CRIME PREDICTION USING MACHINE LEARNING TECHNIQUES

JAMES STEPHEN MEKA Principal, WISTM, Andhra University, (jamesstephenm@gmail.com)
K. HIMANTHA RAO Assistant Professor, WISTM (himantharaokolli@gmail.com)

Abstract- Crime is one amongst the largest and dominating downside in our society and its bar is a vital task. A massive variety of crimes are committed every day. The present day issue is that the maintenance of a correct crime knowledge set. Therefore the analysis of this data to assist within the prediction is the backbone of future criminality. In this paper we are used machine learning algorithm to forecast crime. The goal is to judge a dataset as well as multiple crimes and forecast the type of crime which will occur in the future that is supported by numerous conditions.

Keywords*: prediction of future crimes, machine learning, analysis of crime.*

1.Introduction

Crimes purpose a sizeablerisk to humanity. On a daily basis, numerous crimes take place. maybe it's growing and spreading at a fast and widespread rate. Crime happens in little villages, towns, and huge cities. Robbery, Murder, Rape, Assault, Battery, False Imprisonment, Kidnapping, and kill are every type of crimes. As a result of crime is on the rise, it is important to solve cases as quickly as possible. Crime activity has hyperbolic at a rapid pace, and it is the work of the department of local police to regulate and cut back crime activity. As a result of there's a colossal quantity of crime knowledge available, the police department key issues are crime prediction and criminal identification. Hence, the needfor technology which will help in quicker case solve is there.The goal would be to coach a prediction model. The coaching data set would be used for training, whereas the check dataset would be used for validation. looking on the accuracy, a far better rule is accustomed to build the model. For crime prediction, the K-Nearest Neighbor (KNN) classification and alternative algorithms will be employed. The image of data sets is meted out so as to look at the crimes which will have occurred within the country. This system improves the exactitude with that enforcement agencies predict and observe crimes in Chicago, lowering the crime rate. Machine learning algorithms have advanced dramatically, creating crime prediction supported historical data possible. The goal of this analysis is to use machine learning models to analyse and forecast crime in states. It focuses on developing a model that aid in police work the quantity of crimes by class in an exceedingly given state.The machine learning

model K-NN are going to be utilized to predict crimes during this study. Space to understand the crime pattern, a detailed geographical analysis can be accomplished. Varied mental image approaches and plots are wont to assist enforcement authorities in police work and prediction crimes with bigger accuracy. This may indirectly serve to scale back crime rates and might facilitate to boost security in such important regions.

2.Literature Review

Machine learning is that the scientific study of algorithms and applied mathematics models that systems use to perform a selected task while not victimization specific instructions, wishing on patterns and reasoning instead. Tt's seen as a set of Artificial Intelligence. Machine learning algorithmic rules are employed in a good sort of applications, corresponding to email filters and computer vision, wherever it's troublesome or unfeasible to develop a standard algorithm for effectively playacting the task. Machine statistics, that focuses on creating predictions with computers, is closely involving machine learning. Machine learning advantages from mathematical improvement analysis as a result of it provides tools, theory, and application fields. Machine learning is typically referred to as prophetical analytics once applied to industrial concerns. Machine learning uses knowledge to spot distinct patterns in an exceedingly dataset.

1. It will learn from previous data and improve on its own.
2. it's a technology that's primarily based on data.
3. Machine learning and data processing are terribly similar therein they each agitate massive amounts of data.

Knowledge Analysis

A. KnowledgeSource:

Communities within the US provided the initial data source. the info incorporates socioeconomic data from the 1990 Census, enforcement statistics from the 1990 enforcement Management and body Statistics survey, and crime data from the 1995 FBI UCI. It contains of 147 attributes and 2215 instances in this dataset.

B. Preprocessing:

To fill the vacant cells, cut back extraneous columns, and add numerous helpful features, the initial dataset should be preprocessed. The Fig.1 shows the preprocessed datasets.

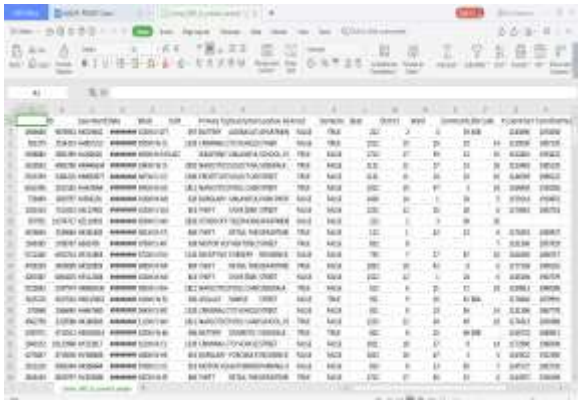


Fig.1Preprocessed Dataset

C. Applied Mathematics Analysis:

The crime dataset illustrated in Figure 2 is sorted by year, month, and day of the week. In Chicago, the mean number of crime outbreaks per year would be almost 31624, 2720 a month, and 90 per day. As the time intervals rise, the dataset tends to follow a standard distribution. The graph of every day, on the other hand, shows an unusual maximum value of 650 occurrences. The crime statistics by (a) year, (b) month, and (c) day is depicted in Figure-2.

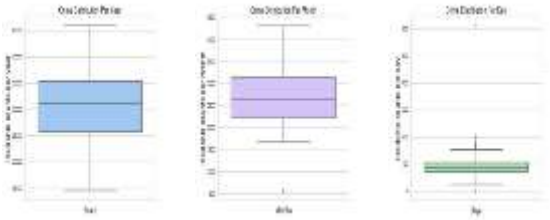


Fig.2 Crime statistics by (a) year, (b) month, and (c) day

D. Trend Analysis:

Per Fig.3, the overall trend shows that the common variety of arrests per month shrunken from 2001 to 2010, climbed in 2011, so considerably decreased to around 3000 events yearly in 2017.

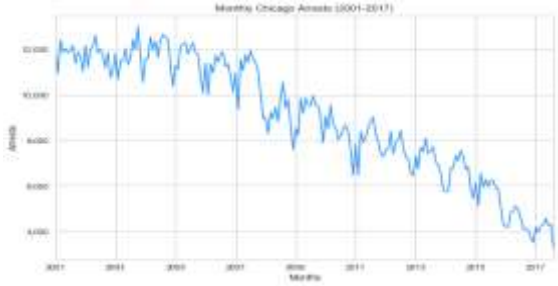
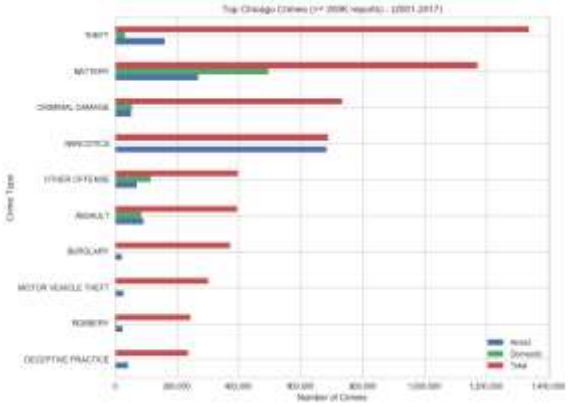
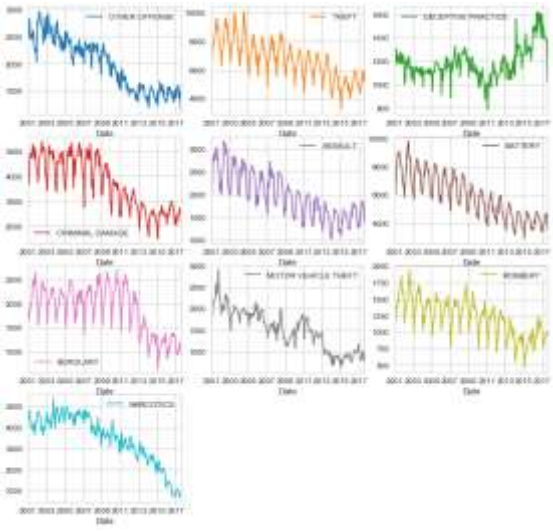


Fig.3 Monthly Chicago Arrests

Arrests the foremost incidents were within the class of stealing from cars, followed by battery. However, whereas auto felony has decreased dramatically in recent years, alternative forms of crime have surged. Figure four depicts the frequency and trends of every crime



(a)



(b)

Fig.4 (a) Number and (b) trend of crimes committed by type.

3.Methodology

The two varieties of supervised learning are classification and regression. Classification is that the task of predicting a distinct class label, whereas regression is the duty of prediction a nonstop quantity. The goal of this analysis is to predict the types of crimes that will occur in a very specific location. The aim of the study is therefore to classify the crime. Some techniques to adopt the classification are K-NearestNeighbor (KNN), Nave Bayesian, Support Vector Machine (SVM),Decision Tree, and Ensemble Methods. Every technique has advantages and drawbacks in terms of complexity, accuracy, and coaching time, and totally different results may be obtained from one dataset. The KNN technique was accustomed

coach our model throughout this study. one amongst the foremost basic categorization strategies is KNN. If the most values nearer to z are from category A, it allocates a sample z to class A; otherwise, it assigns the sample to class B. The likelihood of the take a look at sample happiness to class is calculated victimisation the subsequent formula in KNN:

$$P(x \in C_i) = \frac{\sum_{j \in C_i} {}^{n_j}/_{d_j}}{\sum_{l=1}^{k} {}^{n_l}/_{d_l}}$$

Wherever and are the amount of components depicted by every coaching dataset sample, and also the geometrician norm-based distances between the test and also the associated coaching sampled source. KNN keeps track of all obtainable objects and classifies new ones supported the similarity live by trying to find the input values nearest neighbours.

4.Experimental Results

The consequences of making use of the KNN algorithmic rule to the dataset are displayed below. This map provides an in-depth check out the crimes that occurred in Chicago from 2001 to 2017.

5.Conclusion

Throughout the analysis it's been evident that basic details of a criminal activities in a part contains indicators which will be employed by machine learning agents to classify a

criminal activity given a location and date. Despite the fact that the training agent suffers from unbalanced classes of the dataset, it absolutely was ready to overcome the issue by oversampling and undersampling the dataset. Through the experiments, it are often seen the imbalanced dataset was benefitted by exploitation KNN undersampling.

Exploitation the undersampled data, Adaboost decision tree with success classified criminal activities supported the time and location. With a accuracy of 81.93%, it was ready to outstrip alternative machine learning algorithms. Unbalanced categories are one in all the most hurdles to realize a stronger result. Though the machine learning agent was able to prophetical model out of merely crime data, a demographic dataset would most likely facilitate to more improve the result and solidify it.

6.References

[1] Jacky Yu, Wei Ding, Ping Chen, Melissa Morabito, “Crime Forecasting Using Spatio-Temporal Pattern with Ensemble Learning”, Pacific-Asia Conference on Knowledge Discovery and Data Mining, May 2014, Boston, USA.

[2] Dawei Wang, Wei Ding, Henry Lo, Tomasz Stepinski, Josue Salazar, Melissa Morabito, “Crime hotspot mapping using the crime related factors—a spatial data mining approach”, Applied Intelligence, Volume 39, December 2012, Boston, USA.

[3] Md. AminurRabRatul, “A Comparative Study on Crime in Denver City Based on Machine Learning and Data Mining”, Cornell University Machine Learning, Jan 2020, Ottawa, Canada.

[4] Y. L. Lin, L. C. Yu, and T. Y. Chen, "Using machine learning to assist crime prevention," *IEEE 6th Intl. Congr. on Advanced Appl. Inform. (IIAIAAI)*, Hamamatsu, Japan, Jul. 2017.

[5] X. Zhao and J. Tang, "Exploring transfer learning for crime prediction," *IEEE Intl. Conf. on Data Mining Workshop (ICDMW)*, New Orleans, LA, USA, Nov. 2017.

[6] R. Marchant, S. Haan, G. Clancey, and S. Cripps, "Applying machine learning to criminology: semi parametric spatial demographic
Bayesian regression," *Security Inform.*, vol. 7, no. 1, Dec. 2018.

[7] L. McClendon and N. Meghanathan, "Using machine learning algorithms to analyze crime data," *Mach. Learn. and Appl.: an Intl. J. (MLAIJ)*, vol.2, no.1, Mar. 2015.