Safety App: Crime Prediction Using GIS

Atharva Deshmukh
Computer Engineering
Don Bosco Institute of Technology
Mumbai, India
atharva2six@gmail.com

Sourab Banka
Computer Engineering
Don Bosco Institute of Technology
Mumbai, India
sourabbanka22@gmail.com

Sean Bruno Dcruz
Computer Engineering
Don Bosco Institute of Technology
Mumbai, India
seandcruz@gmail.com

Sana Shaikh
Computer Engineering
Don Bosco Institute of Technology
Mumbai, India
sana.siraj.shaikh@gmail.com

Amiya Kumar Tripathy
Computer Engineering
Don Bosco Institute of Technology
Mumbai, India
tripathy.a@gmail.com

***Abstract*—Crime in Mumbai has always been a concern for both the government authorities as well as the citizens of Mumbai. With the increase in the population, the crime rate has also been on an increase which in turn increases the risk factor for the safety of the citizens of Mumbai. To deal with this, the citizens should not only rely on the government authorities to cater the needs but at the same time, the citizens should also understand and adapt with the applications of technology in this very era of technology, to help provide better safety for all the citizens residing in the dynamic city of Mumbai. Safety while traveling can be considered as the major need in Mumbai, as Mumbai is regarded as the city that never sleeps. There have been several at-tempts made by the government authorities as well as individuals to help provide a model for better safety while traveling but still it has never been completely automated. This research pro-poses a model that works in synergy with various steps that include crime data collection, and then analysis on that data to predict the crimehotspots which are specifically targeted to help citizens to distinguish between safe and unsafeareas while traveling.**

***Index Terms*—Crime Prediction, Safety Route,Web Scraping, Crowd-sourcing, Safety Android Application**

I. Introduction

To counter to common belief, crime is not continually focused in weak areas, with well-to-do elements of the town affected too. Therefore residents and guests would like info to avoid additional vulnerable viewpoints of the city. The idea is to make a reliable source of data instead of supposing episodic news reports that contribute to a way of hysteria. Our goal is to create what is already in public offered statistics accessible and unjust for citizens. Unwary residents can make easy targets because they stand out in a crowd, these citizens are usually unused to their surroundings, or are usually carrying cash, credit vouchers, and costly electronic things like cell phones, cameras. Due to this, the risk of being mugged or robbed is usually high while traveling[17].

In fact, 72% of the criminal activities that take place in Mumbai are due to avoiding the basic signs which could have been prevented if there existed a system that could help identify and distinguish between the safe and unsafe areas in Mumbai.[11] This research model observes and realizes after studying several criminal cases that Mumbai has encountered, that citizens usually ignore the fact to be safe when it comes to reaching a destination on time or in haste, and keeping the same thing in mind our group aims to build an application that helps citizens to distinguish and predict between safe and unsafe areas of Mumbai while traveling.

The analysis is achieved using multiple sources including the web scrapping, official First Information Report (FIR) reports and user ratings for every specific area on which machine learning is applied and then created a Training model based on the 240 official First Information Report (FIR) reports and additional user responses to help achieve an accurate safety index.

II. Literature Review

Nowadays, millions of citizens go through the streets of Mumbai town, taking some particular ways that area unit designed by victimization either public or non-public transit or perhaps walking. Although this area unit is popular routes that voters common frequently regard their traveling, different ways(apparently risky) could study, especial-for immigrants. Particularly, safety is an important characteristic that ought to be thought of for those mobile applications, especially in overpopulated large towns. Mobile demands for routing and planning in town conditions are more and more turning into vital for rising public areas.

Therefore, it means the looks of this next-generation of mobile information operations, during which advice area unit focused on call making order to support the growth of large towns competently.

Researcher, market analyzer, or academicians rather data from different websites for their better improvement. Copying of data on the website to end-user local storage in forbidden by most of the website authority. So that the user wants to manually copy the data from the website to local computer file storage, but such a task is very exhausting and time-consuming. Due to such limitation web internet scraping techniques area unit introduces by exploiting net scraping techniques user will extract info obtainable on multiple web site into a single database or spreadsheets. So data can be easily visualized and analyze for further use. The Web scraping technique is a sub-discipline of web mining technology.

978-1-7281-2339-4/20/$31.00 ©2020 IEEE

We have reviewed a number of safety applications that have been published on the play store/ios AppStore that are freely available to be used by the users. To review some of the applications like ”Raksha ,A Women’s Safety App” and the ”Nirbhaya App” what we could draw the conclusions were that these applications majorly focused on sending out an SOS alert to the guardians of the users in the times of difficulties and also track the user’s location, which in case of the Nirbhaya app could be triggered using a panic button.

On the contrary, our application is a combination and aims to treat safety in a more efficient way, by calculating the crime rate of every particular area based on the official crime First Information Report (FIR) of that particular area, news/articles that are about the criminal activities in that area and also through the user reviews, which then is used to calculate the safety index of that area that can help the user be acquainted with the crime rate of the area, also this is even depicted on the maps for a better user experience.

We have got our app reviewed from a Mumbai Police Personnel: Mushtaque Maner working as Asistant sub Inspector (ASI) in Taloja police station. Whose review is “The application is good when it comes to detecting the crime rate of an area but a certain areas of the app can be worked on or certain features can be added such as the Crime Patterns of every particular area as the way crime might take place in one area might differ from the other, so if this could be added, it would be better”.

III. Proposed Architecture

Data from numerous open sources like news, open Data, or through user input or web scraping knowledge from the web site are going to be kept within the information Server. At that time, feature choice is going to be done on knowledge accessible within the information. Once feature selection is finished, knowledge cleansing is going to be performed on knowledge to get rid of vociferous knowledge and chemical agent. Knowledge having great circle and latitude is going to be premeditated on the QGIS server. Finally, the Prediction algorithmic rule can work on the given knowledge that is plotted within the MAP.[10] The output is going to be a map accessible for the user on android Application and additionally for Agencies, and it’ll be accessible within the sort of an internet site with a completely different visual image of knowledge options.

The Application can use high-level machine learning knowledge to divine crime rates within the town’s areas at completely different times of the day and night. The applying is going to be ready to predict new patterns of crime within the space with the assistance of incoming new crime data set.

In the figure (1) its shown Proposed System Architecture having Data source, Machine Learning Model, and Back-end Components are main essential part.The training of the machine learning models data was collected from 2 data sources. The first one being the official First information report(FIR) crime reports and another one being the data collected using Web scrapping. These datasets were integrated into a single data-set and trained using the Random Forest Algorithm. These trained models

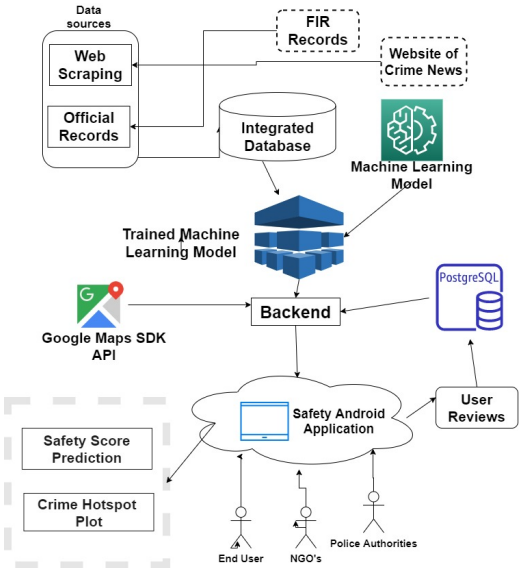

Fig. 1. Proposed System Architecture

were added to the back-end with the Maps SDK for Android Google Map API. The user of the Application can enter his/her reviews, and that gets stored in the back-end PostgreSQL database.

IV. Research Methodology

A. Multi-approach

- Crowdsourcing : The survey results carried out to the point that Lighting, Walk Path, Visibility, Security, and Transport were the important five parameters. The particular area is rated based on the above five considered parameters, and each one of the parameters can be rated in the index of 1-5, based on what scale the parameter that is considered is true. Thus, finally based on rating, it will tell us that the area on the right side has a much higher safety index than the area on its left, thus making it safer and less crime-prone.
- Web Scrapping : Web scraping is a method of automating the uprooting of data in an effective and quick way. With the advantages of web scraping, any end users can extract data from any website, no matter how large the data is, on your computer. Further, websites may have data that user required cannot copy and paste. Web scraping can help us extract any data that any user wants.[4] From the news article from TimesofIndia paper has scraped the critical data we are required for the project. Data can be places, type of crime, day, age of the victim. which is available in any news article of crime-related.[20] These data were extracted from news and appended into a Comma-separated values (CSV)
- Official Reports : Under the public open-source portal, Maharashtra police provide the facility to view published FIR online[16]. FIR can be downloaded in a variety of formats like Comma-separated values (CSV), JSON, PDF. CID Branch also provides the weekly and monthly statistical reports of each District[15]. These reports can be useful to analyze the

crime on a monthly or weekly basis[17]. Maharashtra Police has provided a citizen portal to access the FIR published on the Internet. Portal allows us to view FIR published in any three months range of specifically unit areas. This portal gives the latest records as per the search parameters specified in the search query.

Since First Information Reports (FIRs) are the official crime reports of an area that have been documented, this is given the highest priority when it comes to its contribution to the safety index. Articles/News that have been web scrapped have been then given the next priority and lastly the user reviews have been given the least priority as it's not completely authenticated.

B. Integrated Data

First Information Report (FIR) has some critical fields which can be used in our project system. Some of the critical parameters which we can extract are Age of a victim, Place, Date, Time. Once the parameters are extracted, we can insert into the central database, which is Integrated with other Data source. The News scraping data is integrated with common Comma-separated values(CSV) file based on which Machine learning model is trained. The approach towards building an app is based on three aspects, which are Official Reports, Web scrapping, and Crowd-sourcing. We will train our model initially using Official Reports and then build it up in the sequential phases using Web scrapping and Crowd-sourcing. It has built a system that works on all the aspects and integrates all technologies into one. In the figure (2) given below shows the Integrated Data mapping on Maps.

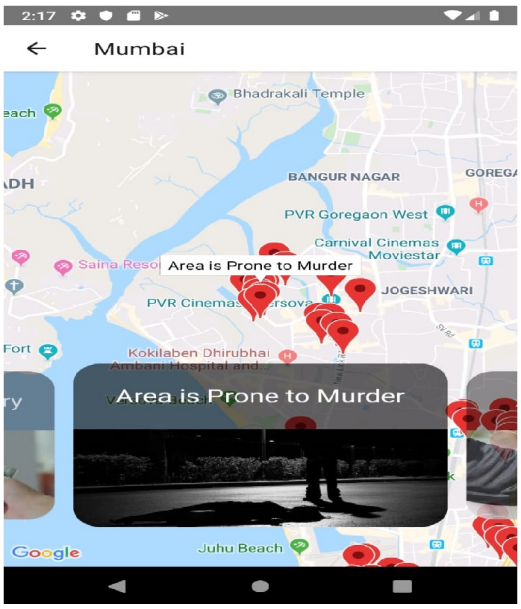

Fig. 2. Crime Density plot of Murders in Mumbai

C. Analysis

The Graph in figure (3) shows the plotting of Crime committed at different time intervals in Mumbai city. Thus, from the graph analysis it is seen that between 12AM to 12PM most of the crimes are committed.

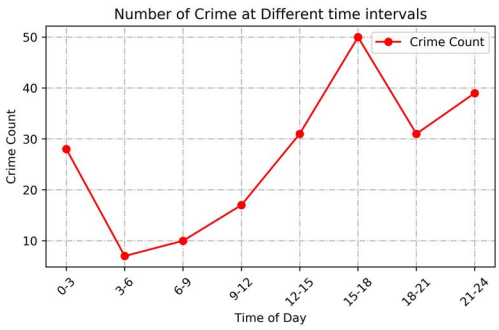

Fig. 3. Estimate of Crimes at Various Interval

The below mentioned graph depicts the crime rates at different parts of the city ranging from low to high based on the number of thefts, robberies and/or murders and also based on the 245 responses we accumulated during the initial data collection. It can very clearly observed that cities like Andheri and Bandra have a high rate of number of thefts, similarly Andheri, Bandra, Dadar have a high rate of number of robberies, and also Andheri, Kurla and Mulund have a high rate of number of murders.

The Graph in figure (4) on next page shows the plotting of various crime events like Murder, Theft and Robbery in several cities of Mumbai. It is clearly visible that Theft is the most committed crime in all the cities and Murder being the least frequent one

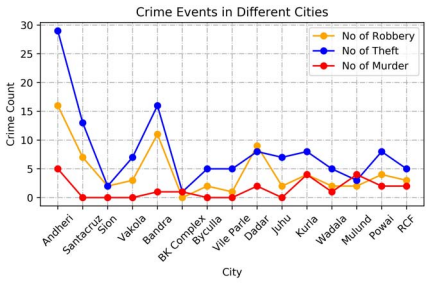

Fig. 4. Crimes in Different Cities

V. Implementation

A. Overview

The mobile application will be using Django as back-end, its flexible, stable and customizable. The Backend API will have a simple model for Location, Rating and Prediction object, which will have its respective serializer class created using Django Rest framework. The Backend API will also have respective API functions in views.py which will server requests at end point. The Frontend Mobile Application is created using React Native. It works with Expo SDK and apps so testing the application is convenient on mobile devices as well as simulator. The Application fetches data from the server and renders it to the view. In Short, this application will only be developed for Android, React Native is powerful framework for building both iOS and Android apps.

B. Algorithms

Random Forest Algorithm has been used to train the Machine Learning models. It is a is a supervised classification algorithm. The larger the number of trees in the Random Forest Algorithm, the more accurate the result gets.

Two Machine Learning Models have been trained here using the crime datasets. First one being the classification model used for the prediction of the type of crime at a specific area and another being the Regression model used for predicting the safety score of the location.

The reason why Random Forest is used for the training purpose here is because it can be used for both classification and regression problems as well as it avoids Over-fitting of a machine learning model and also it is an algorithm that can very well deal with missing values in a data-set.

The figure (6) shows the actual implementation of the Application on above mentioned algorithms.

C. Outputs

The Graph in figure (5) shows the plotting of Safety Index Value in several cities of Mumbai. The lower the Safety value the more crime prone the location. Thus, from the graph analysis it is obvious that Andheri, Kurla and Bandra are unsafe because it has the least safety index values.

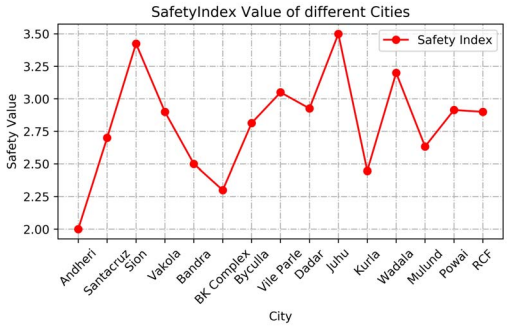

Fig. 5. Safety Index Analysis

VI. Conclusion

In this paper, the complete analysis, classification, and prediction of the Mumbai City is done using crowdsourcing information with the information provided from social networks and also the official crime reports, which in this case is FIR. The spatiotemporal analysis can confirm the situation wherever the crime event occurs. Moreover, the confidence level of a location is going to be outlined. And it'll be used as a parameter for computing the safety index. Finally, we have come to a conclusion based on the research, and the analysis we conducted is that cities like Andheri, Juhu, and Kurla are comparatively not very safe for the users.

VII. Future Improvement

The future scope for our application is quite extensive as it includes majorly 2 important aspects:

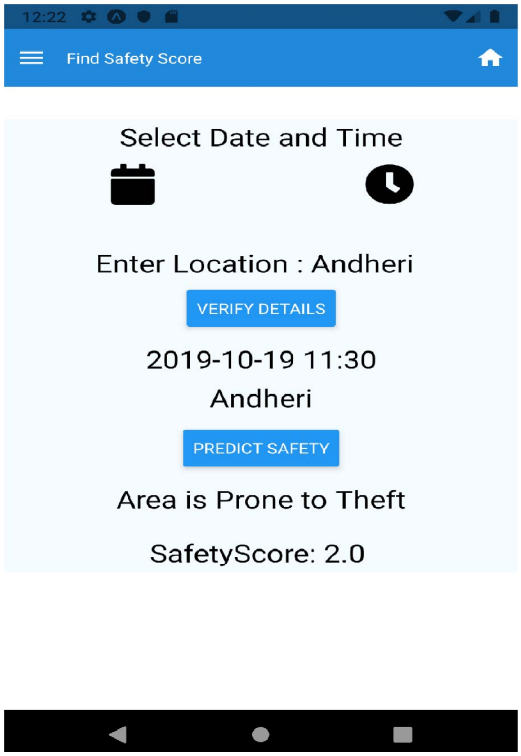

Fig. 6. Model Predication Output on App Interface

- To get the app functional to every minute area of Mumbai, as currently, it focuses only on the key areas of Mumbai.
- To work on the individual crime patterns with respect to every area based on what the crime was so as to get more accuracy and efficiency to help the users understand their safety better.

References

[1] H. Su, K. Zheng, J. Huang, H. Jeung, L. Chen, and X. Zhou, "Crowdplanner : a crowd-based route recommendation system," in Proceedings of the 30th IEEE International Conference on Data Engineering (ICDE'14), pp.1144–1155, IEEE, Chicago, USA, April 2014.

[2] V. Arnaboldi, M. Conti, and F. Delmastro, "Implementation of CAMEO: a context-aware middleware for opportunistic mobile social networks", in Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'11), pp.1–3, Lucca, Italy, June 2011.

[3] Firoz Ahmad, Md Meraj Uddin, Laxmi Goparaju, " Role of Geospatial Technology in Crime Mapping: A Perspective View of India", in World Scientific News Journal 88(2), Apirl 2017, pp.211-226.

[4] M. Nagarajan, K. Gomadam, A. P. Sheth, A. Ranabahu, R. Mutharaju, and A. Jadhav, "Spatiotemporal-thematic analysis of citizen sensor data: challenges and experiences," in Web Information Systems Engineering— WISE2009, vol.5802 of Lecture Notes in Computer Science, pp. 539–553, Springer, Berlin, Germany, 2009.

[5] S.C.M. de S Sirisuriya, "A Comparative Study on Web Scraping" in Proceedings of 8th International Research Conference, KDU, Nov 2015, pp.135-140.

[6] Félix Mata, Miguel Torres-Ruiz, Giovanni Guzmán, Rolando Quintero, Roberto Zagal-Flores, Marco Moreno-Ibarra, and Eduardo Loza, "A Mobile Information System Based on Crowd-Sensed and Official Crime Data for Finding Safe Routes: A Case Study of Mexico City", Hindawi Publishing Corporation Mobile Information System, Article ID 8068209, pp.3-5, 2016.

[7] Varvara Ingilevicha, Sergey Ivanov, "Crime rate prediction in the urban environment using social factors", 7th International Young Scientists Conference on Computational Science, YSC2018, 02-06 July2018, Heraklion, Greece.

[8] Hitesh Kumar Reddy, Bhavna Sainia, Ginika Mahajana. "Crime Prediction and Monitoring Framework Based on Spatial Analysis". International Conference on Computational Intelligence and Data Science (ICCIDS), Volume 132, 2018, Pages 696-705.
[9] Alexandros Belesiotis, George Papadakis, Dimitrios Skoutas . "Analyzing and Predicting Spatial Crime Distribution Using Crowdsourced and Open Data". ACM Transactions on Spatial Algorithms and Systems, Vol. 3, No. 4, Article 12 (2018).
[10] Isha Pradhan, Katerina Potika, Magdalini Eirinaki and Petros Potikas. "Exploratory Data Analysis and Crime Prediction for Smart Cities", Proceedings of the 23rd International Database Applications and Engineering Symposium, June 2019, Article No: 4, Pages 1–9.
[11] Aman Kumar, Nikhil Tiwari, Prakhar Gupta, S.N.Rajan. "WOMEN CRIME PREDICTION", International Research Journal of Engineering and Technology (IRJET), Volume: 04, Apr -2017.
[12] Ginger Saltos, Mihaela Cocea. "An Exploration of Crime Prediction Using Data Mining on Open Data", Article in International Journal of Information Technology and Decision Making, May 2017.
[13] Alkesh Bharati, Sarvanguru RA.K, "Crime Prediction and analysis using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 05, Sep 2018.
[14] Chao Huang, Junbo Zhang, Yu Zheng, Nitesh Chawla, "Deepcrime: Attentive Hierarchical Recurrent Networks for Crime Prediction", Proceedings of the 27th ACM International Conference on Information and Knowledge Management, October 2018, pp.1423–1432.
[15] http://mahacid.com/monthly-clime-review.html last accessed on 31-12-2019.
[16] http://www.mhpolice.maharashtra.gov.in/Citizen/MH/ last accessed on 31-12-2019.
[17] http://ncrb.gov.in/ last accessed on 31-12-2019.
[18] https://data.gov.in/ last accessed on 31-12-2019.
[19] https://www.diva-gis.org/ last accessed on 31-12-2019.
[20] https://timesofindia.indiatimes.com/city/mumbai/mumbai-two-arrested-for-stealing-goods-from-moving-train-coaches/articleshow/71788161.cms last accessed on 31-12-2019.