APPLICATION TO DETERMINE THE SAFEST ROUTE

[1]Asmita Bhat, [2]Vrushali Patil, [3]Madhuri Sonawane, [4]Shital Bagul
[1]Computer Engineering,
[1]Pune Institute of Computer Technology, Pune, India

***Abstract*— As criminal activities are increasing day by day, it is very important to ensure the safety of people to the best possible extent. Current times, demands people to travel to various places, known or unknown. People may not be aware of the unsafe areas and unsafe streets while travelling. In this paper, we attempt to mine the records of 12 years based on geographical locations to determine the safest route amongst all possible routes for given source and destination at a particular time. We use the ID3 decision tree algorithm and determine the risk involved for the streets within a given route. Also, based on the gender and the time of journey, the safest route is suggested so as to provide safety while travelling. This paper discusses the approach, dataset used, the ID3 algorithm and results. This algorithm can be used for various applications such as area wise crime analysis, criminal profiling and assisting in real estate for prospective buyers.**

***Index Terms*— Data mining, Spatial Database and GIS, ID3 Decision tree, Safest Route.**

I. INTRODUCTION

In recent years there is an exponential increase in the criminal activities across whole world. With the help of today's highly developed technology and the availability of criminal data we can easily do crime analysis so that we can increase the security of citizens. We use the dataset of San Francisco to develop a model for determining the safest path from one location to another.

A. Dataset Description

Dataset comprises of San Francisco's record of criminal activities over 12 years. The attributes considered are date and geographical location of occurrence of incidence, the time and address of occurrence of the crime , the description and the category of crime , action taken against the crime .

The dataset ranges from minor activities like traffic violation to severe activities like assault and attempt to murder.

Every single incident that occurred in the city within these 12 years is considered.

II. RELATED WORK

A group created a mining model to find a decision tree based on attributes from criminal cases. [1] This application can be used to find likelihood of involvement of particular suspects with similar criminal activities, by making decision rules from this tree. The main focus in this method is on criminal profiling. [7]

Another group put forth an algorithm that detects patterns of crime committed by the history sheeters. They proposed this by successively adding crimes to patterns based on likeness of attributes of the crimes. [3]

One other application was presented in which the likelihood of future occurrence of residential burglary was determined. [8] The authors attempted to use various classification techniques to determine the best suited attributes for considering the likelihood.

All the works focus on using either grid based spatio temporal analysis or criminal profiling using classification techniques.

III. PROPOSED APPROACH

A. Data Preprocessing

The dataset contains various attributes such as the address of the place where the incident occurred, the geographical location of the incident, the victim and the culprit of the crime,the type of the crime, the category of the crime, the action taken against it, the time of occurrence of the crime, etc. But not all of these fields are needed for application. For this, extraction of the relevant attributes from the dataset is carried out using python script which gives street profile as the output.

A street profile is a table consist of street name and count of the occurrences of each individual category of crime for particular time slot. The time is divided into four subcategories - Morning (5 a.m. -12 p.m. IST), Afternoon (12 p.m. – 5 p.m. IST), Dusk (5 p.m. – 8 p.m. IST) and Night (8 p.m. – 5 a.m. IST). Only the crimes that are relevant to this particular application are considered. Irrelevant crimes such as felony or minor traffic violations are excluded from designing the profiles of streets.

After the prepossessing, the data is cleaned and is represented in a format as shown in Fig 1.

	A	B	C	D
1	Count	Street Name	Category	Time
2	5	10TH AV	ASSAULT	DUSK
3	9	10TH AV	ASSAULT	MORNING
4	15	10TH AV	ASSAULT	NIGHT
5	10	10TH AV	ASSAULT	NOON
6	4	10TH AV	DRIVING UNDER THE INFLUENCE	NIGHT
7	1	10TH AV	KIDNAPPING	NIGHT
8	2	10TH AV	ROBBERY	DUSK
9	1	10TH AV	ROBBERY	MORNING
10	6	10TH AV	ROBBERY	NIGHT
11	1	10TH AV	ROBBERY	NOON
12	3	10TH AV	SEX OFFENSES	DUSK
13	1	10TH AV	SEX OFFENSES	NIGHT
14	1	10TH AV	SEX OFFENSES	NOON
15	19	10TH AV	VEHICLE THEFT	DUSK
16	17	10TH AV	VEHICLE THEFT	MORNING

Fig. 1 Sample of street profiles extracted from the dataset

B. Input for Safest Route

As the proposed application is end user android application, the source and the destination are taken as input using google maps api. Additionally, the age and gender of the user will be also taken as input. Also, the time will be taken as input from the android application. User will provide input in the form of two locations A and B respectively. The route is located using GPS services.From A to B there may be multiple possible routes available. Let route$_i$ be ith route between A and B. Each route will have the precise navigation information as the street name and the distance to be travelled on that street. Each route may consist of one or more streets. For a route$_i$, let jth street on the route$_i$ be labelled as street$_{ij}$. All such street information is send to server where the analysis is carried out to find risk involved in each route.

C. Analysis of Safety of Each Route

There can be one or more streets for each route. Each street within the route will be analysed based on the information available in dataset and the count for each crime by using various server scripts. For each street, value of the offence is considered. The values of offences will be such that the age and the gender of the user and the time will be considered. The dataset will be filtered for age specific and gender specific crimes. Only the data applicable to that particular user will be taken for further analysis. Each street$_{ij}$ has its profile fed to the ID3 classifier. The ID3 classifier will return the safety value for every street involved in analysis. For each route$_i$, the safety of each of its street$_{ij}$ will be calculated in terms of “Yes” or “No”. This information will be used for further analysis.

D. Building the Classifier

The street profiles are reffered by local control experts such as police officers and other law enforcer to determine whether street is safe or not. This data will be used for building a classifier. The classifier output will be in terms of “Yes” or “No” which states whether a street is safe or not.

E. Evaluating the Safest Route

The safety of the route can be determined by combining the results for each street. The already safe streets need not be considered for further analysis of safest route.The processing is carried out on unsafe strrets to evaluate the risk weight. Risk weight is the risk involved in a particular street. Risk weight for a street is calculated by summing up the multiplication of the count of occurrences of nth crime with the crime factor for particular crime. Crime factor is numeric value ranging from 1 to 10 working as a multiplier. By intuitive notion, crime factor is proportional to the punishment prescribed for particular crime. More heinous the crime higher is the crime factor.

Crime factor of crime ∝ Sentence served for that crime. The risk for a route$_i$ is the summation of the product of the risk weight of street$_{ijk}$ and the distance of the street$_{ijk}$ where street$_{ijk}$ is the kth unsafe street$_{ij}$ for a route$_i$. Only the streets with safety values as “No” will evaluated.

[1] Calculation of risk weight of street$_{ij}$:

$$Risk\_weight\_of\_street_{ij} = \sum_{n=1}^{Total} Count_{ijn} * Crime\_factor_{ijn}$$

Where,

Total = the number of types of crimes

Count = the count of criminal incidents for crime$_n$ for street$_{ij}$

Crime_factor = the intensity of the crime

[2] Calculation of risk of a route$_i$:

$$Risk\ of\ route_i = \sum_{k=1}^{tu} Risk\_wieght\_of\_street_{ijk} * distance_{ijk}$$

Where,

tu = total number of unsafe streets on that route$_i$.

The route with the minimum amount of risk will be the safest route which will be given as output to user. The safest route is plotted using GPS services. The user will have the option to share the safest route on social media to let well-wishers know about his route.

IV. IMPLEMENTATION

In paper, only those crimes are considered which can affect the security of a person while travelling from one place to another. Crimes such as felony or cybercrime is considered as irrelevant for our analysis.

The logical work flow map is as shown in Fig. 2. At the beginning, route information will be acquired by processing the input provided by the user. Route information consist of streets in all the routes. User specific dataset that means dataset relevant for that particular user will be extracted from the original dataset. Street profile for all the possible routes will be generated which consist of street name, category of offence, number of occurrences of that offence, and time of occurrence. Then the street profile will be given to the data mining engine. Data mining engine comprises of two parts:

1) ID3 decision tree classifier:

ID3 classifier will check the safety of all the streets in a route. Similarly, all the routes and their integral streets will be analyzed. With the help of this classifier the street profile will be considered with the safety value “Yes” or “No”.

2) Determination of the safest route:

Only those streets having value “No” will be analyzed further. The risk value of the route will be calculated by using and the risk weight which is associated with these unsafe streets. The route with the minimum risk value will be given as the safest route.

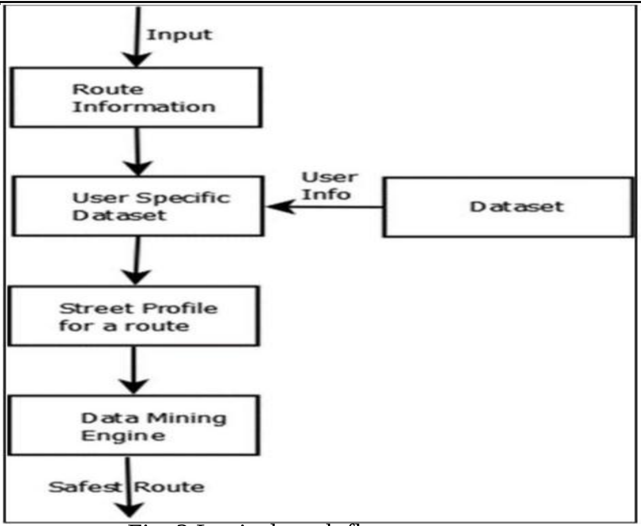

Fig. 2 Logical work flow map

V. ID3 DECISION TREE ALGORITHM:

Iterative Dichotomiser 3 or ID3 is an algorithm which is used to generate decision tree. For training the dataset a table is created in which the rows denotes the street id and the columns denote the categories of the crime. The working of ID3 is based on the principle of attribute selection. The two mathematical terms which lead to attribute selection are Entropy and Information Gain. These two terms are the deciding factor for selecting the node of the decision tree. Here categories of the crime play the role of attributes. All the records in the table are referred as the Collection(c).

Calculation of Entropy: Entropy is defined as the average amount of the information contained in the event, sample or record.

Let S be a set of instance, and let p_i be the fraction of instances in S with the output value i.

$$\text{Entropy}(S) = -\sum p_i * \log_2(p_i)$$

Calculation of Information Gain: The entropy changes when we use a node in a decision tree to partition the training instances into smaller subsets. Information gain is a measure of this change in entropy.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum (|S_v|/S) * \text{Entropy}(S_v)$$

Where,

S is Entropy.

A is the attribute for which gain will be calculated. v is all the possible of the attribute A

S_v is the number of elements for each v for instance $\sum$ is the summation of $((|S_v|/|S|) * \text{Entropy}(S_v))$ for all the items from the set of v.

```
After Decision tree:
Route, Street Name, Distance, Gender, Time, Assault, Driving Under Influence, Kidnapping, Robbery, Sex Offences, Vehicle Theft, Safe?
['1', '10TH AV', '463', 'MALE', 'DUSK', 'HIGH', 'LOW', 'HIGH', 'HIGH', 'LOW', 'LOW', 'YES']
['1', '12TH ST', '333', 'MALE', 'DUSK', 'LOW', 'HIGH', 'HIGH', 'HIGH', 'LOW', 'HIGH', 'NO']
['1', 'ARLINGTON ST', '200', 'MALE', 'DUSK', 'HIGH', 'LOW', 'HIGH', 'LOW', 'LOW', 'HIGH', 'YES']
['1', 'BRITTON ST', '300', 'MALE', 'DUSK', 'HIGH', 'HIGH', 'LOW', 'LOW', 'LOW', 'LOW', 'YES']
['2', 'CARDENAS AV', '756', 'MALE', 'DUSK', 'LOW', 'HIGH', 'HIGH', 'LOW', 'LOW', 'LOW', 'YES']
['2', 'CARNELIAN WY', '845', 'MALE', 'DUSK', 'LOW', 'LOW', 'LOW', 'LOW', 'LOW', 'LOW', 'YES']
['2', 'DELMAR ST', '900', 'MALE', 'DUSK', 'HIGH', 'HIGH', 'LOW', 'HIGH', 'LOW', 'HIGH', 'NO']
['2', 'ELLIS ST', '333', 'MALE', 'DUSK', 'LOW', 'HIGH', 'HIGH', 'HIGH', 'LOW', 'HIGH', 'NO']
['3', 'HIGHLAND AV', '489', 'MALE', 'DUSK', 'HIGH', 'LOW', 'LOW', 'HIGH', 'LOW', 'HIGH', 'YES']
['3', 'HOFF ST', '555', 'MALE', 'DUSK', 'LOW', 'LOW', 'HIGH', 'HIGH', 'LOW', 'HIGH', 'YES']
['3', 'KAREN CT', '100', 'MALE', 'DUSK', 'LOW', 'LOW', 'LOW', 'LOW', 'LOW', 'LOW', 'YES']
['3', 'LIBERTY ST', '899', 'MALE', 'DUSK', 'HIGH', 'HIGH', 'LOW', 'LOW', 'LOW', 'LOW', 'YES']
```

Fig. 3 Safety values for street after Decision Tree

VI. RESULTS

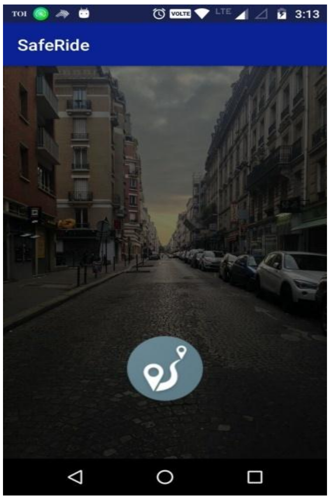

Fig. 4: Welcome Page

The first page for Android App that appears after the App is successfully running on the target device. The Home page is the next page after Welcome Page in the Android App. This page has 3 options for the user of the App. They are:

Find Route: Clicking on this button takes the user to the input page. It also takes the input source and destination by directly tapping on the desired location in the map.

Safest Route: This option gives the safest route from all possible route by sending the request to the user.

Social Support: Social Support option enables user to add and edit the contacts. On long press, the message will be send to added contacts about current location and route on which user is travelling currently.

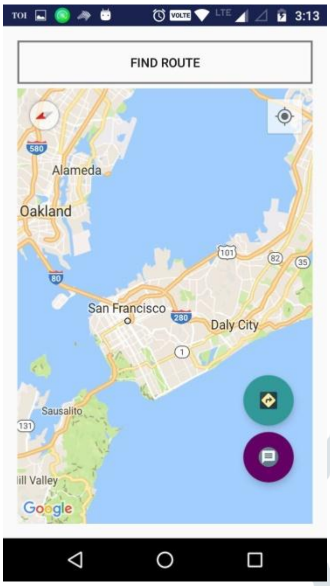

Fig. 5: Home Page

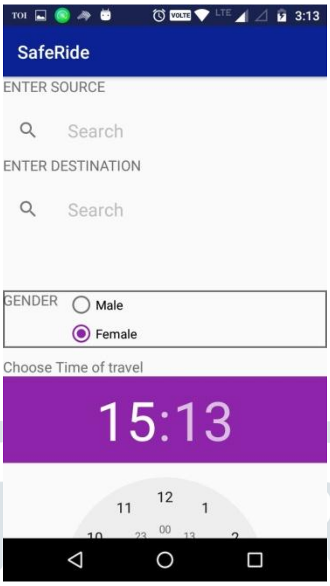

Fig. 5: User Input

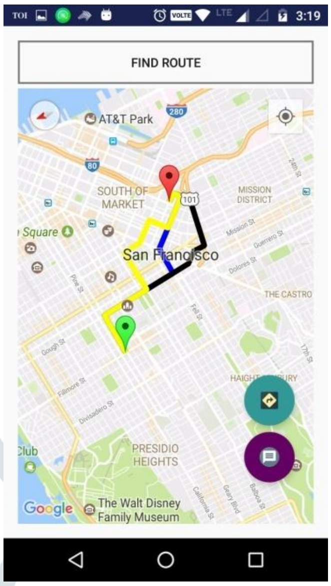

Fig. 6: Possible Routes

The User Input Page takes the input as source location ,destination locatio and some optional input such as gender and time of travel. The App will display all the possible routes for given source and destination after pressing Find Route button. On pressing Safest Route button the request is send to server and safest route among all possible route will be displayed with green colour.

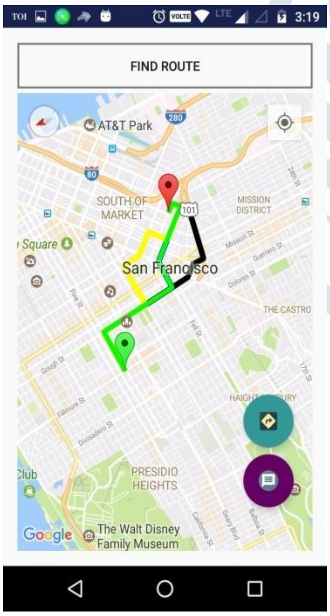

Fig. 6: Safest Route

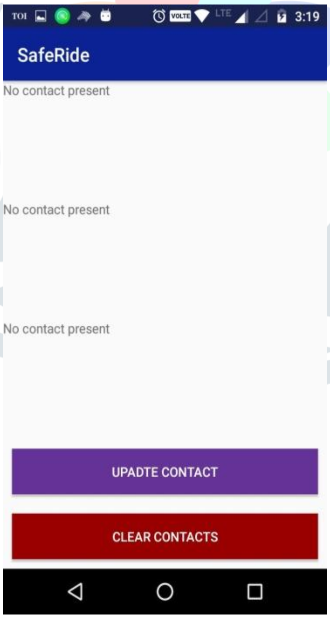

Fig. 7: Social Support

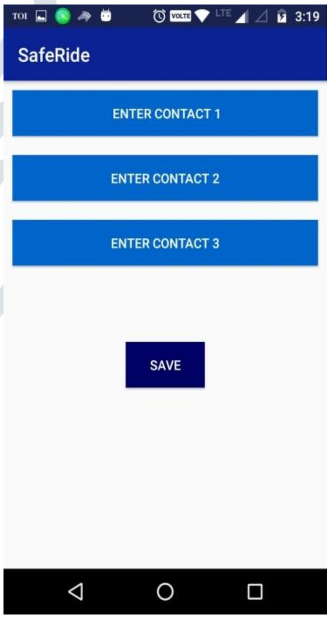

Fig. 8: Edit Contacts

After clicking on Social Support option the options displayed for updating or editing contacts .It provides the facility to add 3 contacts to whom the messages will be send on long pressing the Social Support button.

VII. CONCLUSION AND FUTURE SCOPE

An android application , which will determine the safest route from all the possible routes for given source to destination based on ID3 decision tree algorithm, has been presented. The safety of the user can be assured if this application is used. The application reduces the risk involved while travelling through unsafe area. Social support through social media or messaging will also be provided so that close ones are aware about where the particular user is travelling currently.

In future scope, the application can be used for area wise crime analysis and to reinforce the security in less secure and risk prone areas. In another perspective, it can also be used for criminal profiling. For this, repetitive offenders can be mapped to their corresponding categories of crimes or their areas of activity to help solve future cases. In addition, safer areas from real estate perspective can be determined. Also, verified users can travel together for common destinations.

REFERENCES

[1] Sarang Tarlekar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1536-1540

[2] C.-H. Yu, M. W. Ward, M. Morabito, and W. Ding, "Crime Forecasting Using Data Mining Techniques," 2011 IEEE 11th Int. Conf. Data Min. Work., pp. 779–786, 2011.

[3] TongWang1, Cynthia Rudin1, DanielWagner, and Rich Sevieri. Learning to detect patterns of crime. JECET, 1:124{131, 2012.

[4] J. Han, M. Kamber, and A. K. H. Tung, "Spatial Clustering Methods in Data Mining: A Survey," Geogr. Data Min. Knowl. Discov. vol. 2, 2001.

[5] S. Deerwester, S. Dumais, T. Landauer, G. Furnas, and R. Harshman, "Indexing by Latent Semantic Analysis," J. Am. Soc. Inf. Sci., vol. 41, no. 6, pp. 391–407, 1990.

[6] J. R. Quinlan, "Induction of Decision Trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.

[7] Ruijuan Hu. Data mining in the application of criminal cases based on decision tree. International Journal of Engineering Sciences, 2:24–27, 2013.

[8] Chung-Hsien Yu, Max W. Ward, Melissa Morabito, and Wei Ding. Crime forecasting using data mining techniques. IEEE 11th International Conference on Data Mining Workshops, pages 779{786, 2011.