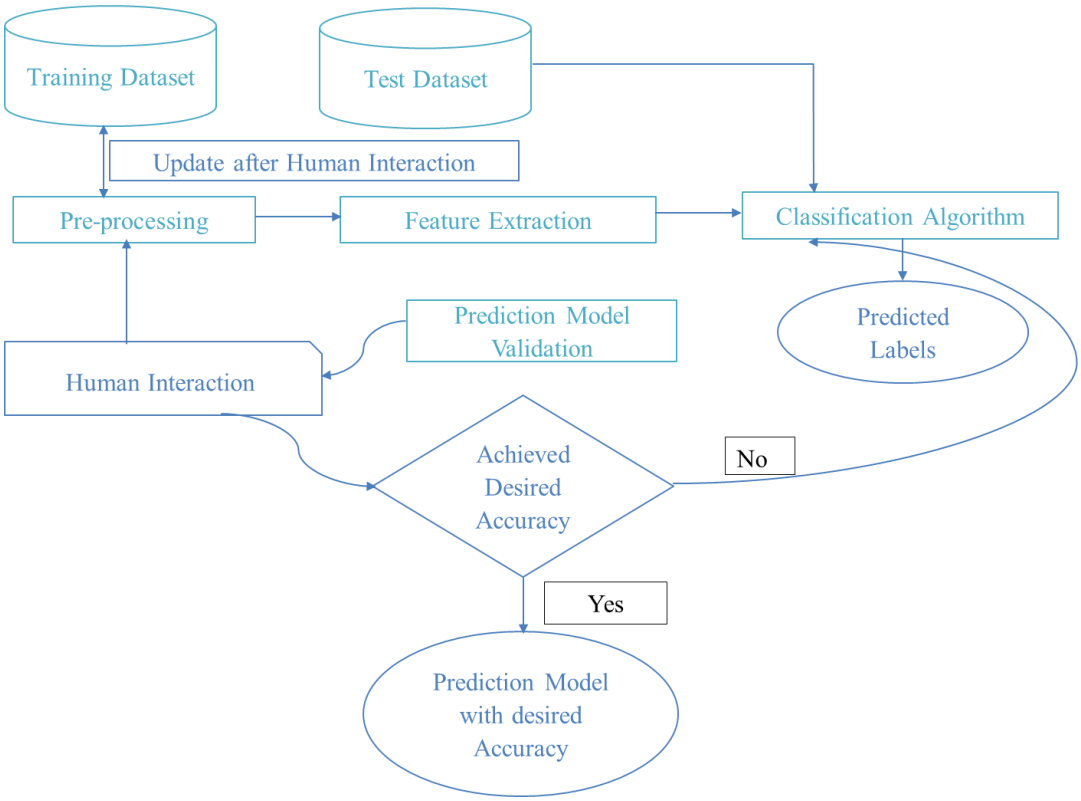

Figure-5 Flow for Prediction Using HITL (Human in the Loop)

Validation of classifier can be done by evaluating following parameters using validation phase generated confusion matrix values:

- True positive (TP): correct positive prediction
- False positive (FP): incorrect positive prediction
- True negative (TN): correct negative prediction
- False negative (FN): incorrect negative prediction

Accuracy = (TP + TN) / (TP + TN + FN + FP)

Sensitivity= TP/ (TP+FN)

Experiment

For web framework we have used PHP 7.0 Xampp server with leaflet , and for algorithm implementation Matlab 2019a used.

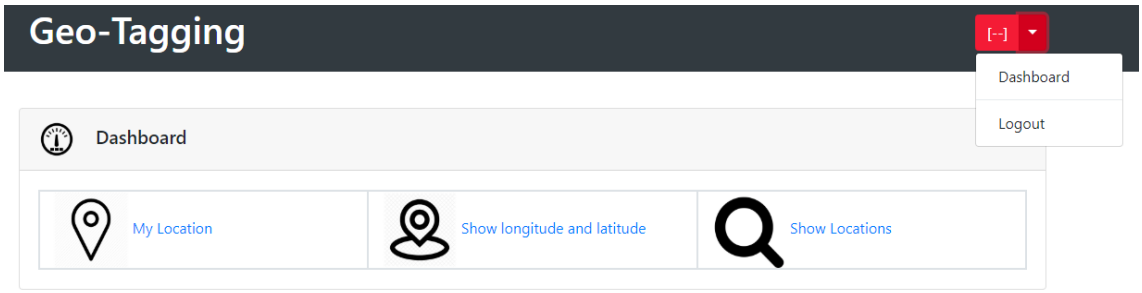

Figure-6 Dashboard Web Framework

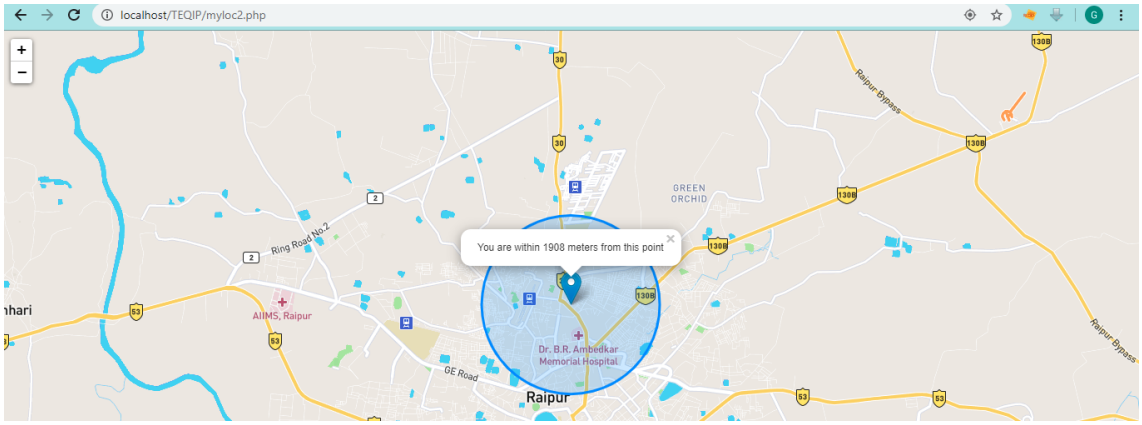

Figure-7 If user click on My Location

Fig.6 shows the dashboard for user, fig.-7 show the user's own location, fig. 8 shows the graphical and numerical longitude and latitude of user's location for saving the location

for historical data, fig.-9 shows the user interface for putting any saved location as vulnerable.

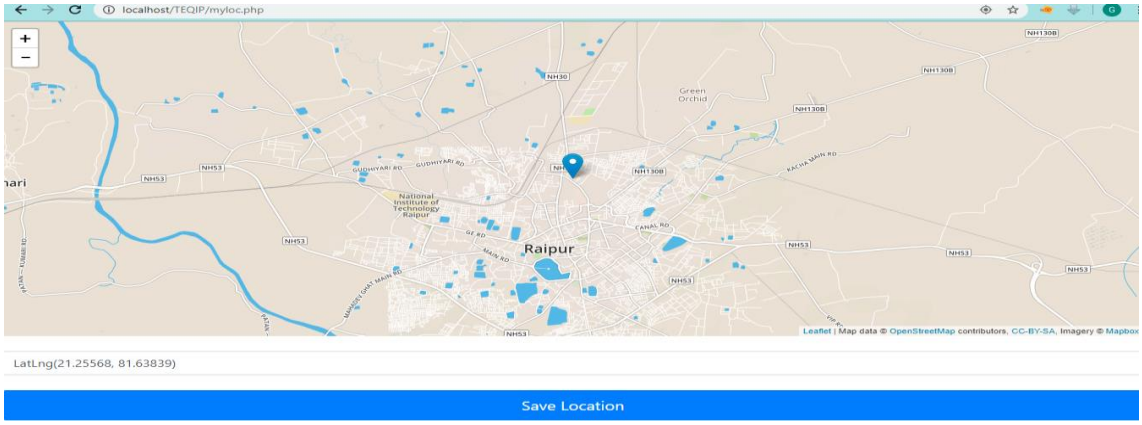

Figure-8 Save Location Visited

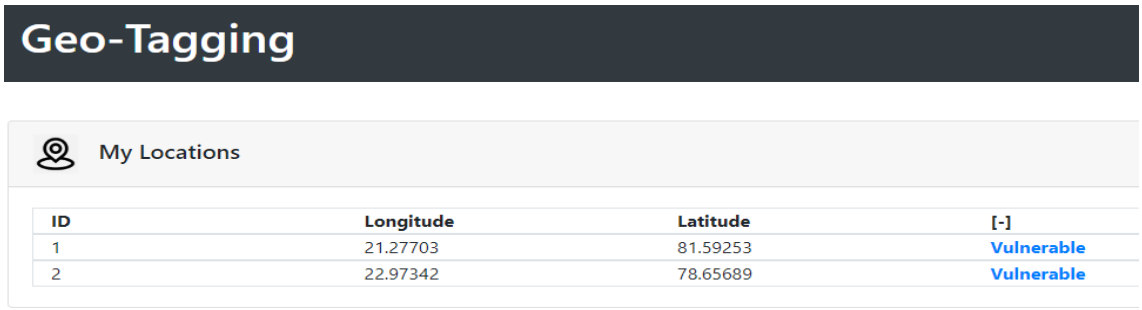

Figure-9 Tagging for Vulnerable Location

After gathering historical data, we have applied several machine learning classifiers.

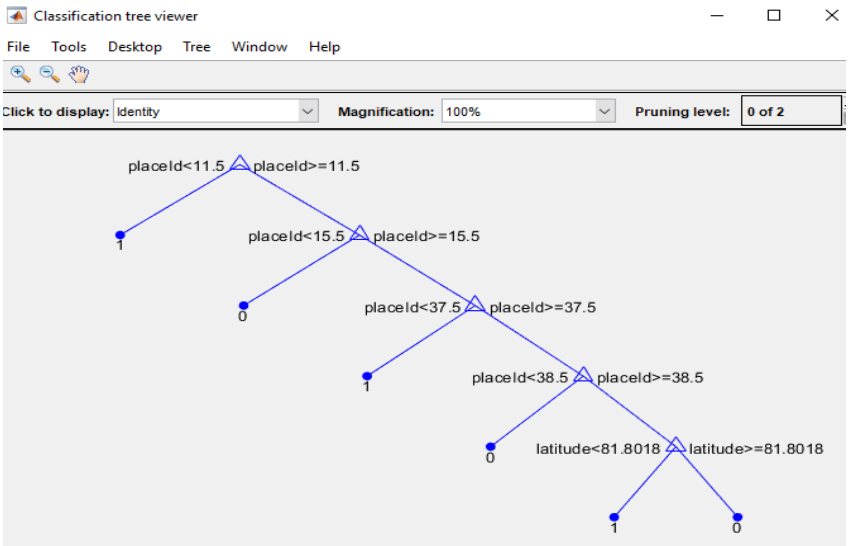

Figure-10 Decision Tree

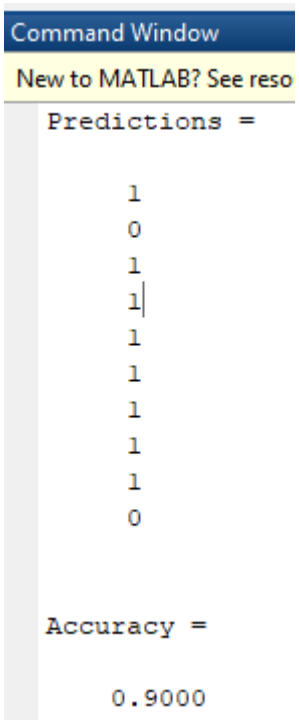

Figure-10 Accuracy of Decision Tree Classifier

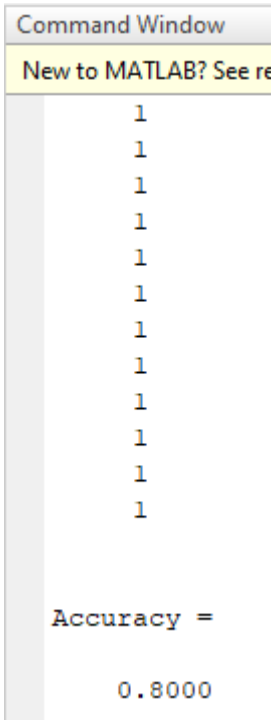

Figure-11 Accuracy of Naïve Bayes Classifier

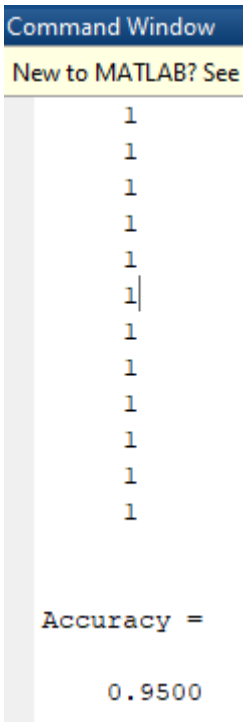

Figure-12 Accuracy of Ensemble Classifier

Table-2 Comparison

S. No.	**Classifier**	**Accuracy**
1.	Naïve Bayes (NB)	0.80
2.	Decision Tree (DTC)	0.90
3.	Discriminant Analysis	0.92
4.	Support Vector Machine (SVM)	0.87
5.	Ensemble Classifier	0.95

Conclusion

We have proposed to use Geotagging by users, in our proposed prediction model, this approach is called HITL system that is Human in the Loop system, to improve the accuracy of system we need to check the result with help of users so as to improve machine learning classifier accuracy, we have achieved 95% accuracy as we have

gathered less amount of training data. In future we can apply deep learning classifier and can use AWS (Amazon Web Service) server for deployment of ML model.

References

John Krumm, Real Time Destination Prediction Based On Efficient Routes, Intelligent Vehicles & Transportation Systems (SP-2027) 2018

Deguchi, Y., et al., Hev Charge/Discharge Control System Based on Navigation Information, in SAE Convergence International Congress & Exposition On Transportation Electronics. 2004: Detroit, Michigan USA.

Tate, E.D. and S.P. Boyd, Finding Ultimate Limits of Performance for Hybrid Electric Vehicles. 2001 SAE Transactions - Journal of Passenger Car - Mechanical Systems, 2001.

Karbassi, A. and M. Barth, Vehicle Route Prediction and Time of Arrival Estimation Techniques for Improved Transportation System Management, in Intelligent Vehicles Symposium. 2003. p. 511-516.

Torkkola, K., et al., Traffic Advisories Based on Route Prediction, in Workshop on Mobile Interaction with the Real World (MIRW 2007). 2007: Singapore.

Simmons, R., et al., Learning to Predict Driver Route and Destination Intent, in 2006 IEEE Intelligent Transportation Systems Conference. 2006: Toronto, Canada. p. 127-132