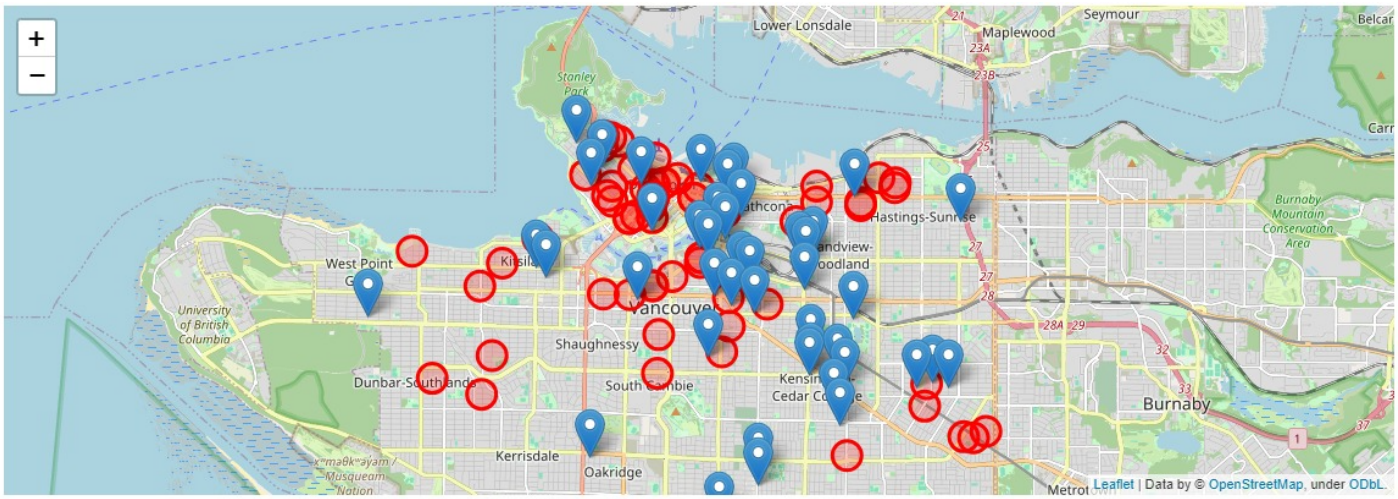
Crime Prediction In Vancouver City



2020-09-26

Red Circle: Predicted Crime Point Circle(1000ft radius)
Marker Point: Actual Crime Point

Abstract

Crime is one of the biggest and dominating problems in our society and its prevention is an important task. Daily there are huge numbers of crimes committed frequently. This requires keeping track of all the crimes and maintaining a database for the same which may be used for future reference. The current problems faced are maintaining a proper dataset of crime and analyzing this data to help in predicting and solving crimes in future. The objective of this project is to analyze a dataset which consists of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions.

In this project, we will be using the technique of machine learning and data science for crime prediction of Vancouver crime data set. The crime data is extracted from the official portal of Vancouver police department. It consists of crime information like location description, type of crime, date, time, latitude, longitude, etc. Before training of the model data preprocessing will be done following this feature selection and scaling will be done so that accuracy obtained will be high. The various machine learning algorithms will be tested for crime prediction and one with better accuracy will be used for training.

Dataset

We have extracted the information of crime data from the official portal of Vancouver police department. The dataset consists of crimes in Crime in Vancouver from 2018 to 2020

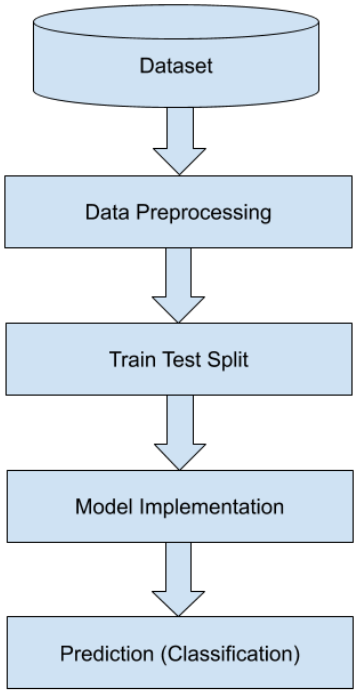
which consists of 81,835. It consists of features like type, date, time, location, latitude, longitude and many more.

Data Preprocessing

Initially we need to preprocess data by removing all null values and removing all columns that are unnecessary. One hot encoded categorical data and below columns are checked for NULL values.

Latitude	0	Count_Break_Enter_Last_30_Days	0
Longitude	0	Count_Mischief_Last_30_Days	0
Days_Since_Last_Crime	0	Count_Other_Theft_Last_30_Days	0
Distance_From_Last_Theft_From_Auto	0	Count_Recovered_Vehicle_Last_30_Days	0
Distance_From_Last_Break_Enter	0	Count_Theft_From_Auto_Last_30_Days	0
Distance_From_Last_Mischief	0	Count_Theft_Of_Auto_Last_30_Days	0
Distance_From_Last_Other_Theft	0	Type_of_Crime	0
Distance_From_Last_Recovered_Vehicle	0	FSA	0
Distance_From_Last_Theft_Of_Auto	0	Time_From_00_06	0
Days_Since_Break_Enter	0	Time_From_06_12	0
Days_Since_Mischief	0	Time_From_12_18	0
Days_Since_Other_Theft	0	Time_From_18_00	0
Days_Since_Recovered_Vehicle	0	Crime_Year	0
Days_Since_Theft_From_Auto	0	Crime_Month	0
Days_Since_Theft_Of_Auto	0	Crime_Day	0
Count_Break_Enter	0	Day_Name_Fri	0
Count_Mischief	0	Day_Name_Mon	0
Count_Other_Theft	0	Day_Name_Sat	0
Count_Recovered_Vehicle	0	Day_Name_Sun	0
Count_Theft_From_Auto	0	Day_Name_Thu	0
Count_Theft_Of_Auto	0	Day_Name_Tue	0
		Day_Name_Wed	0

Methodology



The proposed work is divided into 4 parts:

1. Data preprocessing
2. Data Analysis
3. Data Modelling
4. Evaluation of performance

Model sampling for train and split:

- Training dataset consists 70% data
- Testing dataset consist of 30% data

Machine learning models -

One approach used to implement crime prediction is, Categorical Variables are encoded and then used for training in model. Crime type is our output (target). Since we are going to classify types of crimes, we are going to implement the following machine learning models.

Using features like **LATITUDE, LONGITUDE, POSTAL CODE, DAYS_SINCE_LAST_CRIME, DISTANCE_FROM_LAST_CRIME** we predicted the type of crime that could happen in Vancouver city.

Models implemented -

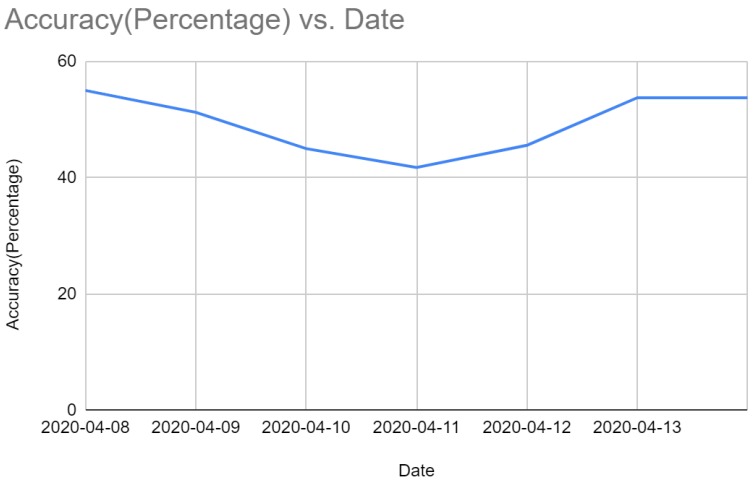
- ➢ Support Vector Classification (SVC)
- ➢ Naive Bayes
- ➢ Decision Tree Classification
- ➢ K Nearest Neighbors (KNN) Classification
- ➢ Logistic Regression
- ➢ Linear Discriminant Analysis (LDA) Classification

Results

Model Name	Accuracy
Support Vector Classification (SVC)	0.47155
Naive Bayes	0.301
Decision Tree Classification	0.413125
K Nearest Neighbors (KNN) Classification	0.454125
Logistic Regression	0.4905
Linear Discriminant Analysis (LDA) Classification	0.477375

As we can see from the results obtained from the table the algorithm which can be used for the predictive modeling will be SVC algorithms with accuracy of 0.47155 best among the rest of the algorithm.

Note: *While updating the dataset, model accuracy can increase or decrease.*



Conclusion

In this research, The dataset for Vancouver city is from 2018 to 2020. After applying machine learning models, the classification prediction obtained is approximately 47%. Different algorithms require different training time. The accuracy can be improved if we include more features in our dataset. After understanding the dataset and results, more features are needed to distinguish different crimes from each other.

Reference

1] Data of actual crime reported to the Vancouver Police Department

https://vancouver.ca/police/CrimeMaps/

2] Research paper on crime prediction

https://pdfs.semanticscholar.org/3bb0/40430edf0ffdef0c93dadc04dcd7e6905637.pdf
https://www.irjet.net/archives/V5/i9/IRJET-V5I9192.pdf
https://www.researchgate.net/publication/275220711_Using_Machine_Learning_Algorithms_to_Analyze_Crime_Data
http://athena.ecs.csus.edu/~shahr/poster.pdf

3] Plotted the value on map
https://blog.dominodatalab.com/creating-interactive-crime-maps-with-folium/