

Crime Analysis: Safe Location and Route

Ravi Sundar K.M

cb.en.u4cse16654@cb.students.amrita.edu

Sai Sruthi V

cb.en.u4cse16640@cb.students.amrita.edu

Akhil S

cb.en.u4cse16601@cb.students.amrita.edu

Abirami K

k_abirami@cb.amrita.edu

Dept of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham
India

Abstract— Safety is an integral part of everyone's life. A model is developed to suggest a safe route. Machine learning techniques are used to predict the routes. An on-screen navigation system which helps to let know the nearest police stations and hospitals during the transit.

Keywords—Crime Rate Navigation, Machine Learning, Android Application, API, Web Application, pickle, flask

INTRODUCTION

During recent times, ensuring safety while commutation has become much more crucial. With technological innovations, we have managed to provide solution for many critical problems that prevail in our society. Despite this, there has not been a concrete method to resolve the need of safety during travel. In the past, various statistical models have been used to analyze crime data sets to gain insights about the pattern and the nature of crime. They are used to predict the crime rates of a particular geographic location in the upcoming years. But unfortunately, these predictions do not actually help in reducing the crime rate neither do they provide safety to a person. Therefore, our project tries to address this issue by utilizing the crime rate data set to help people choose the safest possible route that would ensure a hassle free and secure journey. Though there are certain applications that try to answer this, they end up providing alert messages during times of emergencies instead of aiming to prevent such situations. We aim to design an effective solution that targets to provide a safe journey by presenting the users with the best possible routes of travel based on certain factors like past crime rates, the type of crime and the significance of it so that it diminishes the risk of any toward events. Data set is obtained from Kaggle and it covers Chicago's geographic location. It comprises of more than 100000 entries of records with details regarding the number of crimes, the nature of crimes, the block/area where the crime, its latitude and longitude and a few more. The data is processed by handling missing values and label encoding is used to assign a significance score to the type of crime. We have implemented an android application where the user can input the source and the destination to be reached. It notifies the nearest police stations and hospitals during emergencies without having to cancel the navigation route. The task is to determine the probability safety score from the specified place. Moreover, an area's safety is also determined by the availability of police stations near by for emergencies. Additionally, there is a panic button which when pressed will make a call to the women's helpline.

LITERATURE SURVEY

The data set is obtained from Kaggle and it covers Chicago's geographic location. It comprises of more than 1,00,000 entries of records with details regarding the number of crimes, the nature of crimes, the block/area where the crime, its latitude and longitude and a few more. The data is processed by handling missing values and label encoding is used to assign a significance score to the type of crime. We have implemented an android application where the user can input the source and the destination to be reached. It notifies the nearest police stations and hospitals during emergencies without having to cancel the navigation route. The task is to determine the safe routes among all the possible routes from the specified place. Moreover, an area's safety is also determined by the availability of police stations nearby for emergencies. Additionally, there is a panic button which when pressed will make a call to the women's helpline.

In another research paper [2], crime rates of previous years have been used to predict the crime rate in the coming years. For this purpose, statistical models like the WMA, i.e., weighted moving average model, FCR, i.e., Functional Coefficient Regression model and AGP, i.e., Arithmetic Geometric Progression model have been incorporated.

Weighted Moving Average model: It takes the average of values with weights in descending order for predicting. It adds weights to each value in a descending way. Then the sum of the obtained series is divided by the weights' sum to obtain the prediction.

FCR model: In this, the weights to be assigned for a value is calculated using functions. In this paper, the researchers have utilized two functions, which are x^2 and $1/x$. But the utilized functions should neither linearly increase nor decrease since the crime rate will not have any specific pattern.

AGP : This model is the one in which the terms of an AG progression are considered the weights and assigned to the value of previous years for getting the prediction for the necessary year.

EXISTING SYSTEM

These are systems that focus on determining the trends in crime rate in future for any given area using statistical models. WMA ,i.e., weighted moving average model, FCR, i.e., Functional Coefficient Regression model and AGP ,i.e., Arithmetic Geometric Progression model are some of the statistical approaches to denote weights. These are implemented over the crime data set for prediction. There are mathematical and technical simulations to discover about crime and the related factors concerning it. It is used to compare the previous years' crime pattern and checks if it has any effect in the future years. These kinds of models can find habituated criminals and not-so-frequent law breakers. It also uses statistics and Artificial neural network to know about the nature of law breakers. But as we know, just predicting the crime rate or behavioral trend without utilizing it for safety purposes is not a good solution.

Other solutions include triggering mechanisms that send text messages, e-mails or tweets to the friends of users in android application on the event of pressing a button or shaking the phone during emergencies. GPS enabled tracking system to monitor the movement of user while the person starts journey and constantly informing the whereabouts of him/her to the specified list of people is also seen in many applications. Wearable devices that trace only the users' location instead of providing the safest way of navigation are prevalent in the recent times. There are also applications order vices that emit loud shrills of cry when an emergency button is invoked to capture people's attention instead of reducing the chances of emergencies. So, there are plenty of systems available for monitoring a person but there is no efficient system that helps to decide the safest route of travel before starting a journey,there by averting any possible chances of crime instead of taking action after a crime has occurred. Therefore, our application strive stop provide this smart decision which would be more useful when compared to the existing system.

PROPOSED SYSTEM

A. Present Work

we collected Chicago crime data set from Kaggle. The data is preprocessed using smooth algorithm. Smooth algorithm is a digital filter, that helps to identify the significant patterns in the data. It also helps to extract more information from the data set. We have implemented different supervised machine learning algorithm such as Logistic regression, Gaussian Naive Bayes, Random Forest Classifier, K Nearest Neighbors, XGBoost and Decision tree to predict the safety area. All the above said algorithms are compared and based on the accuracy, precision and recall value. Random Forest classifier fits perfectly for our data set.

B. Implementation:

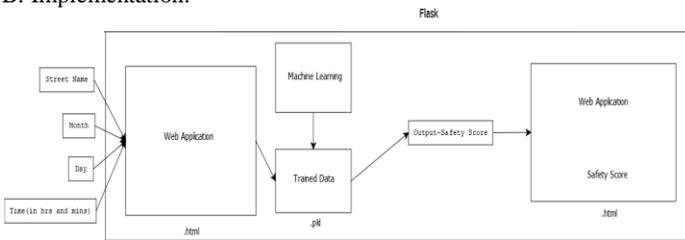


Fig:1 application block diagram

- An android app reads source and destination place. Refer fig 1. All possible routes along with latitude and longitude from source to destination are generated using google maps API.
- Since crime vary in time, date and month refer fig 2, 3, 4. Our application will collect current date, time, month along with the generated latitude and longitude as an input to determine probability score.
- The data will undergo machine learning techniques to find the best probability score
- If the score is below 40% the route will be marked in red color which indicates that the route will not assure safety for transit.
- If the score is between 40% to 60% it will be marked as blue which indicate the route ensure medium safety.
- Above 60% it will be marked as green which indicates the route will be safety for transit.
- An average value of the probability score will be displayed to the user.

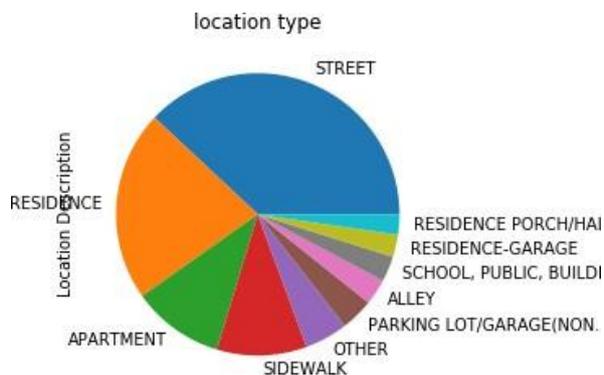


Fig 2: Top 10 location with most crime

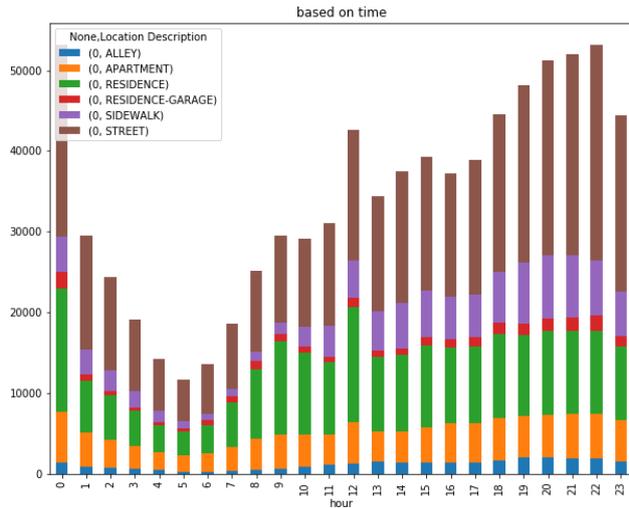


Fig:3 Location description based on time

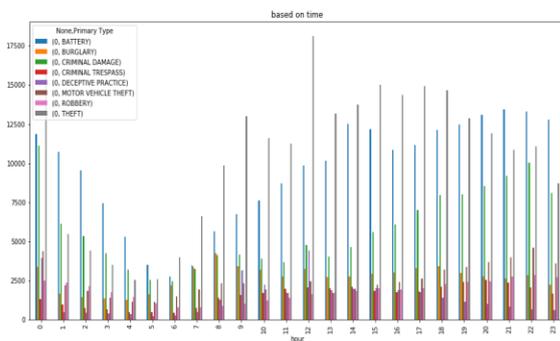


Fig 4:Crime based on time

C. ALGORITHM:

1. Logistic regression:

It is a classification algorithm that is used predict the class in which the data would fall. It is a supervised algorithm.

The features used for prediction is latitude and longitude of crime, the date at which the crime took place and the time of crime. During a crime, whether prompt action is taken or not is used to classify which helps to know if that place is safe or not. After classification, the probability of safety is also found for the surrounding region.

The accuracy thus obtained through logistic regression is 74.2.

2. Gaussian Naive Bayes

It is a supervised classification algorithm that is based on Bayes Theorem. The condition for this algorithm is that the features of data should be independent of each other.

Since our data has independent features, Naive Bayes algorithm was tried and tested. The accuracy reduced by 0.001.

Accuracy of Naive Bayes: 74.10

3. Random Forest Classifier:

It is used for both classification and regression problems.

It essentially builds a set of decision trees for the sample of data. One advantage is that it does not suffer from the problem of over fitting as decision tree does. When we implemented decision tree, the accuracy obtained was 99.98 which is extremely over fitted. Random forest solves this problem.

Accuracy of Random Forest Classifier: 99.98

4. K Nearest Neighbors:

It is yet another classification and regression algorithm that is supervised. This considers that similar data occur near to each that. K is the number

of nearest points to be taken for prediction. If majority of the K points belong to a particular class, then it predicts the new point to that class.

Accuracy: 74.4

5. XGBoost:

It is a gradient boosting algorithm that uses decision trees. It is used to maximize the performance and accuracy. It is an extreme gradient boosting algorithm that reduces the errors by gradient descent. It performs very well for unstructured data too. We secured the highest accuracy through this. Thus this is the chosen algorithm for implementation.

Accuracy: 76.1

Chosen Algorithm: XGBoost

RESULT:

Thus, the model can ensure safety. This is an on-screen device to classify the routes to be safe or unsafe during transit. Refer figure 6,7,8. The prediction is made better by implementing different supervised machine learning algorithms. The results of the algorithms are compared to find the best probability score refer fig5. This highest score enhances more security. Our model unfolds nearby police station and hospitals refer figure 8.

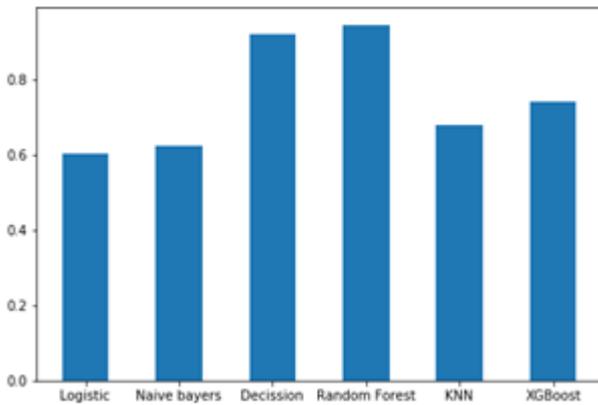


fig:5 different machine learning

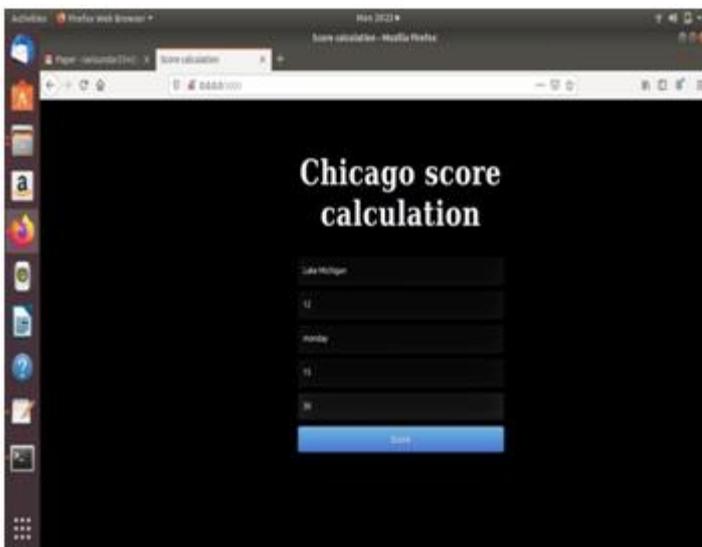


Fig 6:web application user interface.



Fig 7:probability safety score for Chicago

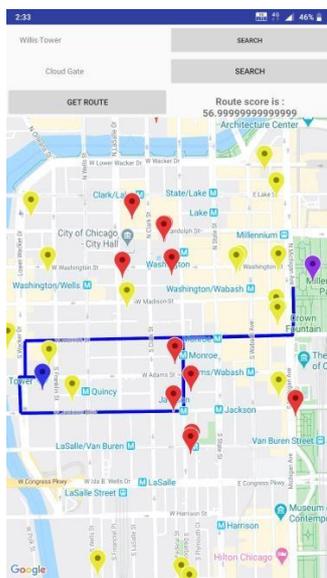


Fig 8:safety route with safety score

CONCLUSION

Thus, the model can provide an on-screen emergency solution while navigation. It is also being extended to classify the locations to be safe or unsafe using the data set provided. The prediction can be made better by trying to implement algorithms that perform better. Additional features like the streetlight, road conditions etc., could be used to give more reliable prediction. Real time inputs such as weather conditions around and issues that occur locally can also be used in conjugation to make the application powerful. User feedback for an area can also be used to determine a safety score that will do a check on real time

References

1. Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav, "Crime Pattern Detection, Analysis and Prediction" IEEE2017.
2. Shubham Agarwal, Lavish Yadav, and ManishK Thakur, "Crime Prediction based on Statistical Models" IEEE 2018.
3. Chhaya Chauhan, Smriti Sehgal, "A Review: Crime Analysis using Data Mining Techniques and Algorithms" IEEE 2017.
4. Naz Mahmud, Khalid Ibn Zinnah, YeasinAr Rahman, Nasim Ahmed, "CRIME-CAST: A Crime Prediction and Strategy Direction Service" IEEE2016.
5. Julio Borges, Daniel Ziehr, Michael Beigl, TECO, Institute of Telematics Karlsruhe Institute of Technology (KIT), Nelio Cacho, Allan Martins Federal University of Rio Grande do Norte, UFRN Natal, Brazil and Simon Sudrich, Samuel Abt, PatrickFrey, Timo Knapp, Michaela EtterandJohannes Popp Karlsruhe Institute of Technology (KIT) Karlsruhe, Germany, "Feature Engineering for Crime Hotspot Detection" IEEE2017.
6. abirami K, harini N, vaidhyesh P.S, kumarp, Analysis of web workload on qos to assist capacity. "Lecture Notes in Computational Vision and Bio-mechanics."
7. Sabarish B.A, Karthi R, Gireshkumar T, A survey of location prediction using trajectory mining.

