Crime Log Analysis And Crime Type Prediction

PIKKU KUMAR SINGH[1], PRAVEEN KUMAR[2], RUCHITHA.S[3], SATYAM MOURYA[4], KANCHAM ARCHITHA[5], MR. MOHAMED SHAKIR[6]

1 Student, Computer Science & Engineering, Presidency University, Bangalore, India
2 Student, Computer Science & Engineering, Presidency University, Bangalore, India
3 Student, Computer Science & Engineering, Presidency University, Bangalore, India
4 Student, Computer Science & Engineering, Presidency University, Bangalore, India
5 Student, Computer Science & Engineering, Presidency University, Bangalore, India
6 Assistant Professor, Computer Science & Engineering, Presidency University, Bangalore, India

ABSTRACT

Crime is among the biggest and most concerning problems of our society and an important task is to prevent it. On daily basis very large number of crimes is being committed at quiet a frequent rate. It is required to keep track of all kind of crimes and maintaining a record of it so that it can be used to prevent crimes in future. The problem which we are currently facing is to maintain a proper record or database of it and to analyze the crime database in order to predict the crimes in future to reduce the crime rate. The main objective of the project is Analyzing the crime dataset which consist of crime log of West Virginia and to predict the type of crime which may happen depending upon various factors. In this project, we will be using machine learning, visualization tools and neural networks for crime type prediction of West Virginia data set.

Keyword: *Crime Prediction1, Visualization2, Neural Network3, Python4 and Machine Learning5.*

1. INTRODUCTION

Crime is one of the major concerns of our society. Crime type analysis and identification are the biggest problems in the police department as there is a lot of crime information available. There is a need for technology where case resolution may be successful.

The purpose of this project is to detect crime and predict the nature of crime using features that are available in the database. Database is issued on official sites. With the help of viewing libraries we can analyze the local crime record to find hidden patterns behind the data and we can predict the type of crime based on a variety of factors.

The purpose would be to present a comprehensive report on West Virginia crime and to predict the type of crime based on the data provided. A comprehensive crime analysis report in West Virginia helps us to understand hidden patterns behind text and numbers. Database observations are made to analyze potential crime in the area. This work helps people better understand where crime is in western Virginia and what the hidden patterns behind it can thus help reduce crime.

2. LITERATURE SURVEY

S. Sathyadevan and S. Gangadharan used the models Apriori Algorithm ,Naïve Bayes, Decision Tree and did crime prediction by data mining and he published this in the year 2014.The merit of this survey was to Predict region having high probability for crime disadvantage was that it can't Predict the time in which crime happens.[1]

Z. Zubi and A. Mahmmud used the models Association rule mining, k-means clustering and did Crime data analysis using data mining and he published this in the year 2014.The merit of this survey was to Analyzing crime patterns become easy due to clusters. Disadvantage was that it Can't predict multi-model data.[2]

Sunil Yadav, Meet Timbadia and RohitVishwakarma used the models k-means clustering, naïve Bayes, Apriori correlation, regression and did Crime pattern prediction and published In the year 2017. The merit of this survey was Prediction based on states and age group with dates. And the demerit was handling insufficient data is difficult.[3]

Jyoti Agarawal and Renuka Nagpal used the models k-means algorithm, and did crime analysis Using k-means clustering and published it in the year 2013.The merit of this survey was Predict crime trends and design precaution methods and demerit was Time taking process and can't predict multi- model data .[4]

McClendon and Natarajan Meghanathan used the models Linear Regression, decision Stump and Using machine learning to analyze crime data, and published in the year 2015. The merit of this survey is Predict the crime Pattern and demerit was can't predict the location. [5]

3. EXISTING SYSTEM

One of the existing methods used to analyze and predict crime was data mining. Data Mining is one of the tools that can be used to analyze and predict crime. In this survey the algorithm used in this method is Apriori Algorithm, Naïve Bayes, and Decision Tree. Using these algorithms they have tried to find a way to improve the data mining process that can help solve crime faster. Crime data analysis uses data mining. In this case the algorithm used in this way is Association rule mining, k-means clustering. Data from police departments were used to develop and evaluate the proposed model. This data was pre-processed to obtain pure and accurate data, using a variety of processing techniques. By analyzing the detection of the criminal pattern and predicting the algorithms used in this approach Apriori, k-means clustering, naïve Bayes, affiliates and retreats. Research shows that only 10% of criminals make up 50% of cases. This information is very useful and can be used to track criminals using Naïve-Bayesian and other algorithms. In the analysis of crime using k-means which combines crime analysis is done by making k-means combined into a crime database using the tool of fast miners.
The main objectives of crime analysis include:
1. Release of crime patterns through crime analysis and crime log.
2. Crime forecasts on local distribution of available data and crime rate prediction for different types of data mining.
3. Crime detection using a study machine to analyze crime data another survey was conducted and the algorithm used in this survey was Additive Regression, linear regression and Decision Stump. In this study, the lineback algorithm performed the best among the three selected algorithms.
The aim of this project is to prove how efficient and improve the machine learning algorithms analysis can be to predict violent crime patterns.

4. PROPOSED SYSTEM

- The approach taken for this project used machine learning and neural networks to predict the type of crime based on various given factors.
- The address is taken from dataset in text format which is building name, city name and nearby landmark in form a string.
- The text address is converted into latitude and longitude with help of geopandas and folium.
- Folium is used in our project for visualizing the crime location in order to get insights of the crime pattern.
- The location of crime is visualized in open street maps and leaf-let maps to find out possible pattern of crimes.
- The analysis of the crimes gave us insights about the crime pattern.
- In order to make the data more useful to the model we converted all the data of the dataset into numbers, and for the same purpose we used label encoder.

- Machine learning models like random forest and k-nearest neighbor, in addition to this multi-level perceptron are used to predict the type of crime based on various given factors.

5. TECHNOLOGIES USED

Python is the language used to program this undertaking, as it is flexible, and it gives a wide scope of library. Programming the crime log analysis and prediction is finished utilizing an online compiler like Google colab which upholds Python. Numpy where we use powerful data structures in Python that ensure effective statistics with arrays and matrices and provides a large library of mathematical operations that work on these same components and matrices. And Next library GeoPandas is an open source project to make geospatial data processing in python easier. GeoPandas expands the types of data used by pandas to allow local performance in geometric types and geopy makes it easy for Python user to find links to addresses, cities, countries, and important places around the world using geocoder. Nominatin It is also an OpenStreetMap data search tool by address or location (geocoding). Nominatim is included in GeoPy in the GeoPy Python library. Folium is a powerful Python module that helps us create multiple types of maps in Leaflet. By default, Folium makes a map in a different HTML file. As Folium effects interact, this library is very helpful in building a dashboard. Ratelimiter Can be used to prevent DoS attacks and limit web scratches. Whenever limits are applied, the limit is usually written down. Plugins are a great way to extend QGIS functionality. Marker cluster It is a form of integration so you can quickly see which areas are full Of crimes .so for our analysis part we have used all these above libraries for crime analysis part .Coming to our prediction part we have used three machine learning models like Random forest classifier, Knn classifier and multi-level perceptron which helps in prediction of crime. Voting classifier is used to aggregate/combine two or more machine learning or deep learning models and it builds these models parallelly to improve the accuracy of the model. The result is predicted by aggregating all the models/classifiers used in the parameter 'estimators'. Weight parameter tells the importance given to each model. Voting parameter has value 'hard' that means the voting is done for the results given by all the models written in 'estimators' parameter and then the final result is declared.

6. SYSTEM DESIGN

The model presented in the report uses machine learning, neural networks and data visualization to Gin incites of crime log of West Virginia.
- Fig.1.1 is the architecture diagram of the proposed model; the user gives the data to the model in form of text which includes address and building name.
- The machine can't process text directly thus it has to be converted into numbers. The text address is converted into latitude and longitude, after which we get data which is called geospatial data.
- This geospatial data is given to the program which plots various maps using folium which uses leaf-let maps and graphs which gives us a deep insight of crimes happening in West Virginia.
- In addition to the analysis of crime log of West Virginia, this report also proposes a model which uses machine learning and neural networks to predict the type of crime happened on the basis of various given factors.
- The model first encodes the text into numbers using label encoder in order to use it while training the model.
- The proposed model uses K-Nearest neighbor classifier to predict the type of crime, in addition to this the model uses random forest classifier and multi-layer perceptron classifier to predict the type of crime based on various given factors.
- The model combined all three classifiers together at last to improve the overall learning of the model.
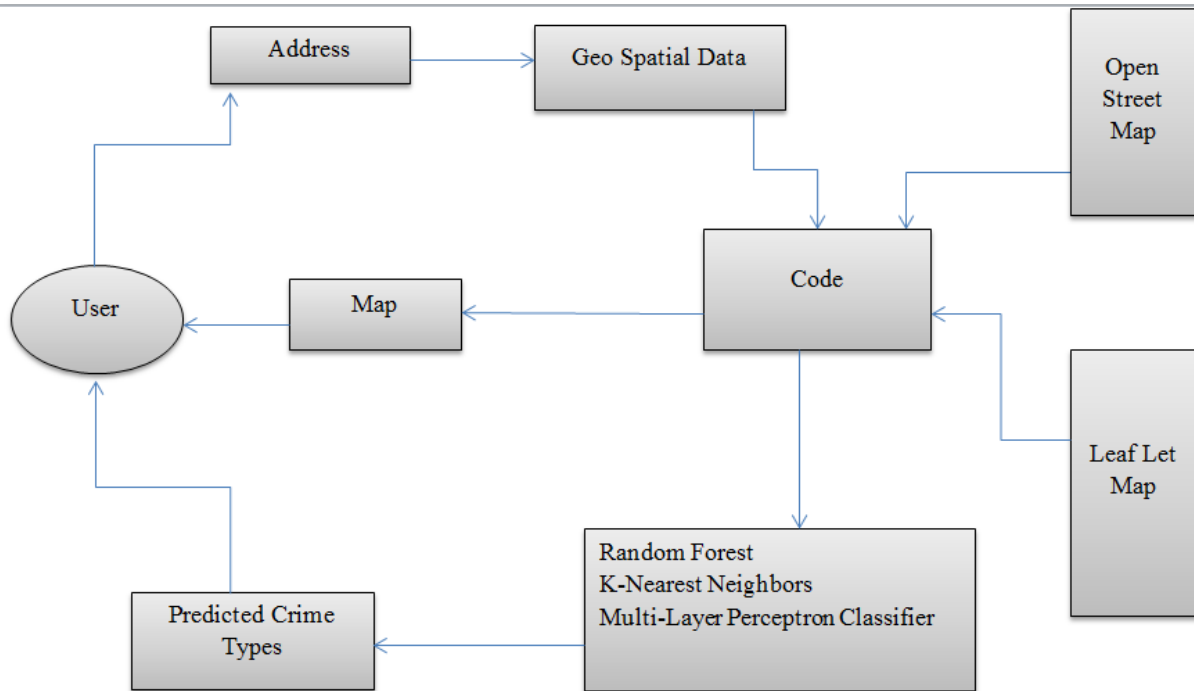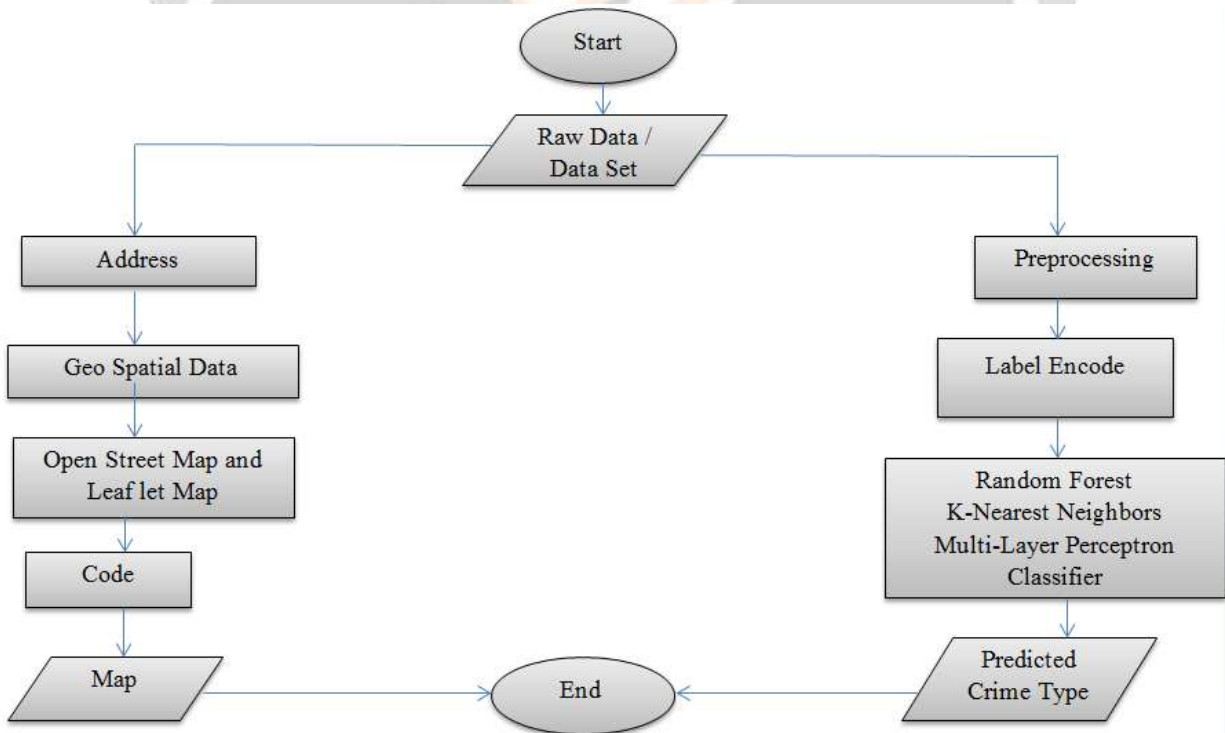
Fig 1.1

FLOW CHART

6.1 SYSTEM IMPLEMENTATION

The model implementation consists of following steps:-

- Firstly, we have to analyse the dataset inorder to get insights of the crime pattern which intern gives us the clue that which features we have to use which training our model.
- Second step is to filter the dataset; we have to remove the redundant and duplicate data to improve the performance of the model.
- Now, the dataset contains all the data in string dataset, so it has to be converted into numbers to improve the performance of the model, for this purpose we used libraries like label encoder and geopy (to convert address into coordinates).
- The model used random forest with 70 trees to predict the type of crime according to the given factors.
- The model also used k-nearest neighbors with 3 neighbors to predict the type of crime.
- Multi-level perceptron is also used to predict the type of crime in out model as neural network gave us better accuracy than any other machine learning methods.
- At last voting classifier is used to aggregate knn, random forest and MLP to increase the accuracy of the model.
- The voting classifier used all three classifiers mentioned above to aggregate the result of classifiers inorder to give us a whopping accuracy of 87%.

7. RESULTS AND DISCUSSION

The work in this work is very focused on analyzing crime log to have a better understanding of the hidden patterns behind the data and predict the type of crime based on specific data provided by machine learning and social networks. The machine learning categories and neural networks used to create the able model predict the type of crime according to certain given factors.

8. CONCLUSION

The data view caught us to find a relationship with the patterns in the middle various data. The work in this work is very focused on analyzing crime log to have a better understanding of the hidden patterns behind the data and predict the type of crime based on specific data provided by machine learning and social networks. I Analysis of crime log data in West Virginia allow us to have a deeper view of crime. Viewing data helps to analyze data set. Visibility of data using graphs and maps has helped to obtain details of crime patterns in West Virginia. The machine learning categories and neural networks used to create the able model predict the type of crime that has helped us to understand western Virginia crime that can be used to reduce crime rates that will keep the peace and keep the community safe.

9. FUTURE SCOPE

This project presented a comprehensive analysis of the criminal case of western Virginia to assist law enforcement agencies to obtain certain details of hidden patterns behind data. Scope of use many different ways of predicting and preventing crime can change the status quo law enforcement agencies. The use of ML and neural network can have a significant impact full functioning of law enforcement agencies. In the near future, with ML and neural network, and security features such as surveillance and surveillance cameras broadly, the machine can read the pattern of past crimes, understand what the crime is in fact, and he predicted future crimes accurately without human intervention. A possibility the default would be to build a system that can predict and anticipate crime scenes tropical areas of the city. Law enforcement agencies can warn and prevent crime which occurs through additional surveillance within the forecast area. This is complete automation can overcome the barriers of the current system, as well as law enforcement agencies may rely heavily on these strategies in the near future. Design a machine to anticipation and identification patterns of such crimes will be the beginning of our future read.

10. REFERENCES

[1]S. Sathyadevan and S. Gangadharan2014 Crime prediction by data mining
https://www.researchgate.net/publication/280722606_Crime_Analysis_and_Prediction_Using_Data_Mining
[2]Z. Zubi and A. Mahmmud 2014 Crime data analysis using data mining
https://www.naun.org/main/NAUN/computers/2014/a022007-096.pdf
[3]Sunil Yadav, Meet Timbadia and RohitVishwakarma 2017 Crime pattern detection analysis
https://www.researchgate.net/publication/321990399_Crime_pattern_detection_analysis_prediction
[4]JyotiAgarawal and RenukaNagpal 2013 Crime analysis using k-means clustering
https://www.researchgate.net/publication/269667894_Crime_Analysis_using_K   Means_Clustering
 [5]McClendon and NatarajanMeghanathan 2015 Using machine learning
https://www.researchgate.net/publication/275220711_Using_Machine_Learning_Algorithms_to_Analyze_Crime_Data