Route-The Safe: A Robust Model for Safest Route Prediction Using Crime and Accidental Data

Article · December 2019

CITATIONS
6

READS
983

3 authors, including:

Venkatesh Gauri Shankar
Manipal University Jaipur
19 PUBLICATIONS **88** CITATIONS
SEE PROFILE

Chaurasia Sandeep
Manipal University Jaipur
29 PUBLICATIONS **78** CITATIONS
SEE PROFILE

International Journal of Advanced Science and Technology
Vol. 28, No. 16, (2019), pp. 1415 - 1428

Route-The Safe: A Robust Model for Safest Route Prediction Using Crime and Accidental Data

Shivangi Soni[1, *], Venkatesh Gauri Shankar[2] and Sandeep Chaurasia[3]

[1, 2, 3] *School of Computing and IT, Manipal University Jaipur, India*
[1]*shivangisoni98@gmail.com*

Abstract

Crimes are rising day by day & thus safety & security is becoming a major concern for people today. Even while travelling, people should be aware & choose the route which is safest to travel from. People who are new to the city, have no idea about the safe routes. Though people rely on google maps for planning their routes; yet it only provides the shortest path & give no consideration for safety of the path. Although several other route planning apps exist which provide the safest route, but these do not consider all the factors that account for safety of the path. Apart from other navigation apps, this paper describes an innovative method to find safest route having lowest risk score. This paper uses updated crime and accident data available on NYC OpenData to determine average risk score of clusters/regions. Machine learning algorithms are used to generate the risk score of a path based upon average risk score of nearby clusters/regions. Also, one can get better results by increasing the number of factors that affect the safety of the path. In future, a better prediction algorithm can be introduced through which traveler can identify probable crimes which he/she might face while travelling on a specific route.

Keywords: *KNN K nearest Neighbor, Semantic Classification, Crime Data, Accidental Data, Clustering, Data Analytics, Machine Learning;*

1. Introduction

Safety & security became the top most priority of people due to rising number of crimes in cities. Even using google maps while travelling, may lead to life threatening & hazardous situations. Many women take different routes from those recommended by google maps & other similar apps due to the safety concerns. People who are local to the city might know which routes are safe to travel but people who are tourists or new in the city rely on their drivers or invest a lot of time in research about the safety of the area. To deal with such growing issue related to safety, people need better & efficient solutions. Need of a solution that suggests safe route is becoming much more important than it was ever before. With the help of such solution, people will feel safer than before while travelling.

Local government authorities are constantly collecting data related to crime offenses, accidents, routes etc. these datasets are constantly updated & maintained by the authorities & these are a great source of information.

In this paper, New York city is chosen being a huge metropolitan city. [1] According to crime data between Year 2010-2019, the crime rate in NYC is 6% higher than the average of the whole of the state of New York, but it is 28% lower than national average. When looking at violent crimes, NYC has 51% higher violent crime rates than those of state of New York average, while it is 41% higher than the national average. The overall score of crimes calculated for NYC with respect to national average are: - violent crimes 65% and property crimes 35%. In this paper arrest and accident datasets from NYC OpenData are used [2] which is maintained and updated by NYC agencies and other partners. Data is updated several times a day which help in making better predictions.

ISSN: 2005-4238 IJAST

The proposed solution for this ever–increasing problem focuses on - First: Predicting a safe route using crime and accidents data as well as considering distance between the source and destination. Second: Dividing NYC into smaller regions of risk by applying nested clustering on the data. It gives better predictions as it takes into consideration smaller crime areas also. Third: Calculating risk score of the routes based on the risk score of the nearby clusters.

Already existing work focuses on finding the safe route in populated cities but have the following disadvantages: uses only crime data that is highly subjective in nature, the data is not updated, and the approaches used ignore the smaller crime areas.

Proposed solution in nutshell: This solution focuses on predicting safest route by calculating risk score of all the routes lying between source and destination. The risk score of the paths are based on the average risk score of the nearby clusters formed using the datasets. The route with the lowest risk score is then suggested as the safest route.

Roadmap: Following is the categorization of the remaining paper. Study of related work is given in section 2. It includes information about single and multi- preference routes. Section 3 shows algorithm and summary of the model. Section 4 shows the methodology used to build this model. Result of the proposed model is explained in section 5. Conclusion of the model is done in section 6.

2. LITERATURE REVIEW

2.1. Safe Route

There are applications that find the safest path by creating a balance between safety and distance. Safepath is one such application. [13]. It uses a crime density map to assign risks to routes and then suggest routes that vary from shortest distance to safety. Another safe route application that was developed for Mexico City uses social crime reports and tweets to do the classification and geocoding of crimes with the help of a Naive Bayes classifier [14]. Its main aim is to find the safest route without considering the geographical distance. Similarly, SocRoutes uses geocoded tweets to suggest safe routes to the users and routes users outside an unsafe region. [21].

Sentiment analysis is applied on the tweets and then the regions are categorized into safe and unsafe depending on the results of sentiment analysis. For suggesting the safest route, first the shortest path is found out and if the route passes through any region that is unsafe it shifts the points out of that unsafe region [18][19].

In comparison to all these mentioned approaches which focus on crimes on a larger and greater scale, this proposed solution focuses on crimes at smaller level by using nested clustering. Work done previously on this topic either neglect crime areas that are smaller in size or use data which is highly subjective in nature.

2.2. Multi-Preference Routes

The routing algorithms that focuses on more than one objective in determining the safest route, such as this proposed model. An application that considers both distance and either quietness, happiness, or beauty. It first generates the k list of paths by ranking them based on shortest paths and then re- ranking the k list according to the second objective [22]. It may be hard and difficult to optimize based on second objective unless the list of k routes is diverse. Another web application is Be-safe travel, it provides the safest route considering the shortest path and security level. [7] However, it directly counts the number of crime points in a path without considering the degree of crime. Waze is another such app that uses various factors like traffic, crime etc. One of the limitations of this app is that it can find safe routes only over 1,000 miles (1600 km). As given on waze website another model includes traffic and distance as factors [5]. For example, the PreGo

model suggests routes which includes preferences that the users focus on such as conditions of roads, scenery and time [8]. Even if this model can suggest route based on factors related to risk, it does not include the distances from risk near edges due to this there is not enough information regarding the street safety. Another system is T-Drive that modify routes based on distance and time using GPS taxi information [9]. However, in this process there is no new way to take crime information into consideration. ARSC algorithm is applied in [10] for generating paths by using graph search and then it selects the best one first. The models listed above consider many factors to generate the safest route but most of these models ignore information about crime when generating the route. This can lead to generation of unsafe paths.

3. Proposed Model

3.1. Summary

Data pre-processing was performed on the data to give better and accurate results. In data pre-processing the following steps were performed: the crime dataset and accident dataset were imported from the Nypd website [2] , from both the datasets the columns that were not required were dropped , rows from both the datasets which had nan or null values were dropped, the outliers were analyzed and dropped, columns were renamed for better understanding, created one new column in the accident dataset named as (A_S) as shown in Algo 1, dropped the offenses from the crime dataset which were of no use in the model , created one new column in the crime dataset named as (C_S) as shown in Algo 1 , assigned the crime scores between 1 to 15 [6] , created a new column in both the datasets named as 'Index' , assigned 'a' to all the rows of accident dataset and 'c' to all the rows of crime dataset , made two new datasets consisting of only Manhattan Brough from accident and crime dataset , concatenated the crime and accident datasets of Manhattan.

Now the next step is the model design: nested K Means clustering was done based on latitude and longitude. K Means clustering performed for the first time: K Means clustering was done after determining the number of clusters to be formed using the elbow method. K Means clustering performed for the second time: K Means clustering was done in one of the already formed clusters from the previous clustering after determining the number of clusters to be formed using the elbow method, centroids of the newly formed clusters were found using K Means clustering [20].

Some of the other steps that were done in the model design part : created one more column in dataset named as 'cluster', assigned the cluster number to which the data belongs to , separated the data again to accident dataset and crime dataset based on index , splitted the crime dataset and accident dataset in 70:30 for training the model , for both accident and crime dataset which point belongs to which cluster was found out and this was done for both training and testing data , in the testing and training datasets formed from the crime dataset (C) as shown in Algo 2 was calculated for all the clusters formed, in the testing and training dataset formed from the accident dataset (A) as shown in Algo 2 was calculated and this was done for all the clusters formed, for both training and testing datasets of crime/accident dataset new dataframes were created consisting of cluster number, centroid (latitude and longitude) of the clusters and 'C'/'A'.

Next step in model design is K Nearest Neighbor Regressor. For crime dataset and accident dataset : the value of 'k' was found using the rmse value, centroid (latitude and longitude) were used for training K Nearest Neighbor Regressor and testing crime value, R2 score was determined , new functions 'find_crime'/ 'find_accident' were created to predict the crime/accident score of a point using the KNN Regressor model.

Last step in model design part is finding the path using google maps directions function, 'chunk_user_route', 'interpolate_points (route_line, line_points, distance)', 'total_score ' and 'min' functions were created.

ISSN: 2005-4238 IJAST

The next part of this model is Evaluation and Assessment: two times K Means algorithm elbow method was used, the rmse value was used to find the 'k' in K Nearest Neighbor Regressor, the R2 score for evaluation of K Nearest Neighbor Regressor was obtained.

The last part of this model is Implementation: route suggested by google map and the route suggested by the model explained in this paper were plotted for comparison and to find the safest route.

Algorithm 1: Data preprocessing Mask Algorithm

1. Arrest data (crime data): AD1 → nypd
2. Accident data: AD2 →nypd
3. RM (law code, jurisdiction code) → nypd
4. Na.rm→ nypd
5. RM (Outlier: OL) → nypd
6. Column: rename→ nypd
7. 'Accident Score of a point'(A_S) →
   a. count of pedestrians injured - P_i
   b. count of pedestrians killed - P_k
   c. count of cyclist injured - C_I
   d. count of cyclist killed - C_k
   e. count of motorist injured - M_i
   f. count of motorist killed - M_k
   g. $A_S = P_i + P_k * 2 + C_I + C_k * 2 + M_I + M_k * 2$
8. RM(Offences)→ nypd
9. 'Crime Score of a point'(C_S) → 1 to 15
10. Dist_manhattan→ (AD1, AD2)

Model Design Mask Algorithm

Algorithm 2: K-Mean

1. K-Mean (latitude, longitude) → Nested (K-Means)
2. Elbow (Clusters) → # of clusters
3. Nested_K Means→ store
4. Centroid (clusters) →Nested (K-Means)
5. Index→ (AD1, AD2)
6. Cross_validation_train→70:30
7. 'Average crime score of a cluster' (C) →Store
8. 'Average accident score of a cluster ' (A) →Store
9. $C = \Sigma^n_0 C_S / n$
10. $A = \Sigma^n_0 A_S / n$
11. Consisting of cluster number →define
12. Store→ (C, A)

Algorithm 3: K Nearest Neighbor

1. Val(rmse) → K
2. K-NN (latitude, longitude) → Training
3. R_2→Store
4. Start→G_M (Google map)
5. Score→ min (total score)
6. Store→ (rmse, R_2)

ISSN: 2005-4238 IJAST

4. Methodology

Different steps involved in computing the score of regions of interest and predicting the safest route are represented in Figure 1. and Figure 2. respectively.

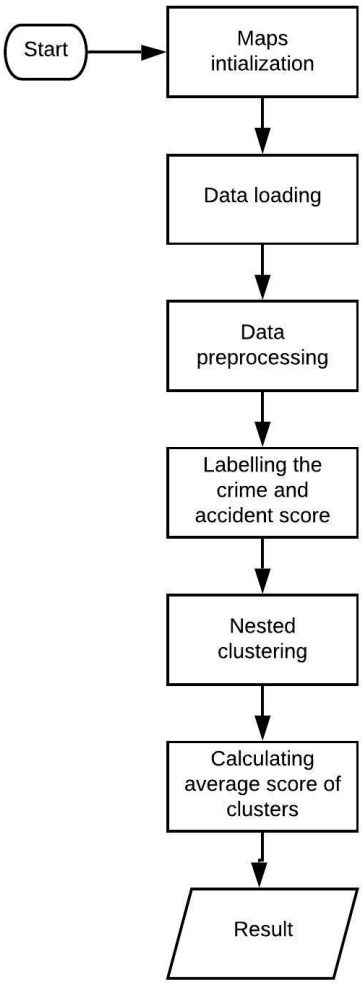

Figure 1. Precomputed Score of Regions of Interest

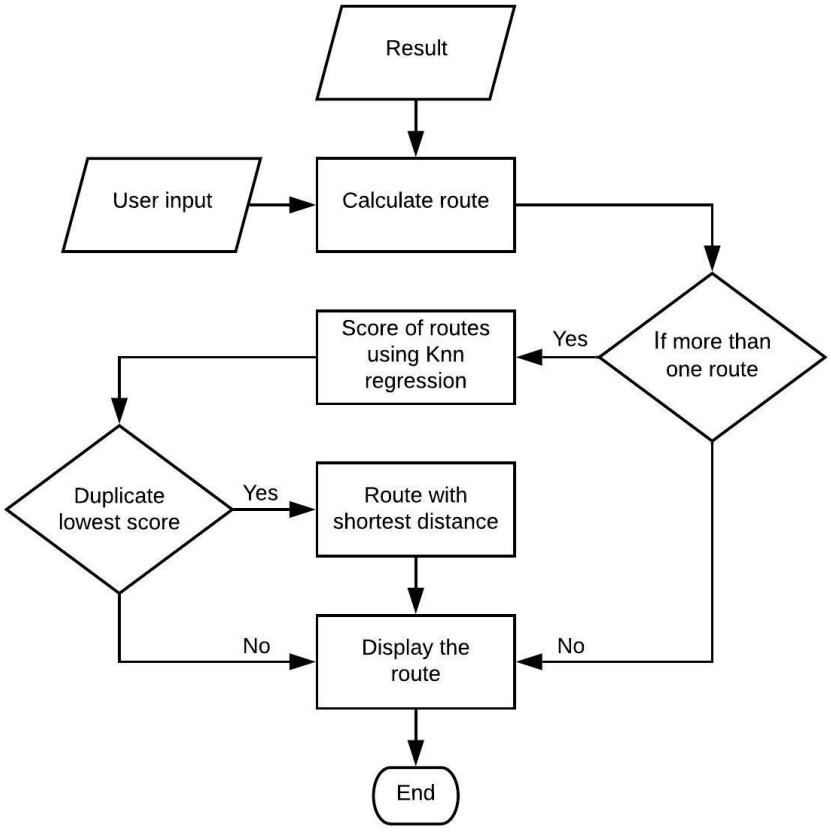

Figure 2. Predicting the Safest Route

4.1. Data

The datasets used in this solution are accident dataset Table 1. and arrest dataset Table 2. which are taken from NYC OpenData [2].

There are eleven attributes in the accident dataset Table 1., borough is the name of the borough in NYC where the accident occurred, latitude and longitude are the coordinates of the exact place where the accident occurred, count of persons injured (I) and count of persons killed (K) is the total count of persons injured and total count of persons killed in an accident as shown in (1) and (2) respectively, count of pedestrians injured (P_I) and count of pedestrians killed (P_K) is the total count of pedestrians injured and killed in an accident, count of cyclist injured (C_I) and count of cyclist killed (C_K) is the total count of cyclist injured and killed in an accident, and count of motorist injured (M_I) and count of motorist killed (M_K) is the total count of motorists injured and killed in an accident.

$$I = P_I + C_I + M_I \quad (1)$$

$$K = P_K + C_K + M_K \quad (2)$$

There are four attributes in the arrest dataset Table 2. borough is the name of the borough in NYC where the crime occurred, latitude & longitude are the coordinates of the exact place where the crime occurred and ofns_desc is the kind of offense happened at the location. In this proposed solution datasets of NYC are used.

Table 1. Accident Dataset

Borough	Q
Latitude	40.739674
Longitude	-73.990950
Count of persons injured	1
Count of persons killed	1
Count of pedestrians injured	0
Count of pedestrians killed	1
Count of cyclists injured	1
Count of cyclists killed	0
Count of motorists injured	0
Count of motorists killed	0

Table 2. Arrest Dataset

Borough	B
Latitude	40.830799
Longitude	-73.82949
Ofns_desc	Rape

4.2. Initializing Google Maps API

The next step of this model is to initialize google maps API. After initializing the google maps API & configuring its gmaps function, allotted API is used to perform various operations.

4.3. Data Loading

The next step involves the loading of datasets. The arrest dataset and the accident dataset are loaded from the NYC OpenData website [2]. The datasets are updated several times a day. The arrest dataset gives complete information about the type of crime offense that took place at a latitude and longitude and similarly the accident dataset gives complete information about the accident that took place at a latitude and longitude [17].

4.4. Data Preprocessing

One of the main steps of the model is the data pre- processing step. [11] This step is very important from the perspective of machine learning as data pre-processing takes 60 to 80 percent of the whole analytical pipeline in a typical machine learning project. Also, the information derived from the data influences the model's ability to learn. In proposed solution data pre- processing is performed to remove missing data & data outliers thus proper clusters can be formed after clustering [16].

4.5. Labelling Score

The next step is the labelling of accident score of a point (A_S) & crime score of a point (C_S). Thorough care is taken while assigning weights (C_S) to different kind of crime offenses. Assigning weights (C_S) to different crime offenses is an assumption based on the class the offense belongs to and the type of punishment or sentence of imprisonment given to the suspect [6]. Arrest dataset is used in this paper to find out which type of crime occurred at which location. Table 3. Shows the weights assigned to the different crime offenses. Accident Score of a point (A_S) is calculated as shown in (3). Weights assigned in this equation were an assumption made with thorough analysis.

$$A_S = P_K *2 + C_K *2 + M_K*2 + P_I + C_I + M_I \quad (3)$$

Table 3. Crime Score of a Point (C_S)

Offense	Crime score of a point (C_S)
Rape	15
Sex crime	15
Homicide-negligent-vehicle	14
Murder & non-negl. Manslaughter	14
Homicide-negligent, unclassified	14
Offenses related to children	13
Offenses against the person	13
Offenses against public safety	13
Anticipatory offenses	13
Assault 3 & related offenses	12
Kidnapping	11
Kidnapping & related offenses	11
Disorderly conduct	10
Harassment 2	10
Arson	9
Criminal mischief & related offense	8
Dangerous weapons	7
Felony assault	6
Escape 3	5
Robbery	5

ISSN: 2005-4238 IJAST

Intoxicated & impaired driving	4
Vehicle and traffic laws	3
Moving and other traffic infraction	3
Fraudulent accosting	2
Petit larceny	1
Grand larceny	1

4.6. Nested Clustering

The next step is the formation of the clusters using nested clustering based on latitude and longitude. The arrest dataset and the accident dataset are concatenated for clustering, so that both the factors can be taken into consideration for determining the risky regions. This process includes two subprocesses: First: Clustering is done based on the latitude and longitude of the locations where the crime and accidents occurred, so that we can divide the NYC into nearby smaller regions of risk. For the initial phase of the project, this paper focuses on Manhattan Borough only instead of whole NYC. Second: To generate more accurate and meaningful regions of risk, clustering is applied in the already formed clusters in the Manhattan Borough. Clustering again is done based on latitude and longitude of the locations where the crime and accidents occurred. Initially clustering is applied only to one of the formed clusters of Manhattan Borough. K Means clustering is used for clustering the points. [12] K Means is a clustering algorithm under unsupervised machine learning. Clustering of this type splits the information/data into clusters where points that lie inside one cluster are of similar nature. K Means is used as it is simple, flexible and works well in case of larger dataset. K is decided based on the lowest root mean square error value. Root mean square error values of k ranging from 1 to 20 were found as generally the range of k varies from 1 to 30.

4.7. Average Cluster Score

Again, after dividing one of the formed clusters in Manhattan Borough to smaller regions/clusters, the average crime score and accident score of all the formed clusters/ regions are calculated. Average accident score of a cluster (A) and crime score of a cluster (C) are calculated as shown in (4) and (5) respectively.

Number of points in a cluster = N

$$C = \sum_{i=1}^{N} C_S / N \quad (4)$$

$$A = \sum_{i=1}^{N} A_S / N \quad (5)$$

4.8. Risk Score (R_S)

After this, input of source and destination are provided for predicting safest route. Now as per google maps the shortest route is displayed using plugin gmaps. DirectionService [3] is a class of google maps API which is used to give directions between source and destination point by displaying all the possible routes. DirectionService produces direction's route most atomic unit which are the steps, which contains a single step that describes a specific, single instruction on the route. E.g. "Turn right at 4th avenue. Now

there are two conditions possible. First: If only one route is possible between the source and destination point then that route is displayed as the safest route. Second: Else if more than one route is possible, risk score must be calculated for all the possible routes. Again, this process includes four subprocesses. First: Through a function, location of the source point and the end location of the all remaining steps are stored. Second: With the help of these stored locations waypoints of the routes are generated at 2 Km apart. Third: K Nearest Neighbor Regressor [4] is used to predict the accident and crime score for all the generated waypoints based on the average accident and crime score of k nearest clusters. The clusters that are near to the waypoint are determined by the distance between the centroid of the clusters and the waypoint. KNN Regressor is used in this model because generally the degree of risk at a location is decided by the degree of risk at the nearby points or areas. K is decided based on the lowest root mean square error value. Root mean square error values of k ranging from 1 to 20 were found as generally the range of k varies from 1 to 30. Fourth: The risk score (R_S) for all the possible routes is calculated as shown in (6).

Number of waypoints in a route = W

$$R_S = \sum_{i=1}^{W} C_S + \sum_{i=1}^{W} A_S \quad (6)$$

4.9. Safest Route

The final step of this model is to display the safest route. This step involves two conditions: First: If there is only one route that has the lowest risk score then that route is displayed as the safest route. Second: Else if more than one route has the lowest risk score then the distances of the following routes are compared & the route which has the shortest distance is displayed as the safest route.

5. Results

The result of K Means clustering in whole NYC is done based on latitude and longitude as shown in Figure 3.

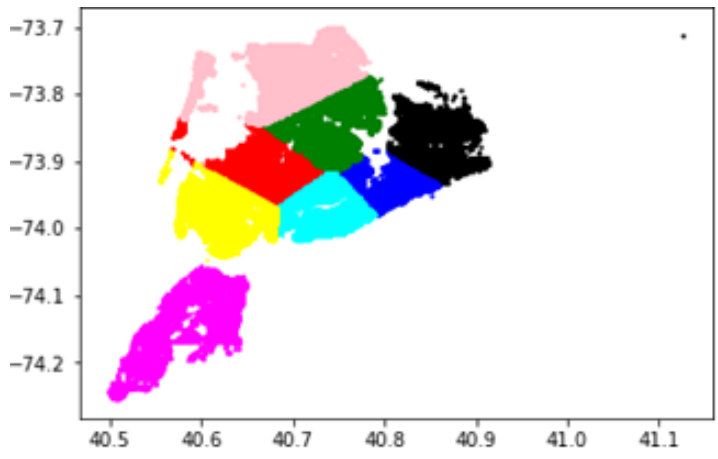

Figure 3. Clustering in the NYC

The result of K Means clustering done in Manhattan Borough based on latitude and longitude is as depicted in Figure 4. Again, K Means clustering is done on the already formed clusters based on latitude and longitude, the result of which is depicted in Figure 5.

The KNN Regressors that predicted the accident and crime score is analyzed using the R2 score. R-Squared also called as coefficient of determination statistically computes how close is the data from the fitted regression line. So, if the value of R-Squared is high the model will fit your data better [15].

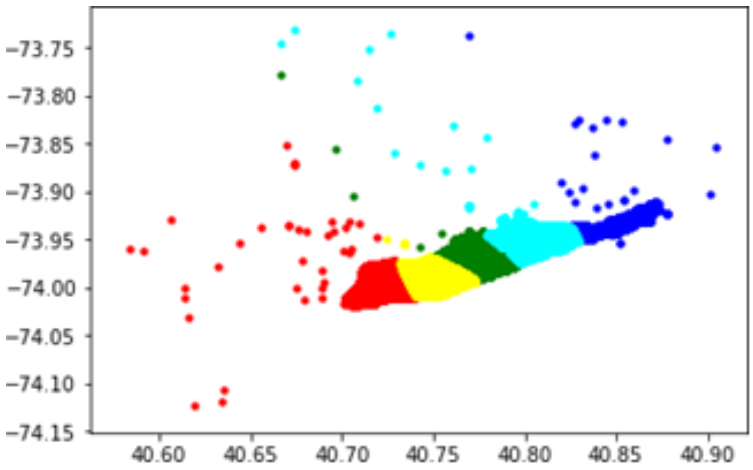

Figure 4. Clustering in the Manhattan Borough

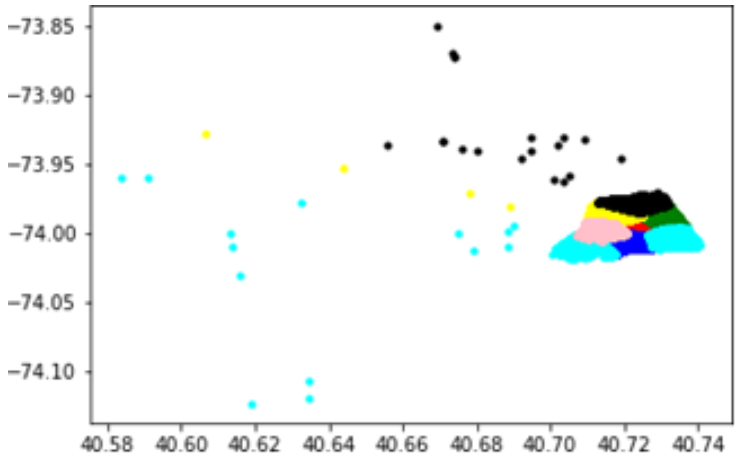

Figure 5. Clustering Applied to One of the Formed Clusters of Manhattan Borough

R2 scores are as shown in Table 4.

Table 4. R_2 Score

Category	R2 Score
Accident Score	0.910
Crime Score	0.974

Taking a case in which more than one route is possible between A and B points. Figure 6. shows the route suggested by google maps when the source A and the destination B are given as input. Google maps suggests the route which has the shortest distance. Figure 7.

Shows the route suggested by this model when the same source A and destination B are given as input. As explained in the methodology for all the routes between A and B, waypoints are determined at 2 Km apart. The accident and crime score for all the waypoints are predicted using the KNN Regressor based on the average accident and crime score of the nearby clusters. At last the risk score for all the routes is calculated using the formula stated in the methodology. The model suggests the safest route by selecting the route which has the lowest risk score. If more than one route has the lowest risk score it suggest the route which has the shortest distance.

For validating the results, one can manually check how many clusters are nearby and determine approximate score.

Figure 6. Route suggested by google maps

Figure 7. Route suggested through proposed model

6. Conclusion

In this paper, a solution is proposed that suggests people the path that is safest to travel from source to destination. Google API technology combined with machine learning models namely K Means clustering algorithm and KNN Regressor algorithm are used in this proposed solution. Arrest and accident datasets of NYC are used to predict the safest route. Google maps suggestion of the routes are purely based on the shortest distance, on

the other hand this solution is modified one & recommends routes which are safe to travel from. Safe path in this solution means route with lowest risk score & this is calculated based on accidents and crimes that happened on that route or in nearby regions. This solution is very useful especially for the people who are new to the city or are tourists. Future work that could be done on this proposed solution are. First: Adding more factors which determines the safety of the route. Second: Developing an android based version for this proposed solution. Third: Adding more features to make user experience amazing.

References

1. https://www.fbi.gov/services/cjisucr
2. https://opendata.cityofnewyork.us
3. https://developers.google.com/maps/documentation/javascript/directions
4. https://www.sciencedirect.com/science/article/abs/pii/S0167865513004 145
5. https://www.waze.com/en-GB/
6. http://ypdcrime.com/
7. Amalia Utamima1, a) and Arif Djunaidy. Be-Safe Travel, a web-based geographic application to explore safe-route in an area. AIP Conference Proceedings 1867, 020023 (2017)
8. Abdeltawab M Hendawi, Aqeel Rustum, Dev Oliver, David Hazel, Ankur Teredesai, and Mohamed Ali. 2015. Multi-preference Time Dependent Routing. Technical Report UWT-CDS-TR-2015–03-01, Center for Data Science, Institute of Technology, University of Washington, Tacoma, Washington, USA (2015).
9. Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. 2010. T-drive: Driving Directions Based on Taxi Trajectories. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10). ACM, New York, NY, USA, 99–108
10. Hans-Peter Kriegel, MaŠhias Renz, and MaŠhias Schubert. 2010. Route skyline queries: A multi-preference path planning approach. In Data Engineering (ICDE), 2010 IEEE 26th International Conference on. IEEE, 261–272.
11. Pyle, D., 1999. Data Preparation for Data Mining. Morgan Kaufmann Publishers, Los Altos,California.
12. J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297
13. Esther Galbrun, Konstantinos Pelechrinis, and Evimaria Terzi. 2016. Urban Navigation Beyond Shortest Route. Inf. Syst. 57, C (April 2016), 160–171
14. Felix Mata, Miguel Torres-Ruiz, and Giovanni Guzman. 2016. A Mobile Information System Based on Crowd-[Sensed and O• cial Crime Data for Finding Safe Routes: A Case Study of Mexico City. Mobile Information Systems 2016, Article 8068209 (2016).
15. Wright, Sewall. (1921): Correlation and causation. Journal of Agricultural Research 20: 557-585.
16. Shankar, V. G., Jangid, M., Devi, B., Kabra, S. (2018). Mobile big data: Malware and its analysis. In Proceedings of First International Conference on Smart System, Innovations and Computing. Smart Innovation, Systems and Technologies (Vol. 79, pp. 831–842). Singapore: Springer. https://doi.org/10.1007/978-981-10-5828-8_79.
17. Shankar, V. G., Devi, B., & Srivastava, S. DataSpeak: Data extraction, aggregation, and classification using big data novel algorithm. In B. Iyer, S. Nalbalwar, & N. Pathak (Eds.), Computing, communication and signal processing. Advances in intelligent systems and computing (Vol. 810). Singapore: Springer. https://doi.org/10.1007/978-981-13-1513-8_16.
18. Devi, B., Kumar, S., & Anuradha, S. V. G. (2019). AnaData: A novel approach for data analytics using random forest tree and SVM. In B. Iyer, S. Nalbalwar, & N. Pathak (Eds.), Computing, communication and signal processing. Advances in intelligent systems and computing (Vol. 810). Singapore: Springer. https://doi.org/10.1007/978-981-13-1513-8_53.
19. Devi B., Shankar V.G., Srivastava S., Srivastava D.K. (2020) AnaBus: A Proposed Sampling Retrieval Model for Business and Historical Data Analytics. In: Sharma N., Chakrabarti A., Balas V. (eds) Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing, vol 1016. Springer, Singapore. https://doi.org/10.1007/978-981-13-9364-8_14.
20. Goel V., Jangir V., Shankar V.G. (2020) DataCan: Robust Approach for Genome Cancer Data Analysis. In: Sharma N., Chakrabarti A., Balas V. (eds) Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing, vol 1016. Springer, Singapore. https://doi.org/10.1007/978-981-13-9364-8_12.
21. Jaewoo Kim, Meeyoung Cha, and Œomas Sandholm. 2014. SocRoutes: safe routes based on tweet sentiments. In Proceedings of the 23rd International Conference on World Wide Web. ACM, 179–182.

22. Daniele ‹ercia, Rossano Schifanella, and Luca Maria Aiello. 2014. Œe shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In Proceedings of the 25th ACM conference on Hypertext and social media. ACM, 116–125.

Authors

Shivangi Soni is currently a Scholar in Computer Science and Engineering, School of computing and IT at Manipal University Jaipur, India. She has participated in many coding and data science events in national level. She has done many data science experiments over many data repository like Kaggle data set. Her research fields are centric to Machine Learning and Data Science. She has well professional knowledge over Jupyter Notebook, Python and R.

Venkatesh Gauri Shankar is working as Assistant Professor in Department of Information Technology, School of Computing and IT, Manipal University Jaipur, India. He has completed his M Tech. in Computer Science and Engineering from Central University of Rajasthan. He has published many research papers in reputed journals and conferences. He has total of 11+ years teaching experience in academics. He got best paper, best reviewer and best innovator award by many professional bodies. He has also reviewer and member of many reputed journals. He has also served as session chair and advisory board in many national and international conferences. His research and specialized areas are centric to Machine Learning, Deep Learning and Data Science. Currently working in the area of many biomedical and neurodegenerative applications of Deep Learning and Machine Learning.

Dr. Sandeep Chaurasia is working as Associate Professor in the department of CSE, School of Computing & I.T. in Manipal University Jaipur. He completed his PhD (Engineering) in 2014 in the area of Supervised Machine Learning and M. Tech in Computer Science in the year 2009. He has done his B.E. in Computer Engineering in the year 2006. He has more than eleven years of rich experience in academics and one year in industry. He has more than more than 20 publications in International / national journals/conference proceedings. He is a SMIEEE, LMCSI, MACM and member of Machine Intelligence Research labs - USA. He is also member of reviewer board of various journals and technical program committee of several reputed conferences. His research interests include Machine Learning and Soft Computing and other areas of interest are Algorithms, Artificial Intelligence. He is associated with Machine Learning for more than 7 years. Currently working in the area of application of machine / Deep learning in natural language processing like semantic analysis & lexical analysis. Currently he is guiding 4 PhD students in the area of NLP, Intrusion detection, food adulteration using AI techniques. He is also active member of special interest group and initiative by MIR labs to connect the researchers & professional across the globe.

ISSN: 2005-4238 IJAST