User Specific Safe Route Recommendation System

Yash S Asawa[1*]
[1]School of Computer Science and Engineering,
VIT University,
Q-348, Mens Hostel,VIT University, Vellore,
Tamil Nadu, 632014, India
*Corresponding Author

Varun V[3]
[3] School of Computer Science and Engineering,
VIT University,
Q-819, Mens Hostel,VIT University,Vellore,
Tamil Nadu, 632014, India

Samarth R Gupta[2]
[2] School of Computer Science and Engineering,
VIT University,
Q-819, Mens Hostel,VIT University,Vellore,
Tamil Nadu, 632014, India

Nikhil J Jain[4]
[4] School of Computer Science and Engineering,
VIT University,
Q-218, Mens Hostel,VIT University,Vellore,
Tamil Nadu, 632014, India

***Abstract*: Criminal activities have reached unprecedented levels in almost every part of the world. Desperate times like these require desperate measures to ensure the safety of people, especially people that need to travel on a daily basis to places, known and unknown. The majority of these criminal offences occur while the victim is travelling, irrespective of the mode of transport: walking, personal vehicles, public transport vehicles, auto-rickshaws, or cabs. This paper proposes a User-Specific Safe Route Recommendation System which presents a safe route visualized on maps to the user based on the past criminal records of the geographical region. Our approach is implemented on two-levels, first to realize the user-specific features using a Decision Network and the latter to actuate the safe route generation using Geospatial Data Analysis. We visualize the determined routes through a colour-code based map interface. Keeping in mind the real-life impact our project needs to create; we have demonstrated our model on San Francisco city data to show the effectiveness of our approach.**

Keywords*: *Safe Route, User-Specific, Colour Coding, Map Interface, Crime Statistics, Geo-Spatial Data Analysis, Decision Network.

I. INTRODUCTION

The reality of the world that we live in today is that parents are afraid to let their kids walk to school. Women are scared of being intimidated by hooligans and eve-teasers as they walk home at night. And families prefer to stay in rather than go out fearing the gauntlet of binge drinkers and criminals with unknown intentions. Crime cripples the quality of life for all of us and affects every section of society. Rising crime has left people across the world feeling insecure and powerless. In a lot of these cases, the police are incapable of making people feel secure being remote and distant, unable to deal with issues that concern their neighborhoods the most.

The authorities often downplay the problems and pretend to be blind to the fact that disorder and violence have become a fact of daily life in many regions. Anyone that has accessed reliable crime statistics will understand and realize the urgent need for a simple, efficient and compact solution that can enable people to visualize the safest route for their transit, taking into consideration their age, gender and the time of their travel ensuring the provision of routes that are specific to the needs of each user.

A Safe-Route Recommendation System uses data, statistics, various algorithms and mathematical models to predict a route that can minimize the risk of becoming the victim to any criminal activity. The efficiency of such a system is of utmost importance since it can often lead to life or death situations.

The visualization of the predicted route is just as essential considering the average level of interpretative capabilities among the users since any safety-based application should have extensive reach and be used by every person that can benefit from it and lead a better life.

II. RELATED WORK

Previously a paper has been published to predict the safest route based on the lowest risk score. [1] They acquired accident and crime data available through NYC OpenData and used it to determine the average risk score for each cluster/region. Their model was based on risk score generation for each path using Machine learning algorithms, namely K Means Clustering and KNN Regressor algorithm.

Another paper [2] attempted data mining from geographically sorted records from over 12 years to ascertain the safest route. The age and gender of the user were also taken into consideration and an ID3 Decision Tree algorithm was used for risk determination but the age and gender parameters were used to selectively extract data and not as part of the prediction model.

A feature-level data fusion method [3] was also proposed that relied on deep neural networks (DNN). They used various online databases including crime statistics, demographic and meteorological data, and images. They performed statistical analysis to select crime data before generating the training data and trained their DNN which incorporated temporal, joint feature representation, spatial, and environmental context layers to achieve benchmark efficiency.

Be- Safe Travel [4] is another such recommendation application that combined Google APIs with a PHP hypertext preprocessor and MySQL Databases to provide a web-based platform. According to their model, the safest path is the path that has traveled through minimum crime points. Their model is based on data from Surabaya city. It is also capable of ranking up to three routes and allotting each of them a color to represent the level of security

Another paper presented SAFEBIKE [5] which recommends routes for bike-sharing services based on factors such as distance and the level of safety. In accordance with its primitive application, the system also provides the number of available bikes and docking stations. The inclusion of crime statistics and the consideration of safety as a parameter for the generation of routes ensure a more comfortable travel experience for the user.

We came across another paper [6] which implemented a risk model for the street urban network in Chicago and Philadelphia. This allows the estimation of the relative probability of criminal activities on any road segment. Their model was aimed at generating a path that is short and safe but instead of achieving both of them simultaneously they generate paths that achieve a tradeoff between both the parameters.

A Bayes Algorithm based implementation [7] was also presented which was realized by integrating crowd-sensed and official crime data through a mobile application. The model is a combination of semantic processing and classifier algorithms which is used to perform safe route generation and crime forecasting primarily using the twitter feed, using a geospatial repository for storing the relevant tweets.

Another model that we came across was TREADS [8] which suggests real-time safe travel itineraries using social media data and points of interest review summarization techniques. It incorporates safety and user interest factors and uses a transportation-related social feed retriever to provide safe, efficient, and convenient transit strategies for travelers. It also uses a text summarization module to summarize the relevant social media data before it undergoes further processing and generates the itinerary.

[9] SB Oh et al. presented a method incorporating the time and day of the user's travel. They computed the crime risk rate of individual facilities and combined the risk rates of the facilities along a particular route in order to acquire the risk rate of the entire route. 5 different types of criminal activities were considered and each facility was allocated a risk level index that ranged from 1 to 10.

In [10] Zhaojian Li et al. proposed a cloud supported model for safe route prediction. A hybrid neural network is used to mine a road and accident database from the highway safety information system to compute an index referred to as a Road Risk Index. They have included factors such as the time of travel, the day of the week the user is traveling and the weather conditions to make the index dynamic. They also inculcate the time required for traveling the entire route as a factor, making their approach multi-objective in nature. Also, vehicle-to-cloud-to-vehicle connectivity was exploited to implement the system.

In [11] two researchers have presented their approach for the selection of safe routes for self-driving vehicles. The vehicles are assumed to be connected to a cloud-based environment. The system analyses real-time data from the vehicle, real-life accident data and employs big data mining to create a safe trajectory in the complete absence of any human intervention. The user only has to give his/her travel preferences and is presented a detailed trajectory along with the time required for the trip, the distance to be traveled and the fuel consumption for the entire journey.

Nima Hoseinzadeh et al. [12] also proposed to use big-data generated by vehicles connected to the cloud. They employ volatility as the primitive idea to quantify road safety and the behavior of the driver. Real-time traffic data from the vehicles and other relevant big data is fed to the framework to generate safety indices. The crash history of the driver spanning the last five years, average speed over the route, volatility in the acceleration and driver volatility are considered as the primary indices. They use a cost function called route impedance computed from a combination of the travel time and safety to determine the optimal route. And weightage can be dynamically allocated based on user preferences.

Bayesian networks have been in use for a long time for risk assessments. [13] gives significant insights and validates the efficiency of Bayesian networks in the determination of risk. Another study, [14] presents the use of Bayesian networks in modeling crime linkage.

III. FINDINGS AND CHALLENGES

The current technologies present the shortest path from a point A to B but the safety score of that path is not mentioned, forcing people to end up in lonely and ill-lit areas. These make them vulnerable to criminals and have been the cause of numerous crimes like Carjacking, sexual harassment, and robberies.

In the present systems and papers that we came across during the literature review, an intuitive notion is considered, where the crime factor is proportional to the punishment prescribed for the crime. More heinous the crime, the higher is the crime factor considered.

Our application takes a step ahead and uses previous criminal records of the locations to generate crime weights of the localities for Longitudes and Latitudes, time of the day and the intensity of the crimes based on the user attributes to visualize the route safety and alert the user and potentially provide an alternative safer route which can help them stay secure.

Most of the existing work focuses solely on women's safety and incorporates data only with respect to the users' gender. Also, the techniques generate only one safe path without providing any alternate routes to the user. None of the existing systems has any alerting mechanism to warn the users of any dangerous locations in their transit.

IV. DATASET DESCRIPTION

For the user risk calculation, we use the FBI UCR [17] database which contains well balanced and curated records of victims related to criminal offenses. This database spans a time-frame of over 15 years.

For the next phase of the algorithm i.e. route profiling, we use the San Francisco dataset which contains criminal incidents over 12 years that have occurred. We take into consideration the category of crime reported, and the geographical location of the occurrence of the incidence. These crimes range from petty traffic violations to assault and attempt to murder. Each incident within the city that occurred within the period of 12 years is considered in the dataset.

V. PREPROCESSING THE DATA

The datasets are preprocessed in such a way to suit the needs of the application. We extracted the right fields from the FBI records, compiled and merged them to fulfil our algorithm needs.

In the case of the San Francisco dataset, for each crime, the reported incident is tabulated and added to a count figure, such that the dataset now contains a particular crime count for a particular geographical location.

VI. PROPOSED WORK

A. User Risk Calculation

Most police records tend to include primitive information about the victims such as their age, gender and other descriptions related to the crime. Last few years have seen law enforcement agencies across the globe invest in modern technology solutions to digitize and maintain their records. Also, most of these records have been made publically available and can be availed in the form of open-source assets.

For our model, the FBI UCR [] database is used which contains well balanced and curated records of victims related to criminal offenses. This database spans a time-frame of over 15 years. The statistics present in the database are used for modeling our algorithm.

B. Data Justification

Keeping in mind the privacy of the users and realizing that the users need to be willing to share their data with us, we have restricted ourselves to model the algorithm using only the essential information. This includes the age, gender, and the users' preferred time of travel since most crime reports contain this basic information. 3 different datasets are used which contain statistical data related to crimes categorized on the basis of age, gender and the timestamp respectively. A thorough analysis of the data suggests that the risk of an individual, becoming the victim of a particular crime significantly depends on the aforementioned attributes.

Table 1. Comparison of Existing and Proposed Work.

Feature	**Existing Work**	**Proposed Work**
User-Specific - Risk Analysis	Most of the existing work focuses solely on women's safety and incorporates data only with respect to the users' gender.	Our proposed work aims to calculate the risk associated for each user taking into consideration their gender, age and the time of their transit with respect to crime statistics from over 12yrs worth of records
Colour Coding	The existing techniques generate only one safe path without providing any alternate routes to the user.	We classify the streets and provide a color code to enhance usability and user experience. We also consider alternate routes from the source to the destination so the user can make individual choices and has a sense of control throughout.
Automated Alert Mechanism	None of the existing systems has any alerting mechanism to warn the users of any dangerous locations in their transit.	Our application provides an interactive dashboard interface to the user to ensure that the user does not miss out on any critical detail during his/her transit.
Data Categorization	Only the required data (with respect to the users' gender) is taken from the dataset and the algorithm is applied to that.	We allocate risk weights to criminal incidents based on user attributes and these weights are used by the subsequent algorithms. The entire dataset is used in order to maximize the efficiency of the model. Also, we use geospatial data analysis to provide a ground-level risk to users.

We have presented the following graphs to indicate the same. Figure 1, 2 and 3 illustrate graphs to depict the relationship between Crime Rate and the Gender, the Time and the Age Group respectively.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 10, October-2020

Figure 1. Crime rate vs Gender/Sex

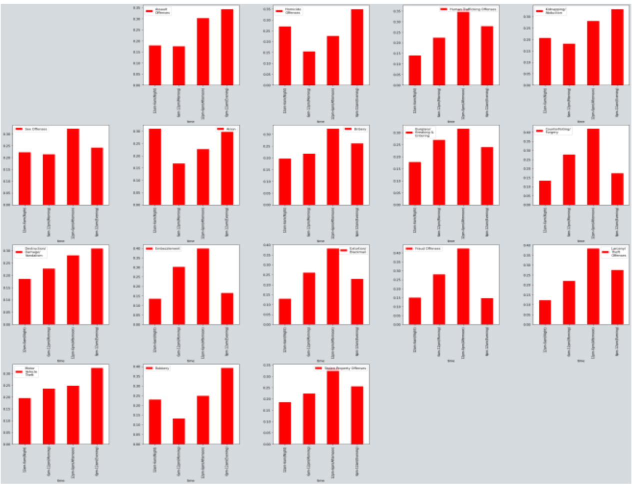

Figure 2. Crime rate vs Time

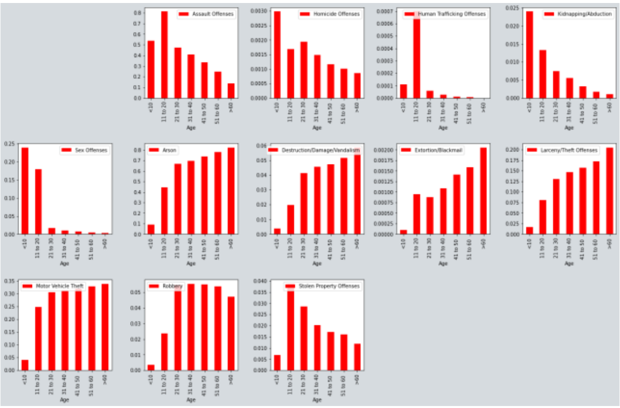

Figure 3. Crime rate vs Age Group

The charts presented above, highlight the relation between the occurrence of a criminal offense and the age, gender, and time of travel corresponding to the victim.

For example, if we consider stolen property offenses, we see that most of the offenses were against teenage male victims that were traveling in the afternoon. This suggests that people with the same attributes are more likely to be subjected to stolen property offenses than others. This correlation forms the foundation for our model and we present the use of a Decision Network to compute a single Risk Index value that quantitatively combines risk based on all the three factors.

C. *Algorithm*

1) *Decision Network*

We propose the use of a Bayesian Network for computing user-specific risk based on his/her profile. A Bayesian Network allows effective comprehension of inter-event dependence. It also allocates a probabilistic value to represent the likelihood of occurrence of that event given that some other event has already occurred. These relationships can then be used to draw conclusive inferences from the random variables in the graphs by using various factors.

From the datasets, we computed three types of conditional probabilities which are, the probability of the victim of a particular crime belonging to a specific age group, belonging to a specific gender and traveling at a particular time.

Figure 4 illustrates the entire decision network implemented in this paper.

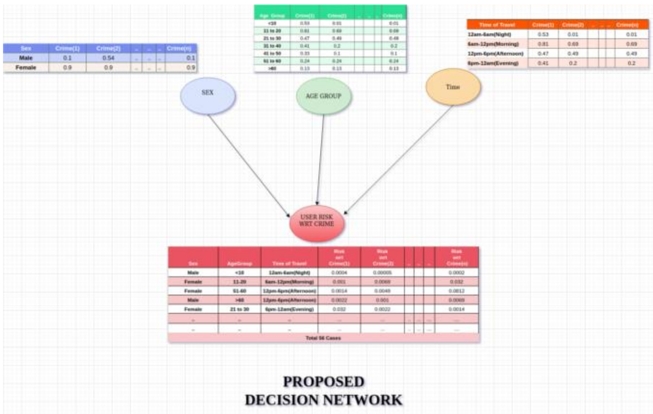

Figure 4. Proposed Decision Network

The network presented above contains age, sex and time of travel as the conditionally independent parent nodes. The final event, i.e., the risk to the user corresponding to a particular crime, is calculated as the joint probability of the parent nodes. Using the Local Markov Independence Assumptions, we represent the joint probability distribution as a factorized representation of these local conditional probabilities as specified by our network.

The decision network provides the likelihood of an individual becoming victim to a certain criminal offense. And is calculated as per the formula -

$$P\ (Risk) = P(Age) * P(Sex) * P(Time) \qquad (1)$$

2) *Crime Score and Risk Index Calculation*

Thereafter, we define a crime score that is assigned to each crime to account for the severity of the crime. A severe crime needs to be treated more sternly than a relatively petty offense and hence it is essential to consider the extremity of the crime for determining the risk any individual will face.

The crime score is assigned based on the category the crime belongs to and the level of punishment adjudged to an offender [6].

The culminating risk index is computed as the product of risk probability from the decision network and the crime score. The risk index is a quantitative risk rating that accommodates the probability of occurrence of a crime and its severity as a single quantified factor which is considered in the later stages of the algorithm.

$$Risk\ Index = P(Risk) * Crime\ Score \qquad (2)$$

Hence, the first phase of the algorithm performs a complete risk assessment for a traveler based on his/her profile and terminates with the risk index which is used as an input parameter for subsequent processing.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 10, October-2020

Table 2. Crime Score.

Crime	Index
Sex Offences	10
Assault	9
Kidnapping	8
Arson	7
Vandalism	6
Extortion	5
Vehicle Theft	4
Burglary, Robbery, Stolen Property	3
Bribery, Forgery, Fraud	2
Larceny	1

3) *Reduced Geo-Spatial Information*

Before further processing, the system needs to identify the coordinate points that lie between the user-defined source and destination. This can be very time consuming and is likely to degrade the overall system performance. Therefore, we use Clustering algorithms to minimize the number of coordinates the system needs to check every time a new use-case is generated. We came across various clustering algorithms implemented for similar purposes. Density-based spatial clustering works best on geo-spatial data, but it is not sensitive to the outliers.

The idea behind using clustering was to identify points that can represent multiple points in smaller subregions as single units. It does not add to or alter the geospatial data, previously acquired by the system. The inclusion of all distinct coordinate points, including outliers, is essential for accurate prediction. So, we use the K-Means clustering algorithm to make the coordinate cluster. We store the cluster crime count data with the centroid of the corresponding cluster, obtained from K-Means clustering as an index attribute.

Figure 5 depicts the results of K Means Clustering in the form of a chart.

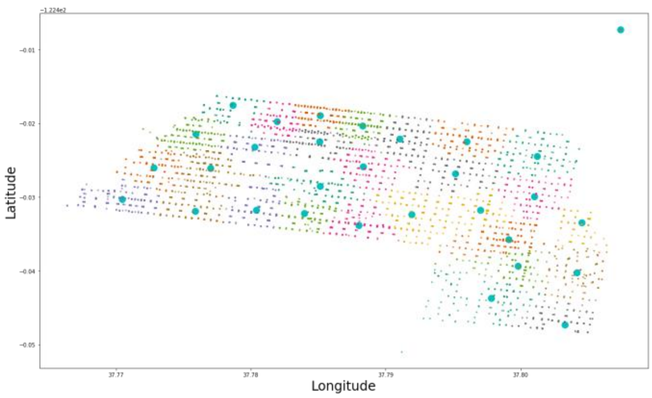

Figure 5. K means clustering performed on a particular region in San Francisco

4) *Optimal Route Profiling*

The system takes the location of the origin and destination as input from the user for route recommendation. Then, using the Google Maps API, three shortest routes from the origin to destination are discerned. The routes are split into small parts, and each part is marked based on the evaluated safety index. The system proposes the optimal route based on the safety index and length of the route. The crime index of each part is deduced by acquiring the crime count of the clusters which fall within a preordained distance from it. The preordained distance is the mean of the radius of all coordinate clusters.

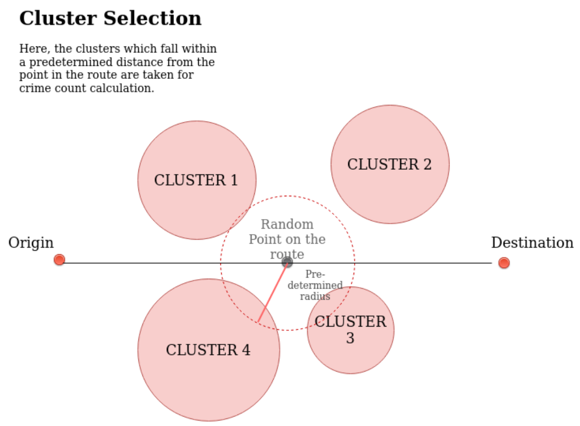

Figure 6. Nearest Cluster Selection

Figure 6 illustrates the process of cluster selection undertaken by the proposed model.

The street risk is defined as the product of the count in crimes of nearest clusters and the risk index calculated in the first phase of the algorithm.

$$\text{Street risk(S)} = \sum_{c=1}^{n} (count(c) * risk\ factor(c)) \quad (3)$$

Where, S = Street Risk, n= Total type of Crimes, c = crime, count = Count of that particular crime reported in nearest clusters and risk index = risk probability of user wrt crime c.

After calculating the risk of each street belonging to the route, route risk is calculated which is a weighted average of all the streets falling within the route.

$$\text{Route risk} = \frac{1}{t}\sum_{i=1}^{t} (Si) \quad (4)$$

where t = total number of streets within the route and Si = Street risk of street i.

5) *Colour Coding Scheme*

A simple color scheme is enforced for easy identification of the safest route. Since the system generates three routes (shortest), we mark the safest route in green and the alternative paths in red. We avoid allocating separate colors to the smaller sections of the route to ensure that the system interface is easy to understand for the user. Apart from the color scheme, the system also displays descriptive text in a pop-up box upon cursor movement over the corresponding route.

Once the Route risk for all the streets is calculated, color coding is done and the routes are accordingly visualized. Below is the complete algorithm for optimal route profiling.

Figure 7 illustrates the route profiling algorithm in the form of a flowchart.

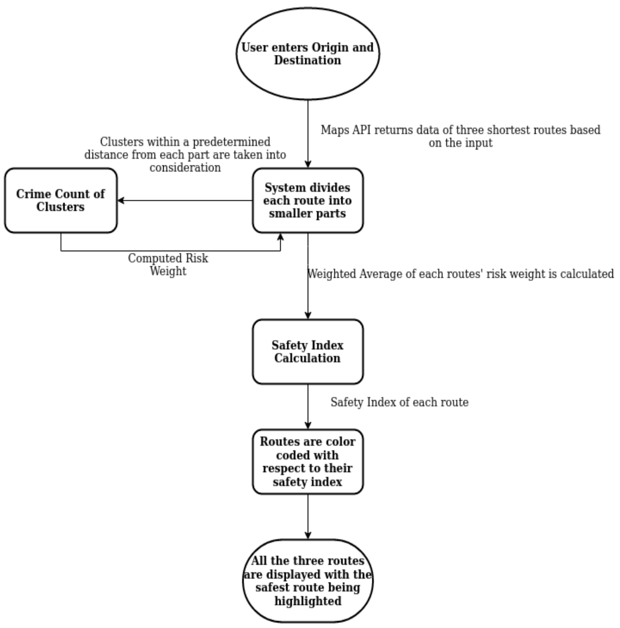

Figure 7. Optimal Route Profiling

VII. IMPLEMENTATION

We have precomputed the risk probabilities, using the Bayesian approach, and also the cluster dataset using K MEANS clustering.

When the project is deployed in real-time on a platform, the precomputed data is stored in databases. When a user makes a query, the data is fetched from the Routes API and appropriate data exchange is done with the databases, such that the route - profiling algorithm provides the optimal path, which is visualized with a safe color signal on the maps and a dashboard interface.

Complete Data Preprocessing, risk Calculation, and, Clustering of data are performed using python. For testing purposes, we have implemented our model over the Northern District of San Francisco. The web-based interface provides the user the capability to query for routes, and assists the user for any warning or alerts using an interactive dashboard.

VIII. RESULTS AND CONCLUSION

Our algorithm is tested on the streets of San Francisco and recommends the safest route between locations. The algorithm is tested for various use cases of user data, and different source and destination locations. The intuitive user interface is designed to meet the real-time needs of the user and is illustrated in Figure 8 and 9.

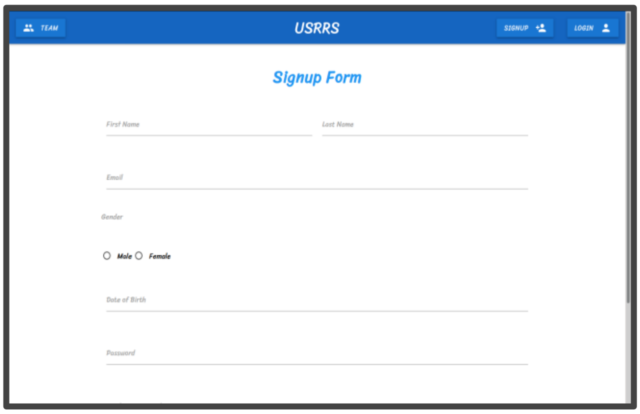

Figure 8. Register form for storing user basic attributes.

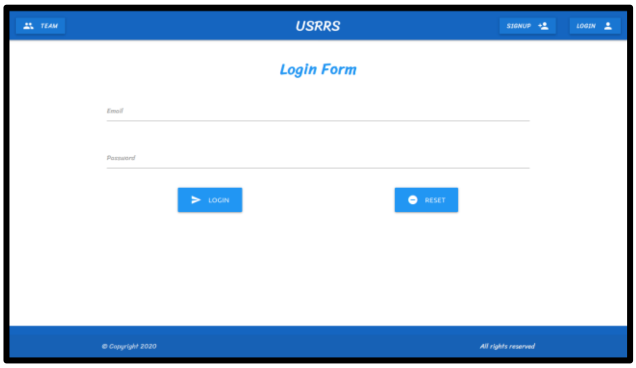

Figure 9. Login form for validating user basic attributes.

The Dashboard interface provides users with smart alerts regarding the criminal offenses in a particular locality, along with detailed information about each path and is illustrated in Figure 10.

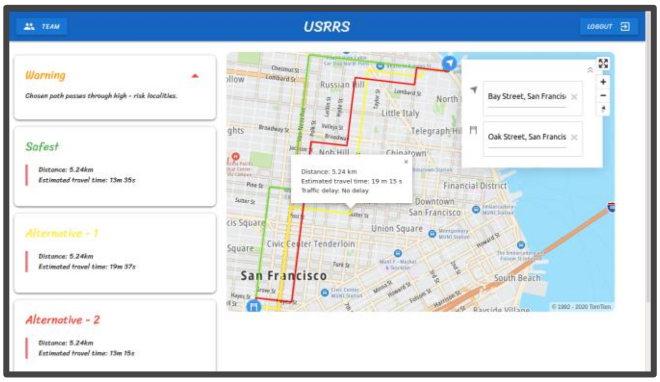

Figure 10. Safe Route Visualization with Smart Alerts

We achieve our project objective to provide safe routes on maps by using a decision network and Geo-spatial data analysis on past crime data.

In our paper, an efficient model is proposed that considers various user-attributes (age, gender, and the time of travel), mines data from a relevant data source, and constructs a Bayesian decision network to compute a risk-factor.

Google API technology is combined with the K-means clustering algorithm to generate street profiles and an optimal safe-route is generated and visualized efficiently through various techniques. The utilization of a user-specific risk factor and criminal statistics make the proposed approach significantly better than its predecessors including the well-known google maps that only consider distance and time-related constraints. The model also surpasses its contemporaries in terms of the number of relevant users' attributes considered for safe route generation.

IX. FUTURE SCOPE

Considering the application of the model there is definite potential for future work that can be done to broaden the scope of the proposed approach. The inclusion of more factors in route safety determination is a potential area of work for future studies. The proposed model can be implemented in the form of an android app, to expand the reach and improve accessibility by making it available across platforms. Also, more interactive features can be added to the interface to enhance the user experience.

REFERENCES

[1] Soni, Shivangi, Venkatesh Gauri Shankar, and Sandeep Chaurasia. "Route-The Safe: A Robust Model for Safest Route Prediction Using Crime and Accidental Data." (2019).
[2] Sarang Tarlekar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3), 2016, 1536-1540.
[3] Kang, Hyeon-Woo, and Hang-Bong Kang. "Prediction of crime occurrence from multi-modal data using deep learning." PloS one 12.4 (2017).
[4] Utamima, Amalia, and Arif Djunaidy. "Be-safe travel, a web-based geographic application to explore safe-route in an area." AIP conference proceedings. Vol. 1867. No. 1. AIP Publishing LLC, 2017.
[5] Zhong, Weisheng, Fanglan Chen, Kaiqun Fu, and Chang-Tien Lu. "SAFEBIKE: A Bike-sharing Route Recommender with Availability Prediction and Safe Routing." arXiv preprint arXiv:1712.01469 (2017).
[6] Galbrun, Esther, Konstantinos Pelechrinis, and Evimaria Terzi. "Urban navigation beyond shortest route: The case of safe paths." Information Systems 57 (2016): 160-171.
[7] Mata, Félix, 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." Mobile Information Systems 2016 (2016).
[8] Fu, Kaiqun, Yen-Cheng Lu, and Chang-Tien Lu. "Treads: A safe route recommender using social media mining and text summarization." Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2014.
[9] Oh, Su Bin, Hyeok Ju Park, Byeong Ki Kang, Ill Chul Doo, and Mee Hwa Park. "Safe Route Recommendation Method to Prevent Crime Exposure." Advanced Science Letters 23, no. 10 (2017): 9579-9583.
[10] Li, Zhaojian, Ilya Kolmanovsky, Ella Atkins, Jianbo Lu, Dimitar P. Filev, and John Michelini. "Road risk modeling and cloud-aided safety-based route planning." IEEE transactions on cybernetics 46, no. 11 (2015): 2473-2483.
[11] Al Najada, Hamzah, and Imad Mahgoub. "Autonomous vehicles safe-optimal trajectory selection based on big data analysis and predefined user preferences." 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2016.
[12] Hoseinzadeh, Nima, Ramin Arvin, Asad J. Khattak, and Lee D. Han. "Integrating safety and mobility for pathfinding using big data generated by connected vehicles." Journal of Intelligent Transportation Systems (2020): 1-17.
[13] https://www.ncbi.nlm.nih.gov/books/NBK396478/
[14] https://www.researchgate.net/publication/271080746_Modelling_crime_linkage_with_Bayesian_networks
[15] https://personal.utdallas.edu/~nrr150130/gmbook/bayes.html
[16] https://cuent.github.io/pgm/notes/2019/05/18/pgm-intro.html
[17] https://www.fbi.gov/services/cjis/ucr

AUTHOR BIOGRAPHIES

Yash S Asawa is currently pursuing his Bachelor's degree in Computer Science and Engineering from Vellore Institute of Technology, Vellore, India. He has previously published a paper on Multi-Document Summarization Techniques. His research interests include Machine Learning, Artificial Intelligence, Natural Language Processing, and Automated Systems.

Samarth R Gupta was born in Gwalior, Madhya Pradesh, India in 1999. He is currently pursuing Bachelor's degree in Computer Science and Engineering from Vellore Institute of Technology, Vellore, India. His Research Interests include Machine learning, Pattern Analysis and Machine Intelligence, and Automated systems.

Varun V is currently pursuing Bachelor's degree in Computer Science and Engineering from Vellore Institute of Technology, Vellore, India. Research Interests include Machine Learning and Artificial Intelligence.

Nikhil J Jain is currently pursuing a Bachelor's degree in Computer Science and Engineering from the Vellore Institute of Technology, Vellore, India. His Research Interests include Cyber Security, Computer Networks, and Cloud Computing.