

CRIME PREDICTION AND PREVENTION USING KNN MODEL

Shubham Kr Singh^{*1}, Rajat Gupta^{*2}, Soumya Choudhary^{*3}, Aashita Chhabra^{*4}

^{*1*}^{*2*}^{*3} Department Of Information Technology, Dr Akhilesh Das Gupta Institute Of Technology & Management, New Delhi ,Delhi, India

^{*4} Assistant Professor, Department Of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology and Management, New Delhi, Delhi, India.

Abstract - Crime is one of the dominant and appalling aspects of our society. Everyday immense numbers of crimes square measure committed, these frequent crimes have created the lives of common voters restless. So, preventing the crime from occurring could also be a significant task. In recent times. However, it's required to keep up a correct info of the crime that has occurred as this data are often used for future reference. the power to predict the crime which might occur in future can facilitate the enforcement agencies in preventing the crime before it happens. the aptitude to predict any crime on the premise of your time, location so on will facilitate in providing helpful data to enforcement from a strategic perspective. However, predicting the crime accurately could also be a difficult task as a result of crimes square measure increasing at associate appalling rate. Thus, the crime prediction and analysis strategies square measure very important to find the long term crimes and cut back them. In recent times, several researchers have conducted experiments to predict the crimes victimization varied machine learning strategies and explicit inputs. For crime prediction, KNN, call trees and a few different algorithms square measure used.

Key Words: K-Nearest Neighbor Support, Vector Machine, Auto regressive moving average, recurrent neural network, Recursive Feature Elimination, National Crime Records Bureau

1. INTRODUCTION

Crimes are a significant threat to humankind. There are many crimes that happen in a regular interval of time. In fact crime is increasing across places at a fast and vast rate. Crimes happen from small village, towns to big cities. These crimes are of different types like theft, murder, kidnapping, rape, assault, false imprisonment, homicide. Since crimes are increasing, it is necessary to solve these cases in a much faster way. The crime activities have increased at a faster rate and it is the responsibility of the police department to control and reduce the crime activities. The primary motive of this project

is crime prediction using the features that is present in the dataset. The dataset is extracted from the official sites. With the help of machine learning algorithms, using python as core we can predict the type of crime which will occur in a particular area.

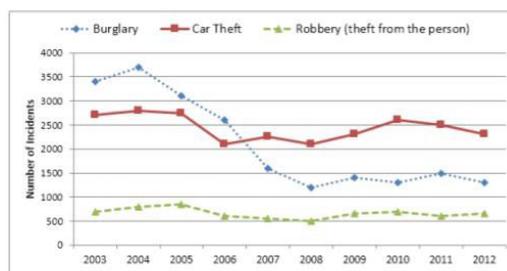


Figure 1: Common Crime Rate

2. METHODOLOGY

Predictive modeling is basically building a model which is trained using a dataset to predict unknown values. This process includes a machine learning algorithm that learns patterns and properties from a training dataset in order to make those predictions. Predictive modeling can be of two types: Regression and pattern classification. Regression models analyses the relationships between variables and trends in to make accurate predictions about continuous variables.

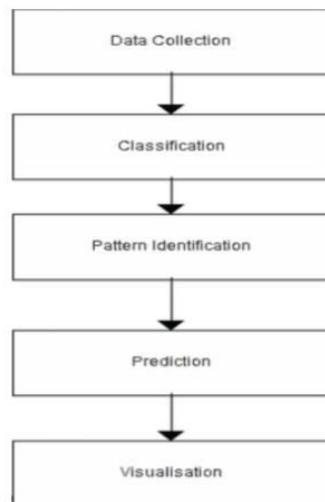


Figure 2: Architecture

3.1 Data Preparation

The dataset used is Communities and Crime dataset available on UCI edu website. The dataset consists of crimes in Vancouver from 2003 to 2017 which consists of 530,652 values. It consists of features like State, Country, District, Community, Population, Race, Age, Income, Hour and month.

3.2 Data Preprocessing

Firstly, the null values are deleted using `df = df.dropna()` where `df` is the data frame. The categorical attributes are converted into numeric using Label Encoder. The date attribute is converted to attributes like month and hour which are used as feature for the model. Features selection is done to build the model. The attributes used for feature selection are State, Country, District, Community, Population, Race, Age, Income, Hour and month.

3.3 Model Selection

We have to select one of the following models based on the defined goal and dataset.

- A. K-Nearest Neighbor uses a database in which the data points are separated into several classes to predict the classification of a new sample point.
- B. Logistic Regression is one of the regression models where the variables which are dependent is either binary is categorical. Continuous data cannot be handled using this model.
- C. Decision Trees is a tree shaped graph which includes outcomes, utilities which helps in making decision.
- D. Support Vector Machine finds a hyperplane which separates two or more classes. It takes maximum time for processing.
- E. Bayesian Methods is based on Naïve Bayes algorithm which constructs models as classifiers and is represented as vectors of all values of features.

3.4 Prediction

After feature selection the dataset is divided into a pair of `xtrain`, `ytrain` and `xtest`, `ytest`. The algorithm models are imported from `sklearn`. After the model is build, prediction is done using `model.predict(xtest)`. The accuracy is calculated using `accuracy_score` which is imported from `metrics - metrics.accuracy_score (ytest, predicted)`.

ALGORITHM	ACCURACY
KNeighbors Classifier	0.78734858681022879
GaussianNB	0.6460296096904441
MultinomialNB	0.45625841184387617
BernoulliNB	0.31359353970390308
SVC	0.31359353970390308
DecisionTree Classifier	0.78600269179004034

Figure 3: Prediction Accuracy of all algorithms used

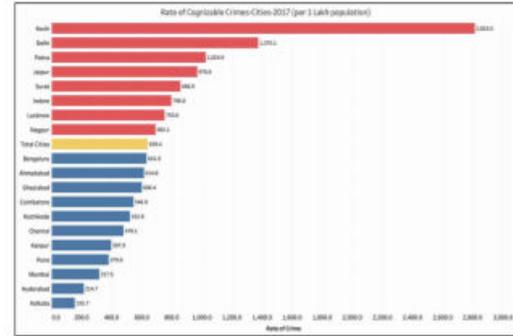


Figure 4: Rate of cognizable crimes

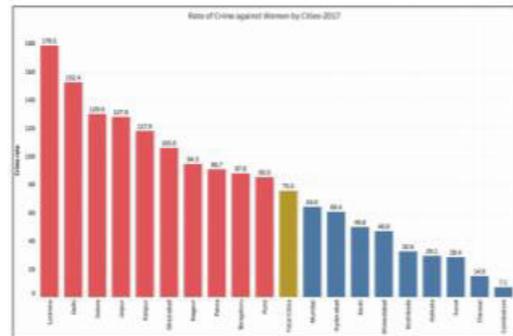


Figure 5: Rate of crimes against women

3. CONCLUSION

Using machine learning algorithms, it has become comparatively easier to find the relations and patterns among various data's. This project mainly involves predicting the type of crime which could happen knowing the location of where it has occurred. We built a mode using machine learning algorithms on training data set that have undergone data cleaning and data transformation. This model predicts crime rate with an accuracy of 0.789.

REFERENCES

[1] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225-230). IEEE.

[2] Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.

[3] Dataset from UCI edu. Url :<http://archive.ics.uci.edu/ml/datasets/communities+and+crime>.

[4] Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat

2006 workshops. 2006 IEEE/WIC/ACM International Conference on (pp. 41-44). IEEE.

[5] Al Boni, M., & Gerber, M. S. (2016, December). Area Specific Crime Prediction Models. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 671-676). IEEE.

[6] Kadhim B. Swadi Al-Janabi, (2011, May), "A Proposed Framework for Analyzing Crime Data Set using Decision Tree and Simple K-means Mining Algorithms", Vol. 1- No. 3, pp. 8-24.

[7] Premalatha, M. & Vijayalakshmi, C.. (2019). SVM approach for non-parametric method in classification and regression learning process on feature selection with ϵ - insensitive region. Malaya Journal of Matematik. S. 276-279. 10.26637/MJMOS01/0051.