A Comparative Study of Decision Tree and Naive Bayes Machine Learning Model for Crime Category Prediction in Chicago

Bshayer S. Aldossari
CCSIT, IAU, KSA
2150009411@iau.edu.sa

Futun M. Alqahtani
CCSIT, IAU, KSA
2150005986@iau.edu.sa

Noura S. Alshahrani
CCSIT, IAU, KSA
2150008026@iau.edu.sa

Manar M. Alhammam
CCSIT, IAU, KSA
2140001814@iau.edu.sa

Razan M. Alzamanan
CCSIT, IAU, KSA
2150009276@iau.edu.sa

Nida Aslam
CCSIT, IAU, KSA
naslam@iau.edu.sa

Irfanullah
CCSIT, IAU, KSA
iurab@iau.edu.sa

ABSTRACT
Crimes have a massive effect on individuals regards their minds, concern, and psyche, along with their indisputable impact on society. As a result, law enforcement keeps monitoring their control areas so they would notice any suspicious activity, which makes them more cautious and raises their ability to prevent potential criminal activity. This study aims to create a machine learning model that will be able to predict the potential crime category in a particular geographic area, by exploring and analyzing the existing repeated incidents data. Decision Tree and Naive Bayes are applied on a dataset, which was extracted from the Chicago Police Department's CLEAR. By applying these techniques, law enforcement agencies will highly benefit as they will be able to predict the potential crimes, which will increase the ability to prevent them. We have applied the two algorithms on the selected top 9 features from the chosen dataset. However, comparing the two algorithms, Decision Tree performed better than Naïve Bayes algorithm.

CCS Concepts
Computing methodologies → Classification and regression trees

Keywords
Crimes Category, Chicago Crimes, Supervised Learning, Decision Tree, Naïve Bayes.

1. INTRODUCTION
In all communities, crimes are source of trouble for the country's development. Crimes have a high impact on societies' peace, satisfaction, and financial growth. Therefore, exploring, analyzing and studying factors that results in criminal activity is an important topic especially for governmental security offices. Since security offices job is to detect and reduce crimes occurrences, criminal analysis and prediction will be a great help to them. However, it is hard for the police stations to do the analysis due to the huge amount of data. As a result, innovation and technology advances should have their contribution to society. Criminal activity is defined as a risk for an individual, which is brought by another individual creature that can deserve penalty according to the government's rules [1].

Machine learning techniques will be used for this purpose, such as classification, regression, and clustering. Many studies have been made for this purpose. R. Kiani, S. Mahdavi, and A. Keshavarzi have made a study on for analyzing and predicting crimes by using clustering and classification [2]. Moreover, R. T.V, J. Joshy, M. R, and A. Soni M have made crime prediction and analysis using clustering and regression [1] and many others. In our study several classification algorithms will be applied and compared to predict the category of crimes. Results of each algorithms will be compared in order to choose the one with the best outcome.

The remaining part of our study is organized as follow. Section 2 contains a review of related literature. Section 3 contains the proposed machine learning techniques i.e. Naïve Bayes and Decision Tree. Section 4 contains empirical studies that include dataset description, experimental setup or methodology. Section 5 presents result and discussion, while section 6 contains the conclusion of our study.

2. LITERATURE REVIEW
A crime is an act which affects the society peace including a violation against human life, and governmental or private organization/property, caused by an individual or group of individuals [3]. This paper presents some of the previous related studies on Crime Prediction, that have been done in different crime-prone societies, using different Machine Learning techniques that have been applied on real datasets to predict and analyze crimes data based on certain criteria.

ICCDE '20, January 4–6, 2020, Sanya, China

ACM ISBN 978-1-4503-7673-0/20/01…$15.00
https://doi.org/10.1145/3379247.3379279

2.1 Crime Analysis

Recently computer has a significant role in every field, also in tracing criminals and tracking their crimes. Since the criminals can have specific properties and may have different crime characteristics and crime careers. Law enforcement agencies take advantage of data analysis techniques to improve Criminology. It refers to the process used for identifying crimes and criminals' characteristics, and to detects the patterns of previous committed crimes in order to predict [4]. As J. Agarwal, R. Nagpal, and R. Sehgal [5] discussed the main aims of crime analysis are:

- Applying crime analysis to find crime patterns based on real criminal information.
- Prediction of crimes and their occurrence using various data mining techniques.
- Crime recognition.

3. Crime Analysis Techniques

The literature review is divided into three categories based on the technique used for detection and analysis:

3.1. Classification

Classification is a supervised machine learning technique that categorizes data into groups (classes) and is used for making the future predictions. In other words, classification is used to predict the classes of a given data based on certain attributes known as predictors [7]. A survey has been made on the detection and analysis of different crime detection mechanism using machine learning [6].

Sathyadevan et al. [8] have applied classification methods on location, date, and type of crime data extracted from Websites, Blogs, Social Media, RSS feeds to predict areas of high probability for crime occurrence. They have used the following classification methods:

- Naive Bayes classification method to create a model by training crime data related to vandalism, murder, robbery, burglary, sex abuse, gang rape, arson, armed robbery, highway robbery, snatching, etc.
- Apriori algorithm to identify crime patterns that occur frequently.
- Decision Tree to predict the crimes areas. The techniques used are easy to apply and interpret, but they can only predict the crime spot without predicting the crime occurrence time.

K. Bogahawatte and S. Adikari [9] developed a system called Intelligent Crime Investigation System (ICIS), which can predict a possible criminal based on the evidence and observations collected from the crime scene. Colombo crimes and criminal database records is trained using Naive Bayes classification algorithm to distinguish potential suspects.

Decision tree that uses an improved feature A. Chandrasekar carried out an attempt to predict and classify the crimes in San Francesco city. The dataset of crimes that took place in San Francisco city over twelve years. First, they preprocessed the data by removing some features, replacing some indexed features with a number, and decomposed timestamp feature into five features. The dataset is enriched the feature by adding features from United States Census data. Finally, they performed multiple classification techniques in order to classify the crimes into 39 categories (original dataset classes). The study results in high testing error, which indicated a high variance, so they merge their dataset classes into two classes and re-applied the same classification techniques i.e. Naive Bayes, Random Forests, Support Vector Machines, and Gradient Boosted Decision Trees. High accuracy and precision were achieved using Gradient Boosted Trees and Support Vector Machines [10].

Another significant work on crime prediction was proposed by A. Babakura, M. Sulaiman, and M. Yusuf on predicting the crime category with the labels "Low", "Medium" and "High". A comparison was made between the classification algorithms namely, Naive Bayesian and Back Propagation on a real and authentic dataset obtained from UCI machine learning repository. The results indicated that the Naive Bayesian algorithm achieved better accuracy and precision [11]. M. Sharma [12] has presented a tool, which applied an enhanced Decision selection method for enhancing an efficiency of the algorithm on a real and open e-mail dataset in order to detect a suspicious e-mail of potential criminal activities.

Iqbal et al. has applied classification methods to predict the crime category in different states of USA [13]. Decision Tree and Naïve Bayes classification algorithms were applied to a real dataset in order to predict the crime category. However, the results of the comparison in terms of the accuracy of the classification algorithm used shows that the Decision Tree algorithm has a higher performance in predicting the crimes category.

To sum up, in order to take advantage of the classification methods in predicting the crimes, the data should be trained, and the classes should be known in advance.

3.2. Regression

Regression is a type of predictive modeling for analyzing and examining the associations among the variables and trends for the sake of performing the predictions on the continuous variables [14]. The cluster of data is given to one of the regression methods, such as Linear Regression, Ridge Regression, and Naive Bayes for crime predictions [1]. The study has been made on several cities in India aiming to predict and analyze crimes by applying various clustering and regression methods.
The following regression methods were used in their study:

a. Linear Regression

A. Bharati and S. RA. K [14] describe Linear Regression as a statistical procedure for predicting the associations between variables. Also, linear Regression is known among the other regression methods as the most simple and easy to use [1]. Similarly, P. Gera and R. Vohra [15] used the Linear Regression method on their study for analyzing a dataset that contains data related to the past 59 years for predicting the incidence of crimes in Delhi. In P. Gera and R. Vohra study [15],

the system should predict the data for the next 15 years after training using Linear Regression method by the past year data.

b. Ridge Regression

In some cases, Linear Regression can cause high variation. As a result, on R. T.V, J. Joshy, M. R, and A. Soni M study [1], they decided to use Ridge Regression. To improve the accuracy of their model, they decided to use biasing and variance to overcome overfitting and underfitting. In order to reduce the errors and enhance accuracy for crime prediction, Python was used to decide on the values for alpha. The value of alpha that was used for dacoity is 15, for murder is 0.5 and for rape is 0.001 [1].

c. Naive Bayes Regression

Naive Bayes algorithm is one of the widely used algorithm and is considered as the fast algorithm, where the theory of Bayes is applied for classifying various classes [1]. The regression method that gives a higher accuracy for R. et al. study [1] was Naive Bayes Regression as the values were closer to 1. Root mean square method was used to decide the efficiency of every regression method [1].

3.3. Clustering

Clustering is an unsupervised data mining technique used to split a group of items and data into clusters based on certain characteristics, each cluster contains a group of similar data. Clustering is used in undefined and unfixed classes and without supervision when it comes to grouping the objects [7].

R. et al. [1] have used diverse clustering methods to analyze crimes in the dataset State Crime Records Bureau (SCRB) of Tamilnadu, India. The data has been obtained which has approximately 38 several cities and districts crime data. The clustering method that has been used for the analysis was DBSCAN and K-means. To resolve the efficiency of the two clustering approaches, they used Silhouette coefficient. The study concluded DBSCAN clustering with a high accuracy and resulted in more accurate clusters.

P. Vrushali et al. [16] focused on the frequency rate during diverse years to classify clustered crimes and the study was conducted on real data which offers an innovative framework for clustering and predicting crimes. To analyze the crime data, in addition, to categorize crimes by grouping the related patterns, several clustering methods were applied.

In conclusion, several crime categorization and prediction methods were discussed that were applied to different crimes datasets. All the discussed papers have applied more than one machine learning technique to achieve the expected results. Therefore, we will need to apply many techniques to our dataset to analyse the data behavior, in order to create an efficient module to predict the crime category with the optimal accuracy.

4. METHODOLOGY

4.1 Description of the Proposed Techniques

This section presents the description of the machine learning techniques that is used in our study to predict the crime category based on behavioral data. In our study we used supervised learning approach and chose Naïve Bayes and Decision Tree algorithms, which fall under the classification technique.

4.2 Na ïve Bayes Technique

Naïve Bayes is a simple supervised classification algorithm. It uses the concept of Bayes Theorem, which classifies tuples to a class label related to a dataset based on the calculations of the conditional probability for each class label and attribute. The Naïve Bayes algorithm works by the given probability rule.

$$P(H \mid X) = \frac{P(X \mid H)\, P(H)}{P(X)}$$

Where X is a data sample of dimensions $x = < x_1, x_2, x_3 \ldots . x_n >$ with an unknown class label, and X belongs to H hypothesis class label. Where $P(H)$ is the prior probability (initial probability) associated with hypothesis H [17].

4.3 Decision Tree Technique (J48)

Decision Tree is a supervised learning technique, which recursively partitions the instance space. Its main use is to predict the class labels. The Decision Tree is composed of internal nodes, which represents set of predictors (attributes), edges, which represent a specific value or range of values of the input predictors (attributes), and leaf nodes, which represent the class labels. The internal nodes along with their edges split the instance space into two or more partitions and each terminal node (leaf node) of the tree is a class label [18].

J48 classifier is a simple C4.5 Decision Tree for classification. In this technique, a binary tree is constructed to model the classification process. Once the tree is built, it is applied to each tuple in the database and results in a classification for that tuple. J48 splits the data into range based on the attribute (predictors) values. J48 allows classification by Decision Tree or the rules generated from them [17].

The Basic Steps in J48 classifier:

- For the instances of the same class, the tree represents a leaf labeled with the class.
- By applying a test on each attribute, the potential information is calculated for all of them.
- Then, a test is applying on the attribute to calculate the information gain.
- Then, based on the selection criterion, the best attribute is selected and used for branching.

The “Entropy” is used in this process, which is a measure of the data disorder. The Entropy of $\vec{y}$ is calculated by

$$Entropy\ (\vec{y}) = -\sum_{j=1}^{n} \frac{|y_i|}{|\vec{y}|}\ log\ (\frac{|y_i|}{|\vec{y}|})$$

$$Entropy\ (j|\vec{y}) = \frac{|y_i|}{|\vec{y}|}\ log\ (\frac{|y_j|}{|\vec{y}|})$$

And Gain is

$$Gain\ (\vec{y}, j) = Entropy\ (\vec{y} - Entropy\ (j|\vec{y}))$$

The objective is to maximize the Gain, dividing by overall entropy due to the split argument by value j [19].

5. EMPIRICAL STUDIES

This section describes the dataset, statistical analysis of the dataset, and experimental setup.

5.1 Description of Dataset

The selected dataset reflects reported incidents that occurred in the City of Chicago from 2013 to 2017. Data is extracted from the Chicago Police Department's Citizen Law Enforcement Analysis and Reporting (CLEAR) system [20]. The dataset contained 12109 records and 18 attributes/features, including the target. The attributes contain nominal and numeric data. The techniques used in our study are Naïve Bayes and Decision Tree.

5.2 Experimental Setup

The experiment was conducted using Weka. Weka was used for data preprocessing and applying machine learning techniques. Moreover, the dataset was split into 70% for the training data, and 30% for the testing data to investigate the performance of the proposed technique. For data preprocessing, multiple preprocessing steps have been applied to the dataset, which include:

5.2.1. Data Cleaning

The dataset was pruned from the missing values. Since some the data are missing completely at random, listwise deletion (deleting rows) method was applied. Also, we manually reduced the dataset records to achieve a balanced distribution of the class labels.

5.2.2. Feature Selection

We have reduced the features from 22 to 10, by dropping the features, which describe the crime itself, such as case number, arrest, and description of the crime, since they only known after the occurrence of the crime and keeping the spatial features since the classification model relies on them. The selected features were: Block, Location Description, Domestic, Beat, District Ward, Community Area, X Coordinates, Y Coordinates Primary Type is the target variable. Since the number of the classification labels in the original dataset was too high, and several crime categories can be labeled under four main categories, so we reduced them as follow for better prediction.

- **Forbidden practices:** Narcotics, prostitution, gambling, obscenity, and other narcotic violation.
- **Theft:** Burglary, deceptive practice, motor vehicle theft, and robbery.
- **Assault:** Crime sexual assault, offense involving children, sex offense, homicide, and human trafficking.
- **Public peace violation:** Weapons violation, criminal defacement, criminal trespass, arson, kidnapping, stalking, intimidation, public indecency.

6. RESULTS AND DISCUSSION

This section demonstrates the applied feature selection and shows the results of the proposed techniques.

6.1. Results of Investigating the Effect of Feature Selection on the Dataset

In feature selection, the features were selected by applying backward feature selection. Initially, the number of features was ten. They were ranked based on Information Gain Attribute Evaluation algorithm, then the attribute with the least rank was dropped until left with 9 attributes since the accuracy began to be reduced.

Table 1 shows the accuracies of features subset for both Naïve Bayes and Decision Tree.

Table 1: Accuracies of Different Features Subset

Using features	Na ïve Bayes	Decision Tree
All	83.40 %	91.68 %
Top 9 features	83.34 %	91.58 %
Top 8 features	82.82 %	90.71 %
Top 7 features	82.60 %	88.66%

As shown in the table above, Decision tree classifier gives better accuracy than Naive Bayes with 9 features and same accuracy was achieved with all the features. Therefore 9 features were selected since it consumes fewer resources. For Naïve Bayes highest accuracy was achieved with all the features.

Table 2 summarizes the results for Correctly Classified Instances (CC), Accuracy (AC), Receiver Operating Characteristic Curve (ROC), Precision, and Recall for Naïve Bayes and Decision Tree.

Table 2: Results with the selected number of features

	Na ïve Bayes	Decision Tree
CC	7063	7763
AC	83.33 %	91.59 %
ROC	0.969	0.976
Precision	0.851	0.926
Recall	0.833	0.916

- **In Decision Tree:**
  - 2033 instances from class “Theft” were classified correctly and 70 instances were classified incorrectly.
  - All the instances belong to class “Public Peace Violation” were classified correctly.
  - 1936 instances from class “Forbidden Practices” were classified correctly and 160 instances were classified incorrectly.

- 1584 instances from class “Assault” were classified correctly and 557 instances were classified incorrectly.

- **In Naïve Bayes:**
  - 1949 instances from class “Theft” were classified correctly and 145 instances were classified incorrectly.
  - 1688 instances from class “Public Peace Violation” were classified correctly and 448 instances were classified incorrectly.
  - 2024 instances from class “Forbidden Practices” were classified correctly and 72 instances were classified incorrectly.
  - 1402 instances from class “Assault” were classified correctly and 739 instances were classified incorrectly.

7. CONCLUSION

This study aimed to predict the category of crimes using “City of Chicago from 2013 to 2017” dataset. The techniques that were used in our study were Naïve Bayes and Decision Tree. In the preprocessing stage, missing values were handled using listwise deletion. Feature Selection was performed using backward feature selection technique. Decision Tree classifier performance was better than Naïve Bayes, its prediction accuracy reached 91.59 %. However, Naive Bayes accuracy is 83.40 %. Therefore, comparing the two results, Decision Tree outperformed Naïve Bayes.As part of future work, though Decision Tree is giving high accuracy, we could improve it using ensemble methods or using different feature selection schemas to get higher accuracy.

8. REFERENCES

[1] R. T.V, J. Joshy, M. R and A. Soni M, *“Crime Prediction and Analysis using Clustering Approaches and Regression Methods”*, International Journal of Computing and Technology, vol. 5, no. 4, 2018.

[2] R. Kiani, S. Mahdavi and A. Keshavarzi, *“Analysis and Prediction of Crimes by Clustering and Classification”,* International Journal of Advanced Research in Artificial Intelligence, vol. 4, no. 8, 2015.

[3] A. K. Singh, N. Prasad, N. Narkhede, and S. Mehta, “*Crime: Classification and Pattern Prediction”,* Int. Adv. Res. J. Sci. Eng. Technol., vol. 3, no. 2, pp. 2393–2395, 2016.

[4] H. B. Fredrick David and A. Suruliandi, *“Survey on Crime Analysis and Prediction using Data Mining Techniques”,* ICTACT J. Soft Comput., vol. 7, no. 3, pp. 1459–1466, 2017.

[5] J. Agarwal, R. Nagpal and R. Sehgal, *“Crime Analysis using K-Means Clustering”,* International Journal of Computer Applications, vol. 83, no. 4, pp. 1-4, 2013.

[6] S Prabakaran and Shilpa Mitra, "Survey of Analysis of Crime Detection Techniques Using Data Mining and Machine Learning" 2018 J. Phys.: Conf. Ser. 1000 012046.

[7] K. Taha et al., *“Analysis and Prediction of Crimes by Clustering and Classification,”* Int. J. Adv. Res. Artif. Intell., vol. 11, no. 4, pp. 1–15, 2015.

[8] S. Sathyadevan, D. S and S. S., *“Crime Analysis and Prediction using Data Mining”,* 2014 First International Conference on Networks & Soft Computing (ICNSC2014), 2014.

[9] K. Bogahawatte and S. Adikari, “*Intelligent Criminal Identification System”,* 2013 8th International Conference on Computer Science & Education, 2013.

[10] A. Chandrasekar, A. S. Raj and P. Kumar, *"Crime Prediction and Classification in San Francisco City".*

[11] A. Babakura, M. Sulaiman and M. Yusuf, *“Improved Method of Classification Algorithms for Crime Prediction”,* 2014 International Symposium on Biometrics and Security Technologies (ISBAST), 2014.

[12] M. Sharma, *“Z - CRIME: A Data Mining Tool for the Detection of Suspicious Criminal Activities based on Decision Tree”,* 2014 International Conference on Data Mining and Intelligent Computing (ICDMIC), 2014.

[13] R. Iqbal, M. Murad, A. Mustapha, P. Panahy and N. Khanahmadliravi, *"An Experimental Study of Classification Algorithms for Crime Prediction",* Indian Journal of Science and Technology, vol. 6, 2013.

[14] A. Bharati and S. RA. K, *“Crime Prediction and Analysis using Machine Learning”*, International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 9, 2018.

[15] P. Gera and R. Vohra, *“Predicting Future Trends in City Crime using Linear Regression”,* vol. 14, no. 07, pp. 58–64, 2014.

[16] P. Vrushali, M. Trupti, G. Pratiksha, and G. Arti, *“Crime Rate Prediction using KNN”*, Int. J. Recent Innov. Trends Comput. Commun., vol. 6, no. 1, pp. 124–127, 2018.

[17] T. Patil and S. Sherekar, *“Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification”,* International Journal of Computer Science and Application, vol. 6, no. 2, pp. 256,257, 2013.

[18] W. Dai and W. Ji, *“A MapReduce Implementation of C4.5 Decision Tree Algorithm”,* International Journal of Database Theory and Application, vol. 7, no. 1, pp. 49-60, 2014.

[19] G. Kaur and A. Chhabra, *“Improved J48 Classification Algorithm for the Prediction of Diabetes”,* International Journal of Computer Applications, vol. 98, no. 22, pp. 13-17, 2014.

[20] *"Chicago Crime",* Kaggle.com, 7 April 2018. [Online]. Available: https://www.kaggle.com/chicago/chicago-crime. [Accessed: 25- Sept- 2019].