A Proposed Approach for Crime Type Prediction using Machine Learning Techniques

Suham adnan Abdul alkareem
Baghdad University, Baghdad, Iraq
Seham.adnan@uobaghdad.edu.iq
Orcid.org/0000-0002-2584-9946

Kadhim B. S. Aljanabi
Ashur University, Baghdad, Iraq
kadhim.aljanabi@au.edu.iq

Salam Alaugby
Faculty of Computer Science and Mathematics University of Kufa
salam.alaugby@uokufa.edu.iq

DOI: http://dx.doi.org/10.31642/JoKMC/2018/120105
Received Jan. 9, 2025. Accepted for publication Mar. 22, 2025

Abstract**—** *Crime and suspect prediction represent an immerging field where machine learning can be useful when applied efficiently. The available data can be used to train the proposed prediction model (building the model, classifier for example) and then this model can be tested and used to predict the crime type and suspect information such as sex, race and age category. The work in this paper presents a proposed approach for such prediction by using real world dataset available on the internet. The gathered data were preprocessed(deletion of the rows with missing and unknown attribute content, converting some attribute values into nominal or categorical data and concept hierarchy techniques) and reduced the number of attributes by choosing the most important features using different approaches and algorithms, then different Decision Tree types, Naïve Bayes Classifier and Association Rules prediction techniques were used to create the required models and to find out associations between different attributes of the given data. Testing phase shows high efficiency and effectiveness of the proposed approaches which in turn provide models that can be used as reliable predictors. An accuracy of more than 85% was achieved when using different classification techniques and the decision tree an accuracy of more than 85% due to its ability to classify the data based on important features, Naïve Bayes is less accurate than the decision tree in some cases, but usually achieves good accuracy ranging from 75% to 85% depending on the data. The decision tree is a powerful algorithm for extracting features that separate classes well, which leads to higher accuracy the Naïve Bayes uses probabilities to extract features but assumes independence of features, which may limit its accuracy in some cases. WEKA software, Microsoft Excel and XLSTAT mining software were used to analyze the given data.*

Keywords— Machine learning, Classification, Decision Tree, Association Rule Mining.

I. INTRODUCTION

Data Mining (DM), also known as Knowledge Discovery from Databases (KDD), refers to the process of extracting valuable knowledge from large datasets. It involves uncovering hidden insights, unexpected patterns, and new rules within data. Data mining consists of two main components: discovery and exploitation. In the discovery phase, useful facts are identified and represented as information, while in the exploitation phase, these facts are used to solve specific problems by creating models and applying them to new, unknown data.

The process of data mining involves several key steps: problem formulation, data evaluation, feature extraction and enhancement, prototyping, and model evaluation. A basic classification of knowledge discovery techniques includes

manual search, OLAP (Online Analytical Processing), knowledge engineering, data visualization, automated search methods, auto-clustering, link analysis, regression, and rule induction. These techniques help in discovering and utilizing patterns to make informed decisions or predictions [1, 2,3,4].

Data mining is a rapidly growing field with a wide range of applications across various industries, such as marketing, banking, city planning, health insurance, and many others that have a significant impact on society. One important application of data mining is crime analysis. The field

encompasses several key tasks and techniques, including classification, association, clustering, prediction, and link analysis. Each of these techniques plays a vital role in solving different types of problems and is applied in various contexts to uncover valuable insights. [5,6,7,8].

Crime analysis involves examining criminal activities by breaking down acts that violate laws to understand their nature and reporting the findings. The primary goal of crime analysis is to extract meaningful insights from large volumes of data and share this information with law enforcement, including officers and investigators, to aid in apprehending criminals and reducing criminal activity. For instance, analysis may reveal details such as the suspect's age group, race, and sex. Additionally, crime analysis plays a crucial role in crime prevention by identifying patterns and trends that can inform proactive strategies. [5, 9, 10]. Preventing crime is more cost-effective than attempting to apprehend criminals after the crime has occurred. Crime analysis can be defined as a systematic process focused on providing timely and relevant information about crime patterns and trend correlations. This information helps both operational and administrative personnel plan the strategic deployment of resources to prevent and reduce criminal activities.

The reasons behind conducting crime analysis can be summarized as follows:

1- To provide law enforcement with up-to-date information on general and specific crime trends, patterns, and series in a timely manner.
2- To analyze crime data in order to leverage the vast amount of information available within law enforcement agencies, the criminal justice system, and the public domain. process is divided into the steps shown in figure (1).

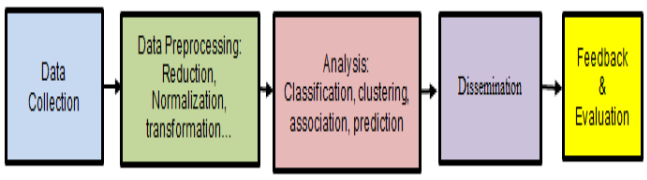


Fig. 1. Crime Analysis Process

1. Classification techniques

Among different classification algorithms, Decision Tree (DT) and Naïve Bayes algorithms represent the most popular and commonly used data mining algorithms in different fields and applications.

Naïve Bayes classification algorithm is based mainly on Byes theorem, Bayes' theorem is the also known as **Bayes' Rule** or the **Bayes' law,** which is used to determine the probability of a hypothesis with prior of knowledge. It depends on conditional the probability. The formula for the Bayes' theorem is Given as:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \qquad \dots (1)$$

Where,

P(A|B) is the posterior probability, which represents the probability of hypothesis A given the observed event

B.P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is the Prior probability: Probability of a hypothesis before observing evidence.

P(B) is the Marginal probability: Probability of the Evidence.

Apriori algorithm represent the familiar association rule technique, it depends mainly on defining the factors Support and confidence [21,22].

Association rule, the mining consists of 2 steps:

1. find all frequent itemset.
2. generate the association rules, from above frequent itemset.

Types of crime analysis:

The Crime of analysis can be classified, into following [5, 15,16,17].

- OPERATIONS ANALYSIS
- TACTICAL THE CRIME OF ANALYSIS
- INVESTIGATIVE CRIME OF ANALYSIS
- STRATEGIC CRIME ANALYSIS
- ADMINISTRATIVE CRIME ANALYSIS
- INTELLIGENCE, ANALYSIS

Both classification techniques (Decision Tree for example) and Association Rule mining are of great importance in crime detecting field.

Table(I) a comparison of Literature review on crime prediction using machine learning techniques

number	Study name	Authors	The year	The approach used	Accuracy and results
1	Advanced Crime Prediction and Analysis Using Machine Learning and Quantum Networking	Jeremy Gideon J., J. Jefrin, S. Dhamodaran	2024	Analysis of 150 studies on crime prediction using deep learning and machine learning	A comprehensive review of different approaches to crime prediction
2	An Analytical Comparison of Crime Prediction Using Machine Learning Techniques	A. Gangwar, Deepak Singh Bisht, S. Choudhary, Vivek Chauhan, Vivek Tomar	2024	Comparison between Polynomial Regression, Random Forest, LightGBM, XGBoost	XGBoost was found to be the best in accuracy.
3	Predictive Analytics for Crime Prevention in Smart Cities Using Machine Learning	Ponugoti Kalpana, Sarangam Kodati, K. Adi Narayana Reddy, Hassan Ali, AC Ramachandra	2024	SVM based model with RBF Kernel	Up to 89% accuracy in predicting crimes
4	A Comparative Analysis of Machine Learning Algorithms for Crime Rate Prediction	A.D. Dileep, K. Ramalakshmi, R. Venkatesan, G. Naveen Sundar, Golden Nancy, S. Shirly	2024	Comparison between Logistic Regression, Naïve Bayes, KNN, Decision Tree, Random Forest	Contribution to reducing crimes against women in India

Crime prediction results comparison the results from previous studies as well as the new study indicate a consistent trend in the use of machine learning techniques for crime prediction,

with a strong focus on classification models. Study 1 uses deep learning and quantum networks to analyze crime prediction, while the proposed study focuses on decision trees and Bayesian classifiers, achieving an accuracy of over 85%. Study 2 compares different machine learning techniques, identifying XGBoost as the most accurate. This is consistent with the decision tree approach in the proposed study, which also emphasizes feature selection to improve classification performance. Study 3 uses an SVM-based model with an RBF kernel, achieving an accuracy of up to 89%, slightly outperforming the decision tree results in the proposed study. Study 4 compares multiple algorithms, including naive Bayes, random forests, and decision trees, similar to the methods in the proposed research. However, it also explores crime rates associated with specific demographics. While the proposed study focuses on feature extraction and model optimization, these studies highlight broader applications and comparative analysis of multiple algorithms. Overall, the results suggest that decision trees and ensemble methods such as XGBoost tend to provide high accuracy, enhancing the effectiveness of structured feature selection and model improvement in crime prediction tasks.

II. DATA PREPROCESSING

Data cleaning and data reduction where applied on the collected data to get clean, complete and minimized dataset from the given data. This will improve the analysis process [15,18,19,20]. Removing the records with missing values and reducing the attribute values through the use of concept hierarchy and converting the data into categories are the main processes applied in this phase. The data were reduced from more than one million records into about 227,000 records with 14 most important attributes shown below:

Crime ID, Crime Name, Offence, Crime type, City, Crime Date, Criminal ID, Criminal Gender, Criminal Age, Criminal race, Victim ID, Victim Gender, Victim race and victim age. The resultant data were grouped into three categories; Crime attributes, Suspect attributes and Victim attributes. Some of the attributes mentioned above were extracted because they have lower effects on the defined models.

III. PROPOSE APPROACH

Start by providing a brief overview of the proposed approach. Explain the problem or need that the approach addresses and why it's relevant. This will set the stage for understanding the purpose of the different phases. The proposed approach consists of the following phases shown the figure (2)

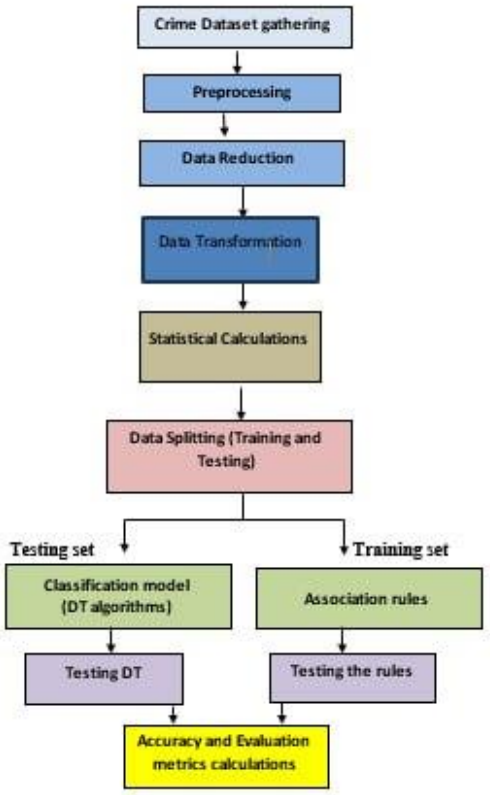


Figure (2). Block Diagram for the Proposed Approach.

Phase 1: Data gathering

In this phase, dataset available on the internet with more than 6 million records and more than thirty attributes was collected.

Phase 2: Phase 2: Data preprocessing

In this phase the collected data was reduced through the deletion of the records with missing and unknown attribute values and some selected attributes where chosen. Sample of the preprocessed data is shown in table(I). More than 200000 records with eight most important attributes were taken as samples. The selected attributes represent that attribute suitable for classification, clustering and finding out associations between different attribute values [4,5,20,21,22].

Table (II). Sample of the Collected Crime Dataset After Preprocessing Phase

Crime Date	Crime key	Offence	City	Suspect Age Group	Suspect Race	Suspect Sex	Victim Age Group	Victim Race	Victim Sex
12/31/2011	353	UNAUTHORIZED USE OF A VEHICLE	QUEENS	45-64	BLACK	M	65+	BLACK	F
12/31/2011	353	UNAUTHORIZED USE OF A VEHICLE	BROOKLYN	18-24	BLACK	M	45-64	WHITE	M
12/31/2011	353	UNAUTHORIZED USE OF A VEHICLE	BROOKLYN	18-24	BLACK	M	45-64	WHITE	M
12/31/2011	105	ROBBERY	BROOKLYN	45-64	WHITE	M	65+	WHITE	M
12/31/2011	355	OFFENSES AGAINST THE PERSON	BROOKLYN	25-44	BLACK	M	25-44	BLACK	F
12/31/2011	361	OFFENSES	MANHATTAN	<18	BLACK	F	45-64	BLACK	F
12/31/2011	361	OFFENSES	QUEENS	18-24	WHITE HISPANIC	F	25-44	WHITE HISPANIC	F
12/31/2011	361	OFFENSES	BRONX	25-44	WHITE HISPANIC	M	25-44	WHITE HISPANIC	F
12/31/2011	361	OFFENSES	BRONX	65+	BLACK	M	45-64	WHITE	F
12/31/2011	361	OFFENSES	BROOKLYN	25-44	WHITE HISPANIC	M	25-44	WHITE HISPANIC	F

Phase 3: Statistical analysis of the collected data

In this phase and in order to have a general knowledge about the behavior and the effect and weight of each attribute values,

a general statistic for these values were carried as show in the figures (3), 4 and 5.

Figure (3) Shows That Maximum Frequencies of The Crimes Are the Crimes with Codes 27, 18, And 25 (Harazment2, Assault 3 & Related Offenses, And Offenses Agnst Pub Ord Sensblty) With Occurrences 86904, 42057, And 38003 Respectively

Fig. 3. Frequency (Occurrence) of Each Crime

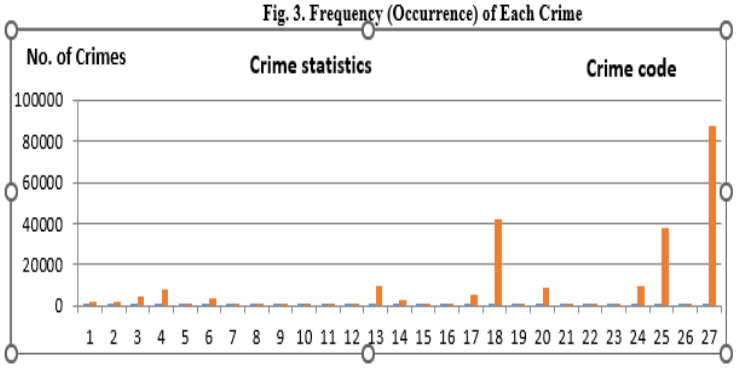
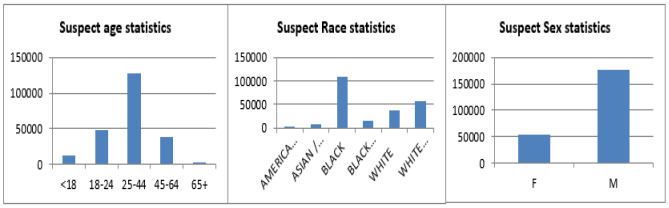
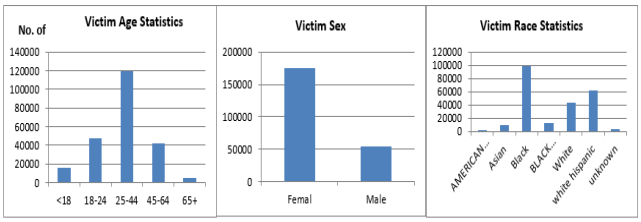


Figure (4). Suspect Statistics According to Age, Race and Gender.



The suspect statistics show that highest crime frequencies occur for black race, age category between 25 and 44 years, and suspect sex is male.

Figure (5). Victim Statistics According to Age, Race and Gender.



The victim statistics show that highest crime frequencies occur for black race, age category between 25 and 44 years, and victim sex is female.

Phase 4. Naïve Bayes Approach

Naive Bayes algorithm is a probabilistic classification model based on Bayes' theorem, and assumes that all features are independent of each other, when given a target class. This algorithm is widely used in classification and prediction, especially in natural language processing and big data analysis, The Naive base algorithm enhances the performance of the decision tree and is a strong complement to it, including crime prediction based on the data in the following table (III) Predicted Offense (Naïve Bayes): This column shows the crime type predicted using Naïve Bayes Classification [20,21].

Table (III) Crime Data Prediction Using Naïve Bayes in XLSTAT

Crime Date	Crime Key	Predicted Offense (Naïve Bayes)	City	Suspect Age Group	Suspect Race	Suspect Sex	Victim Age Group	Victim Race	Victim Sex	Prediction Probability
12/31/2011	353	UNAUTHORIZED USE OF VEHICLE	QUEENS	45-64	BLACK	M	65+	BLACK	F	87%
12/31/2011	353	UNAUTHORIZED USE OF VEHICLE	BROOKLYN	18-24	BLACK	M	45-64	WHITE	M	79%
12/31/2011	105	ROBBERY	BROOKLYN	45-64	WHITE	M	65+	WHITE	M	82%
12/31/2011	355	OFFENSES AGAINST THE PERSON	BROOKLYN	25-44	BLACK	M	25-44	BLACK	M	90%
12/31/2011	361	OFFENSES	MANHATTAN	<18	BLACK	F	45-64	BLACK	F	76%
12/31/2011	361	OFFENSES	QUEENS	18-24	WHITE HISPANIC	F	25-44	WHITE HISPANIC	F	80%
12/31/2011	361	OFFENSES	BRONX	25-44	WHITE HISPANIC	M	25-44	WHITE HISPANIC	F	85%
12/31/2011	361	OFFENSES	BRONX	65+	BLACK	M	25-44	WHITE HISPANIC	F	73%

- Predicted Offense (Naïve Bayes): This column shows the crime type predicted using Naïve Bayes Classification.
- Prediction Probability: The confidence level of the prediction based on Bayes' Theorem probabilities.
- Analysis of Influencing Factors:
- Suspect and victim attributes (age, race, sex) are used as predictors for crime type.
- Past crime patterns are analyzed to extract frequencies and correlations between variables.

Phase 5. Decision Tree Approach

Different Decision Tree Approaches were applied on suspect data to understand the behavior of the suspect [21,22]. Three types of such trees were applied including suspect Race, Sex and Age as shown in figure (6) (a), (b), and (c) respectively. It is very well known that Entropy and Information Gain play crucial factors in building Decision Trees. Calculating Entropy and Information Gain is show in equations 2 and 3. For the dataset, that has C classes and a probability of the randomly choosing, data from class, is Pi. Then entropy E(S) can be the mathematically represented as [22]:

$$E(s) = -\sum_{i=1}^{c} P_i \ Log_2(P_i) \quad \dots (2)$$

And the Information Gain (IG) is given by:

$$IG = Entropy_{parent} - Entropy_{children} \quad \dots (3)$$

Decision Tree is Easy to interpret, handles both numerical and categorical data [21].

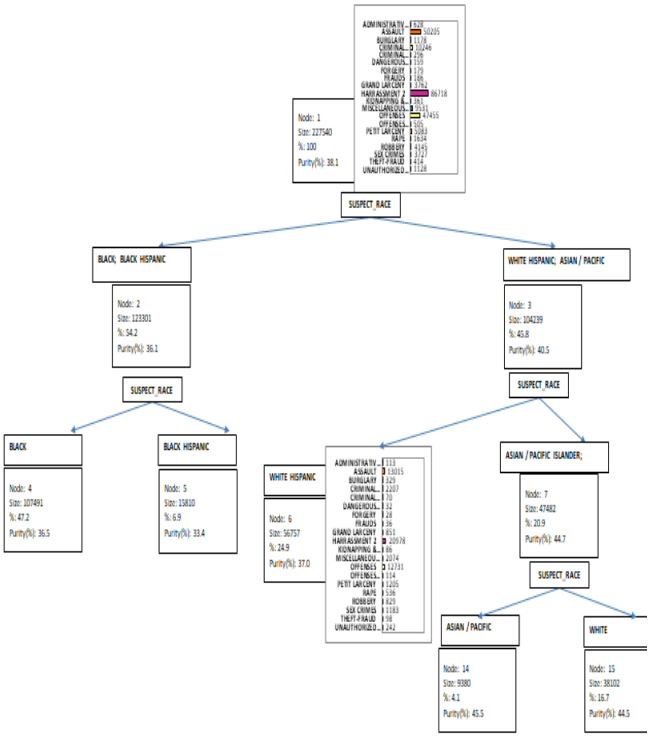


Figure (6) (a). Decision Tree for Suspect Race.

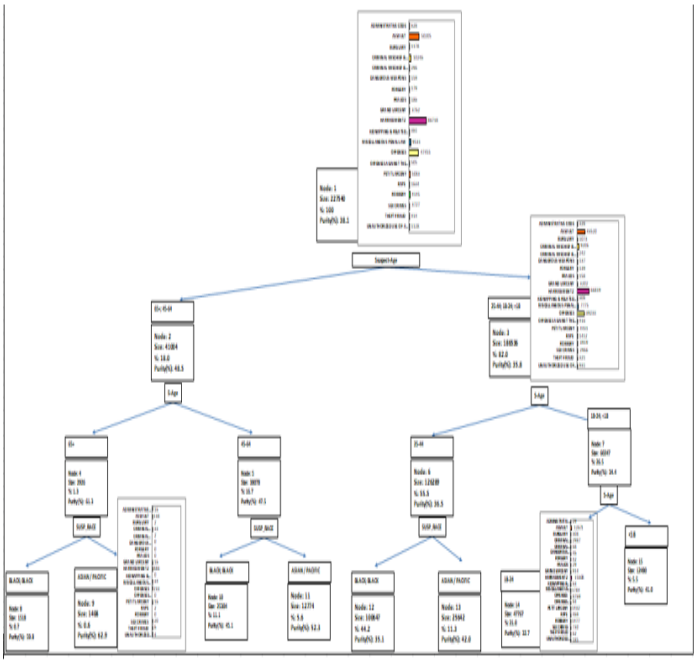


Figure (6) (c). Decision Tree for Suspect Age.

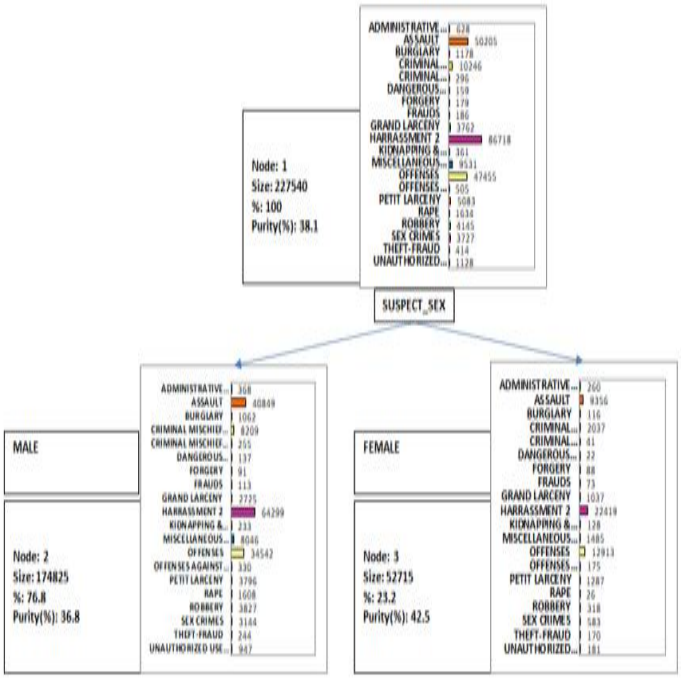


Figure (6) (b). Decision Tree for Suspect Sex.

Phase 6:

Association Rule Mining. In Crime analysis, mining association rules represents one of the most important techniques in crime analysis since it shows the relationships between of crime and the different attributes in the dataset. In this work a minimum support and the minimum confidence of 0.05 were taken into consideration.

Support(X)= (Number of transactions containing X)/ (Total number of transactions) … (4)

Confidence (X => Y) = (Number of transactions containing X and Y) / (Number of transactions containing X) … (5)

And two another the factors Lift and conviction. Lift is a measure that tells us whether the probability of an event B increases or decreases given event A, equation (6) shows the calculation of Lift factor [22].

Lift(A→B) = Confidence(A→B) / Support(B) …. (6)

conviction is another way to measure, the association which Compares probability that A appears without B if they were independent vs, actual the frequency of A's appearance without B, equation (7) shows the calculation of the conviction factor [21].

Conviction(A→B) = (1 - Support(B)) / (1 - Confidence(A→B)) …(7)

Association rules mining is highly dependent on support and confidence of the items in the dataset. Equation 4 and 5 give the calculation of these two factors [22].

Table (IV) Summary of association rules

Antecedent	Consequence	Confidence	Support	Lift
ASSAULT	BLACK	0.522	0.115	1.105
HARRASSMENT 2	BLACK	0.452	0.172	0.957
WHITE	HARRASSMENT 2	0.445	0.075	1.169
OFFENSES	BLACK	0.410	0.086	0.868
WHITE HISPANIC	HARRASSMENT 2	0.370	0.092	0.970
BLACK	HARRASSMENT 2	0.365	0.172	0.957
OFFENSES	WHITE HISPANIC	0.268	0.056	1.076
ASSAULT	WHITE HISPANIC	0.259	0.057	1.039
BLACK	ASSAULT	0.244	0.115	1.105
HARRASSMENT 2	WHITE HISPANIC	0.242	0.092	0.970
WHITE HISPANIC	ASSAULT	0.229	0.057	1.039
WHITE HISPANIC	OFFENSES	0.224	0.056	1.076
HARRASSMENT 2	WHITE	0.196	0.075	1.169
BLACK	OFFENSES	0.181	0.086	0.868

Table (V). Matrix of items

Antecedent	ASSAULT	BLACK	HARRASSMENT 2	OFFENSES	WHITE	WHITE HISPANIC
ASSAULT	0	0.383	0.000	0.000	0.000	0.244
BLACK	0.383	0	0.409	0.295	0.000	0.000
HARRASSMENT 2	0.000	0.409	0	0.000	0.321	0.306
OFFENSES	0.000	0.295	0.000	0	0.000	0.246
WHITE	0.000	0.000	0.321	0.000	0	0.000
WHITE HISPANIC	0.244	0.000	0.306	0.246	0.000	0

The influence charts shown in figures (7) and (8) ensures the results mentioned previously where the maximum occurrence of the crimes Assault, Harrassment2 and Offence are with suspect race Black and suspect sex is Male.

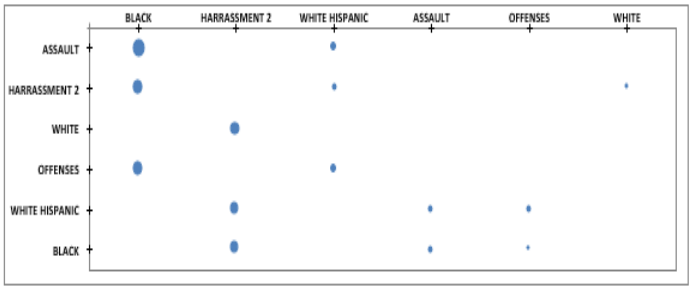


Figure (7). Influence Chart.

Table (VI) Summary of association rules

Antecedent	Consequence	Confidence	Support	Lift
ASSAULT	M	0.814	0.180	1.059
HARRASSMENT 2	M	0.741	0.283	0.965
OFFENSES	M	0.728	0.152	0.947
F	HARRASSMENT 2	0.425	0.099	1.116
M	HARRASSMENT 2	0.368	0.283	0.965
OFFENSES	F	0.272	0.057	1.175
HARRASSMENT 2	F	0.259	0.099	1.116
F	OFFENSES	0.245	0.057	1.175
M	ASSAULT	0.234	0.180	1.059
M	OFFENSES	0.198	0.152	0.947

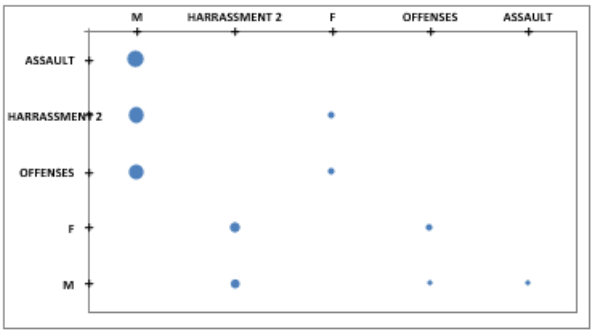


Figure (8). Influence Chart.

IV. RESULTS ANALYSIS

It is clear from figure (4) that the most frequent crimes are related to suspect age between 25 and 44, suspect race is black and male, and the most frequent crimes are related to victim age between 25 and 44, victim race is black and female, which appear clearly in table (II) and Table (III). Decision Trees shown in figure (6) give statistics for the most frequent crimes (assault, harassment 2 and offenses) according to the race, age and sex of the suspect respectively. The influence charts shown in figures (7) and (8) ensures the results mentioned previously where the maximum occurrence of the crimes Assault, Harrassment2 and Offence are with suspect race Black and suspect sex is Male.

V. RESULTS AND DISCUSSION

The results in Figure 4 clearly indicate that the majority of crimes are committed by suspects between the ages of 25 and 44. This finding is consistent with the data in Tables II and III, which show that the most common suspects are male, Black, and in the median age range of 25 to 44. These demographic trends are important because they help identify potential risk factors and patterns of criminal activity in specific age and racial groups. The results revealed that the most frequent victims are also in the 25 to 44 age range, with female Black victims being overrepresented. This is consistent with the general trends observed in the suspect data, which may indicate a potential relationship between suspect race and age and violence. Figure 6 shows a decision tree analysis, which provided deeper insights into the relationship between suspects (age, race, and gender) and the types of crimes committed (assault, harassment, and felonies). The decision tree classifies the most common crimes committed by black male suspects, ages 25 to 44, as shown in Figures 7 and 8. These numbers provide further evidence that suspect demographic characteristics play an important role in predicting crimes.

VI. CONCLUSIONS

Both classification techniques and association rule mining are of great importance in predicting the crime type and the suspect information including sex, age and race which in turn have great importance in helping the police authority. From the results obtained it is clear that most of the crimes were done by male, middle age suspects (25-44 years) and black race, whereas other attributes have low occurrences.

First of all, the raw data were collected and preprocessed to get clean, complete and smaller size so as to optimize the classification (Decision Tree) and association rule mining techniques. It was very important to use different data reduction techniques to reduce the data into about 227000 records such as concept hierarchy and other techniques.

Decision Tree algorithms were applied to classify the preprocessed data from which the conclusion mentioned above is obtained. Association rule mining using Apriori algorithm was also used to, find the Relationships between different suspect of attributes and the crimes, the results showed that the higher support and confidence were obtained with male, middle age and black suspects, as shown clearly in table II, III and IV and the figures 6 to 8.

VII. REFERENCES

[1] Agarwal, Shivam. "Data mining: Data mining concepts and techniques." 2013 international conference on machine intelligence and research advancement. IEEE, 2013.

[2] Hartama, Dedy, Agus Perdana Windarto, and Anjar Wanto. "The Application of Data Mining in Determining Patterns of Interest of High School Graduates." Journal of Physics: Conference Series. Vol. 1339. No. 1. IOP Publishing, 2019.

[3] García, Salvador, et al. "Big data preprocessing: methods and prospects." Big Data Analytics 1.1 (2016): 1-22.

[4] N. Shah, N. Bhagat, and M. Shah, "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention," 2021. ncbi.nlm.nih.gov

[5] P. Stalidis, T. Semertzidis, and P. Daras, "Examining Deep Learning Architectures for Crime Classification and Prediction," 2018. [PDF]

[6] U. Muneer Butt, S. Letchmunan, F. Hafinaz Hassan, and T. Wei Koh, "Hybrid of deep learning and exponential smoothing for enhancing crime forecasting accuracy," 2022. ncbi.nlm.nih.gov

[7] M. Aminur Rab Ratul, "A Comparative Study on Crime in Denver City Based on Machine Learning and Data Mining," 2020. [PDF]

[8] Bokaba, Tebogo, Wesley Doorsamy, and Babu Sena Paul. "Comparative study of machine learning classifiers for modeling road traffic accidents." Applied Sciences 12.2 (2022): 828.

[9] K. Jenga, C. Catal, and G. Kar, "Machine learning in crime prediction," *Journal of Ambient Intelligence and Humanized Computing*, 2023. springer.com

[10]https://pmc.ncbi.nlm.nih.gov/articles/PMC10990207/April 3,2024.

[11] Alamro, Reham, and Abdou Youssef. "Impact of data reduction techniques on classification." 2018 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, 2018

[12] RM Aziz, A Hussain, P Sharma, P Kumar - Karbala International Journal of …, 2022 - iasj.net. Machine learning-based soft computing regression analysis approach for crime data prediction. iasj.net
Cited by **43**

[13] S. Hussain, R. Atallah, A. Kamsin, and J. Hazarika, "Classification, clustering and association rule mining in educational datasets using data mining tools: A case study," in Advances in Intelligent Systems and Computing, 2019, vol. 765, pp. 196–211.

[14] B. B. Miftachul Huda, Andino Maseleno, Pardimin Atmotiyoso, Maragustam Siregar, Roslee Ahmad, Kamarul Azmi Jasmi, Nasrul Hisyam nor Muhamad, Mohd Ismail Mustari, "Big Data Emerging Technology: Insights into Innovative Environment for Online Learning Resources," Int. J. Emerg. Technol. Learn., vol. 13, no. 1, pp. 23–36, 2018.

[15] R. Ari Setyawan, E. Prasetyo, and A. S. Girsang, "Design and Implementation Data Warehouse in Insurance Company," Journal of Physics: Conference Series, vol. 1175, no. 1. 2019.

[16] M. Subekti, J. Junaidi, H. L. H. S. Warnars, and Y. Heryadi, "The 3 steps of best data warehouse model design with leaning implementation for sales transaction in franchise restaurant," 2017 IEEE International Conference on Cybernetics and Computational Intelligence, Cybernetics COM 2017 - Proceedings, vol. 2017-Novem. pp. 170–174, 2018.

[17] K. B. S. Aljanabi, Haider Alsharqi, "An Optimal Warehouse Design for Crime Dataset," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 5, pp. 9080–9088, 2020.

[18] X. Wu et al., "Top 10 algorithms in data mining," Knowl Inf Syst, vol. 14, pp. 1–37, 2008.

[19] https://www.barnesandnoble.com/w/data-mining-jiawei-han/1141054718 october3,2022-data mining.

[20] H. K. Fatlawi, A. F. H. Alharan, and N. S. Ali, "An efficient hybrid model for reliable classification of high dimensional data using k-means clustering and bagging ensemble classifier," J. Theor. Appl. Inf. Technol., vol. 96, no. 24, pp. 8379–8398, 2018

[21]. Himani Sharma , Sunil Kumar"A Survey on Decision Tree Algorithms of Classification in Data Mining", April 2016, International Journal of Science and Research (IJSR) 5(4).

[22] https://journal.trunojoyo.ac.id/edutic/article/download/28150/10268