ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

COMBATING CRIME AGAINST WOMEN IN INDIA: ANALYZE THE CAUSES AND CONSEQUENCES

Anjum A. Patel[1] and Dr. Pooja Kulkarni[2]

[1]Research Scholar, Department of Computer Science, Faculty of Science and Technology, Vishwakarma University, Pune Maharashtra, India
anjum.patel-015@vupune.ac.in

[2]Assistant Professor, Department of Computer Science, Faculty of Science and Technology, Vishwakarma University, Maharashtra, India
pooja.kulkarni@vupune.ac.in

ABSTRACT

Crime Against Women (CAW) is a major concern in India. Physical assault is the most common form of violence against women in India, followed by emotional and sexual abuse. The National Crime Records Bureau reported 15.3% more crimes against women in 2021 than 2020. Rape, domestic abuse, dowry violence, and human trafficking are crimes against women. Federal and state legislation and initiatives have not reduced crimes against women. Proper reporting of rape and assault cases, efficient law enforcement agencies, exemplary punishment, zero tolerance, an effective Indian Police System, legislative awareness among women, and appropriate training for women are needed to combat crime against women in India. To eliminate or reduce this issue, several steps are necessary. News routinely exposes child abuse and women's harassment. Crime against women and children has escalated. Women face cultural, regional, national, and communal violence. Violence against women is physical, sexual, psychological, and economic. Delhi police reported 2,155 rapes in 2016, quadrupling from 572 in 201 [1].Indian Express reported 34,651 rapes in 2015[2]. With total events recorded in India, megacities had a higher crime rate than domain states. Metropolitan Pune is in Maharashtra. It has 15.642 km2 and 6,975,000 people. Our prediction model analysis of 30 regions (police stations) can be applied to all megacities in India and help lawmakers refine the 'Women safety' policy for the entire nation.

Study introduces CAW analysis. The work should use statistical and machine learning methods. Overcrowding, poverty, and ignorance cause women's crimes. To reduce crime, programs, seminars, and campaigns should raise awareness. Chance-rich cities should promote evening women's safety. Old procedure laws are weak. The accused frequently gets anticipatory bail or adjournment. Slows court rulings. Address legislative implementation and improvements. Therefore, it will improve protections for women and our civilized nation. This paper shows the various standard as well as advanced techniques / methodologies used to curb CAW.

This study analyses conventional and advanced Customer Attrition Wastage (CAW) methods. The study uses exploratory data, association rule mining (ARM), machine learning models, and inferential analysis. 'Eli5' is used to improve CAW analysis in a new model. K-Nearest Neighbor, Logistic Regression, and Random Forest classifiers were compared. The Random Forest (RF) classifier surpasses K-nearest neighbor (KNN) and Logistic Regression (LR) in the study's extensive testing. The three classification models were assessed using 'Feature Importance'. CAW scenario understanding and model improvements may depend on feature relevance.

Keywords: *CAW, physical assault, regression, K-Nearest Neighbor, Logistic Regression, and Random Forest classifiers, ARM, Legislative awareness, Feature Importance, Model enhancements.*

INTRODUCTION

Addressing the pervasive issue of Crime Against Women (CAW) in India necessitates a comprehensive approach that considers various aspects of sexual violence and physical assault. The National Crime Records Bureau has observed a troubling 15.3% rise in reported crimes against women in 2021, despite the implementation of legal measures and programs at both federal and state levels, as compared to the previous year. The prevalence of rape,

domestic abuse, dowry violence, and human trafficking has experienced a concerning increase in recent years, necessitating immediate intervention.

In order to deal with and reduce this issue, a number of measures must be taken, such as ensuring accurate reporting of events, establishing efficient law enforcement agencies, imposing severe penalties on perpetrators, adopting a policy of absolute intolerance towards crimes against women, enhancing the effectiveness of the Indian Police System, and promoting legislative awareness among women. These steps are crucial to tackle the intricate and widespread issue of Crime Against Women (CAW) in India. Legislative awareness is essential for ensuring that women are well-informed about their rights and the legal safeguards that are accessible to them. By fostering a deep understanding of legislation, women have the ability to assert their rights and find legal recourse, so enhancing societal safety and fairness. Hence, it is crucial to integrate "Legislative awareness" into the complete strategy to combat CAW in order to attain significant and enduring advancements in addressing violence and discrimination against women in India.

Furthermore, it is imperative to prioritize the promotion of legislation awareness and the implementation of necessary enhancements in order to strengthen protections for women. These actions are essential elements of any successful approach. This work aims to examine both traditional and innovative approaches for addressing CAW, employing techniques such as exploratory data analysis, association rule mining (ARM), and machine learning models such as regression, K-Nearest Neighbor, Logistic Regression, and Random Forest classifiers.The study presents a new model that incorporates the 'Eli5' tool to enhance the analysis of CAW, with the goal of achieving a more nuanced comprehension.

The study aims to assess and compare the effectiveness of several machine learning models, specifically focusing on k-Nearest Neighbor, Logistic Regression, and Random Forest classifiers. Moreover, the evaluation of these classification models includes the consideration of 'Feature Importance' to potentially improve the overall performance of the models and gain a deeper understanding of the complex dynamics of the CAW scenario. This research aims to enhance the strategy for reducing the increasing number of crimes against women in India by incorporating various analytical methods.

LITERATURE REVIEW

The publication authored "C. Chakraborty et al. (2021)"[3] to discover the elements that can effectively mitigate crime against women in India. The article is published in the Journal of International Women's Studies. One notable attribute of this paper is its strength. The research employs panel regression analysis to find the determinants that can effectively mitigate crime against women in India. The report additionally presents an exhaustive examination of gender-based crimes in India, focusing on the regional level. Moreover, the research emphasizes the significance of efficient policy mechanisms in managing crime targeting women. Nevertheless, there are two notable limitations in the research. Firstly, it lacks a precise delineation of crime against women, resulting in potential ambiguity. Secondly, it fails to offer a comprehensive account of the methodology employed for data collection and analysis. The research findings may lack generalizability to other countries or locations. Fluctuations in crime reporting and recording methods throughout time can potentially impact the research results. The study can be enhanced by offering an elaborate exposition of the employed panel regression technique, as well as by providing a lucid account of the data collection and analysis procedures. Collectively, this study offers valuable insights into the determinants that exert influence over the occurrence of crimes targeting women in India. Nevertheless, the study would benefit from enhancing the exposition of the employed panel regression techniques, offering a more precise elucidation of the concept of crime against women, and providing a thorough account of the data collection and analysis procedures.

The authors of the study, R. Singh et al. (2020) and et. al. on "K-means Clustering Analysis of Crimes on Indian Women" emphasized the imperative for governments to prioritize the safeguarding of women. The Indian government has enacted several legislations and rules to safeguard women against assault and various forms of mistreatment. Presently, technological improvements predominantly benefit the government. Data mining

facilitates the application of several methodologies, such as regression analysis, classification, and clustering, to extract valuable insights and generate accurate predictions from a given dataset. The K-means clustering technique[4] is employed to classify a dataset of female inmates in this research study. As part of the data preparation procedure, researchers compiled a separate dataset consisting of all the records containing specific information about crimes committed against women. Subsequently, the study employed K-means clustering to further arrange the data. Our investigation was conducted utilizing quick mining, a commonly employed clustering method. Upon completion of the clustering analysis, the paper has thoroughly examined the findings and engaged in the exchange of opinions. The authors recommend that future studies utilize the publicly accessible derived dataset generated from this research.

The research paper, titled "Prevalence and risk factors of violence against women and children during COVID-19, Germany," published in 2021 by C. Ebert et al., aims to examine the occurrence and factors contributing to violence against women and children in Germany during the COVID-19 pandemic [5]. The study examines a significant matter, as the pandemic has generated apprehensions over heightened susceptibility and hazards for violence inside households. This critical analysis will evaluate the notable aspects, shortcomings, and impacts of the study. The key points of this study are as follows:

1. Extensive Data Collection: The paper employs a thorough data collection approach, encompassing both quantitative and qualitative data, to offer a more nuanced comprehension of the frequency and determinants of violence against women and children. This methodology enables a comprehensive examination of the matter.

2. Extensive Sample Size: The study incorporates a significant sample size, hence augmenting the generalizability of the findings. The researchers gathered data from several regions in Germany, taking into account differences in demography, socioeconomic status, and cultural aspects.

3. Multidimensional Analysis: The paper utilizes a multidimensional analysis to explore the intricacies of violence against women and children during the pandemic. It takes into account many risk factors, including socioeconomic status, household composition, and pre-existing violence, in order to acquire a comprehensive understanding of the issue. This strategy enhances the comprehensiveness of the research findings.The research findings indicate that out of the 3818 survey respondents, 118 individuals (3.09%; 95% confidence interval, CI: 2.54 to 3.64) reported engaging in physical fights, 293 individuals (7.67%; 95% CI: 6.83 to 8.52) reported experiencing emotional abuse, and 97 individuals (6.58%; 95% CI: 5.31 to 7.85) out of the 1474 participants who had children reported using child corporal punishment. The study determined that 3.57% of the individuals (with a 95% confidence interval ranging from 0.33 to 7.46) reported engaging in extramarital sex inside their relationship. We discovered analogous results for alternative forms of violence. The results of our regression analysis indicate that there is an increased risk of physical conflict associated with home quarantine (odds ratio, OR: 2.38; 95% confidence interval, CI: 1.56 to 3.61), financial concerns (OR: 1.60; 95% CI: 0.98 to 2.61), poor mental health (OR: 3.41; 95% CI: 2.12 to 5.50), and having young children (10 years old) (OR: 2.48; 95% CI: 1.32 to 4.64). Inadequate awareness and utilization of suitable support services. Nevertheless, the subsequent are the shortcomings:

1. Absence of Longitudinal Data: The study specifically examines the occurrence and factors contributing to risk within a certain timeframe of the COVID-19 pandemic in Germany. Nevertheless, the absence of longitudinal data hinders the capacity to determine causality or monitor fluctuations in violence over a period of time. Subsequent investigations should incorporate longitudinal designs in order to gain a more comprehensive understanding of historical patterns.

2. Potential Bias in Self-Reported Data: The paper mainly depends on data that is reported by individuals themselves, which might lead to biases such as recall bias and social desirability bias. These biases have the potential to impact the precision and dependability of the findings. To address this shortcoming, the researchers may have augmented the self-report data with additional objective metrics.

3. Insufficient emphasis on intervention strategies: Although the study offers useful information regarding the frequency and causes contributing to violence against women and children, it lacks a discussion on intervention tactics or policies that could effectively reduce or prevent such violence. Providing suggestions for interventions would improve the practical applications of the research. The paper's contribution is as follows:

1. Policy and Intervention Implications: The study paper's findings can inform the creation of policies and initiatives to address violence against women and children in public health crises such as the COVID-19 epidemic. Gaining insight into the risk factors linked to such violence can facilitate the creation of focused interventions and support services.

2. Addressing the Knowledge Gap: This study contributes to the scarce body of literature on violence against women and children in Germany during the COVID-19 pandemic. By presenting empirical data on the frequency and causes that contribute to a certain phenomenon, it enhances the current understanding and emphasizes the necessity for additional investigation in this field.

3. Raising Awareness: The study brings attention to the issue of violence against women and children during crises, elucidating the particular vulnerabilities and hazards experienced by these groups. Such consciousness can promote open discussion, support for a cause, and active involvement in society to effectively tackle the issue.

The article, entitled "Implementation of Predictive Crime Analytics in Municipal Crime Management System in Calauan, Laguna, Philippines," by J. Asor (2020) investigates the application of predictive crime analytics in enhancing the current crime management system in Calauan, Laguna. The study investigates the necessity for a proactive and effective strategy in crime prevention and control. Furthermore, this investigation scrutinizes the barangays of Calauan, Laguna. Dayap has a high vulnerability to a range of index offenses, including rape, murder, robbery, and theft [6] . The researchers commence by addressing the existing challenges faced by law enforcement authorities in Calauan, including constraints in resources and reliance on reactive crime management tactics. Their proposal suggests using predictive crime analytics, which employs past crime data, statistical analysis, and machine learning algorithms to forecast future criminal behaviors. The implementation method comprises three fundamental stages: data acquisition and preprocessing, model creation and assessment, and integration with the existing criminal management system. The researchers collected relevant crime data from several sources, such as police reports and public records, then conducted preprocessing to ensure its quality. Multiple machine learning algorithms, such as decision trees, logistic regression, and neural networks, were employed to build the predictive models. Integrating predictive crime analytics into the existing crime management system empowered law enforcement agencies to allocate their resources more effectively and strategically deploy them in regions with a high risk of crime. The researchers found that the predictive models achieved a high level of accuracy in crime prediction, enabling law enforcement authorities to proactively implement actions to avoid crimes and improve overall public safety. The paper asserts that incorporating predictive crime analytics into Calauan, Laguna's municipal crime management system holds promise for enhancing crime prevention and control measures.

A limited number of features, such as where and what sort of crime, are the focus of all of these studies that have been stated above. These parameters are inadequate for evaluating and preventing crimes committed against women and minors. The age group of the victim and the accused, the relationship between them, the qualifications of the victim, the culture and environment of the city, and other factors are also taken into consideration in this research apart from these characteristics. In this literature review, there is a dearth of work on age groups, which is regarded to be a research gap in this subject area.

METHODOLOGIES USED

The NCRB data pertaining to 568 districts in India and 30 police stations in Pune city comprise the corridor utilized for this study. The contributions presented in this paper align with the definition of the research problem:

1. In this research study, inferential analysis was conducted through the utilization of the chi-square test to test hypotheses. A CAW analysis is conducted in order to ascertain the interrelated elements that contribute to criminal activity. The probable origins of the incident crisis and the most influential correlated contributing attributes/variables were identified by the chi-square test. These characteristics can be used to categorize crimes by location.

Statistical analysis involves evaluating hypotheses. The null hypothesis (Ho) states that there is no correlation between the location and kind of crime, while the alternative hypothesis (H1) asserts that there is a strong connection between the two variables. Previously, we determined that the chi-square test $\chi 2$ is employed to ascertain the independence of the variables loc and Toc[7]. The concept of Toc can be divided into four distinct categories: 1. Physical attack or violent act. 3. Abduction 2. Sexual assault 4. Spousal violence. Pune city (Police Chowki) is divided into 30 distinct "LoCs," denoted by the letter A in the table provided. Below are two test results.H0: There is no correlation between the criminal patterns of molestation and variables such as age group, location, and type of crime.

H1: There exists a robust correlation between the occurrence of crimes targeting women and individuals belonging to the minor and specific age groups, as well as the geographical region.

H2: A correlation exists between the type of crime and the age group.

H3: There is a correlation between the geographical location and the nature of criminal activities.

After the computations, we received the number of degrees of freedom is 87. The level of significance, represented by alpha, is set at 0.05. The estimated value of the chi-square statistic is 8377.9272. The tabular value for 87 with a significance level of 0.05 is 109.77 for the chi-square test.

Consequently, the calculated $\chi 2$ value is greater than the tabular $\chi 2$ value (Yes), showing that we cannot accept the null hypothesis. We establish the premise that there is a link between crime kinds and geographic locations. Analysis of crime statistics on a location-specific level can help predict areas of concern. Put simply, the predictive value of a crime is more influenced by its geographical location rather than the specific date on which it occurred. The table below presents a concise overview, categorized by location, of the percentage distribution of each type of crime. We employed association rule mining to analyze the spatial distribution of several crime categories in Pune.

Risk Zone	Type of Crime	% of crime severity to total crimes	Location with high risk
High Risk	Assault	29%	Hadapsar(A25), Yerwada(A23), Chatushrungi(A21), Kondhawa(A27) Wanworie(A30)
Medium	Rape	15%	Kondhawa (A27), Hadapsar(A25), Yerwada(A23), Wanworie(A30)
High Risk	Kidnapping	31%	Yerwada (A23), Hadapsar (A25), Kondhawa (A27), wanworie(A30)
Medium	Cruelty by husband	26%	Yerwada(A23), Hadapsar(A25), Kondhawa(A27),

Table I: Percentage of severity of crime in the Pune city

Pune has been separated into two regions, high and medium, based upon the frequency of reported violent crimes, as shown in Table C. Based on the statistics, yerwada (A23), Hadapsar (A25), kondhawa (A23), and wanoworie

(A30) exhibited the greatest incidence rates of kidnapping (31%) and assault (29%) correspondingly. Reported incidents of Cruelty by Husband accounted for 26% and Rape accounted for 15% in the areas of Kondhawa, Hadapsar, and Wanworie. Additional prediction criteria can be discovered by conducting this research with monthly data.

2. Machine learning models, such as regression, were employed to detect the dark spot.

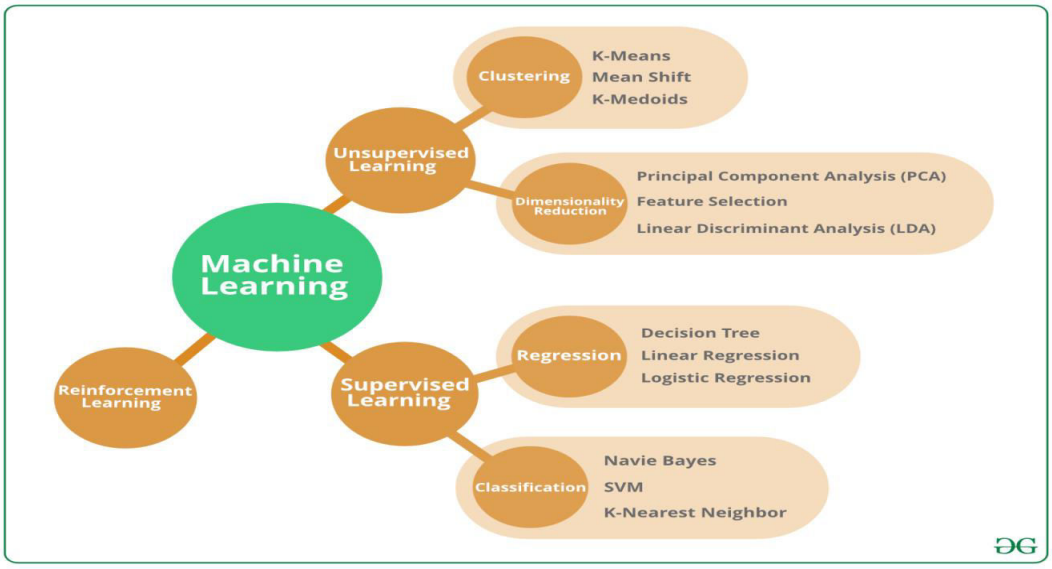

Figure 4.8: Attributes with Correlation coefficient

The evaluation of algorithm efficacy was depicted through the utilization of Root Mean Square Error (RMSE). On the basis of model validation, a regression model is constructed for the purpose of forecasting[8]. Data analyses were conducted in this study utilizing Python. Using a variety of graphs, exploratory data analyses (EDA) were performed to examine relationships, patterns, and the validity of the assumptions underlying the use of regression. The following characteristics have been identified as having a significant impact on the location of criminal activities: "Population growth," "Literacy rate of males and females," "Age group, and working status of males and females." These factors collectively contribute significantly to the identification of "dark spots."

We shall investigate numerous dataset-derived input features. We want to find the "Dark Spot" and predict the dangerous area by studying its characteristics. Start by importing the dataset and doing exploratory analysis. The output will show 4827 cases and 5 columns. A statistical analysis of the dataset is shown:

Statistics	Location	Assault	Rape	Kidnapping	Cruelty by husband
Count	31	1401	701	1475	1250
Mean	15.5	90.387	45.226	95.161	80.645
Max	31	1401	701	1475	1250
Min	1	9	3	12	10
Std dev	8.803	245.358	123.32	258.275	219.041

Table II: Statistics of given dataset

Correlation Coefficient	1
Mean Absolute Error (MAE)	0.4695
Root Mean Squared Error (RMSE)	1.4597
Relative Absolute Error (RAE)	0.1189
Root Relative Squared Error (RRSE)	0.12

Model research predictions were mostly accurate. Mean Absolute Error and Root Mean Squared Error measure model performance. Prediction standard deviation is RMSE. The RMSE is 1.46, 10% higher than the MAE of 0.46. This implies a linear relationship, found. More training data is needed to improve the model's predictions. Model shows Type of Crime (ToC) adversely correlates with year and positively with place. Second, predict from test results and plot estimated and actual values. Note the discrepancy between real and expected. Our model predicts well since real and anticipated values are similar. As seen, the Root Mean Squared Error (RMSE) is 1.4597, 10% greater than the MAE of 0.4695. Our algorithm made precise predictions, indicating it was accurate. This study implemented multivariate linear regression using Python's Scikit-Learn machine learning framework.

3. The 'Elbow method' was employed to determine the optimal number of clusters, after which each cluster was characterized. In this study, we present a proposed algorithm that utilizes k-means clustering to partition the total events of data points into four discrete subgroups, subsequently classifying them into severity zones. An in-depth examination of each cluster reveals accident variables. Consequently, the clustering method shed light on the numerous causes and factors that have had a significant impact on traffic safety.

Line charts of the SSE for each k value help us determine the right number of Clusters. We utilized the "Elbow method" to determine "k" to classify accident locations by frequency counts in this study. It suggests an appropriate k clusters. Sum of squared distances (SSE) between data points and cluster centroids determines this. The SSE curve flattens and forms an elbow in Figure 1, so we must choose 'k'. This figure illustrates the x-axis cluster count and y-axis within-cluster sum-of-squares (WCSS).

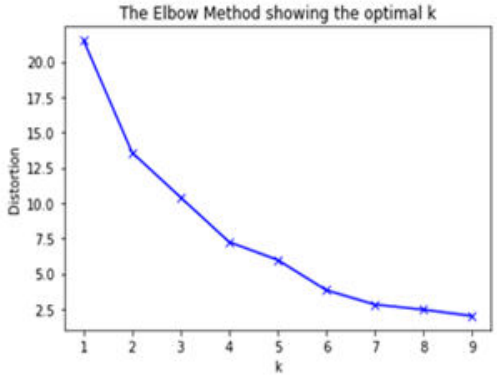

Figure 1: The Elbow method

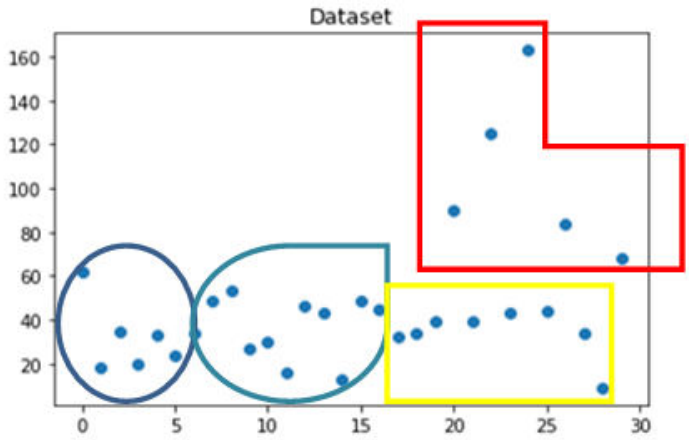

Figure 2: Optimal Number of Clusters

The value of k, also known as the optimal value, can be found at the "Elbow" of the line chart when it is shaped like an arm, as demonstrated in Figure 2. Upon conducting an extensive examination of each cluster, it was discovered that accident variables classified areas with high and low crime rates according to socioeconomic factors such as GIP, FLIT, Mlit, and age group.

The NCRB information was used to analyze all 568 Indian districts. Clusters 0 has 216 districts, 1 238 districts, 2 106 districts, and 3 eight districts.As shown in the table above, composite variables are:Cluster C-0 has 216 districts, 127 of which have a lower crime rate (58.80%) due to modest population growth and 89 of which have a higher crime rate (41.20%). Cluster C-0 has two groups.Out of 238 districts in Cluster C-1, 120 have a low crime rate (50.42%) and 118 have a high crime rate (49.58%).Cluster C-2 evaluates 106 districts, with 71 having a low crime rate of 66.98% and 35 having a higher rate of 33.02%. The fourth cluster C-3 has eight districts, five of which have extraordinarily low crime rates (62.50%) and three exceptionally high crime rates (37.50%).In similar fashion, the composite variable literacy rate in all clusters reveals a crime rate of 55.65% in low crime regions with high female literacy and 44.35% in high crime regions with low female literacy.

4. The support, confidence, and lift values of our database have been computed. Additionally, the findings were displayed as regulations through the implementation of association rule mining. According to the first criterion, a positive correlation exists between location and the nature of the crime. According to the second rule, there is a correlation between age group and the category of crime. And the third rule states that areas with rapid population growth and low levels of literacy tend to have a higher incidence of criminal incidents. The aforementioned model can therefore be employed to assess the efficacy of a safety measure through the detection of the dark mark.

5. Proposed is a novel model based on the concept that employs the instrument 'Eli5' to enhance the analysis of CAW. For debugging Machine Learning classifiers and describing predictions, the Eli5 Python package is utilized. In this instance, the binary classification model produced by the Eli-5 utility was 75% accurate. The results of the Binary Classifier, Confusion Matrix, and Correlation Matrix are found to be highly correlated and positive, with a probability score exceeding 5.

6. An analysis of the reported crimes against women was conducted by assigning the most appropriate supervised machine learning technique. With the aid of potentially pertinent indicators, crimes were analyzed. An evaluation was conducted to compare k-Nearest Neighbor, Logistic Regression, and Random Forest. In order to evaluate the performance of a model, three accuracy metrics are utilized: the classification report, the confusion matrix, and the area under the receiver operating characteristics (AUROC). Put simply, the metrics and tools that are most frequently employed to evaluate a classification model are the confusion matrix, area under the ROC curve, and accuracy (classification report). Overall, this study provides comprehensive experimental findings that demonstrate the superior performance of the Random Forest (RF) classifier in comparison to the K-nearest neighbour (KNN) and Logistic Regression (LR) classifiers. These results suggest that the RF classifier has the potential to bridge the existing or future knowledge divide.

7. The performance of the aforementioned three classification models was assessed by employing the 'Feature Importance' metric. Feature importance (FI) can contribute to a more comprehensive comprehension of the CAW scenario and may inspire additional model enhancements. FI scores are a critical component in the pursuit of predictive modeling. The three most significant features that demonstrate positive importance are population growth, male and female literacy rates, and the employment status of both sexes. These factors can aid in the early detection of issues. According to the other attributes, such as the year, criminal activity occurs on any given date and is uniformly distributed throughout the year.

RESULTS AND DISCUSSION:

This study paper provides a thorough examination of many conventional and innovative methods used to tackle the problem of Customer Attrition Wastage (CAW). The study encompasses:

- Apply inferential analysis to investigate and assess links
- Machine learning models, such as regression leveraging Root Mean Square Error (RMSE)
- Conducting exploratory data analytics (EDA) to ascertain the most suitable number of clusters
- Computation of support, confidence, and lift coefficients
- Leveraging association rule mining to present discoveries in the form of rules

REFERENCES

[1] NCRB report, “2016-2018”.

[2] Indian express, “indian express report-Crimes recorded against woman.”

[3] C. Chakraborty, A. Afreen, and D. Pal, “Crime against Women in India: A State Level Analysis,” 2021. [Online]. Available: https://vc.bridgew.edu/jiws

[4] R. Singh and B. Hazela, "A Survey on Machine Learning And Data Mining Methods And Applications," vol. 20, no. 5, pp. 61–65, doi: 10.9790/0661-2005016165.

[5] Cara Eberta Janina, "Prevalence and risk factors of violence against women and children during COVID-19, Germany," *Bull World Health Organ.*, pp. 429–438, 2021.

[6] Sherwin B. Sapin and Jonardo R. Asor, "Implementation of Predictive Crime Analytics in Municipal Crime Management System in Calauan, Laguna, Philippines," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1, pp. 150–157, Jun. 2020.

[7] A. A. Patel and P. Kulkarni, "GENDER DISCRIMINATION AND CRIME PREDICTION: DEVELOPING WOMEN'S SAFETY AND SECURITY THROUGH AUTOMATED LEARNING TECHNIQUES," *Journal of the Asiatic Society of Mumbai*, vol. 97, no. 12, 2023.

[8] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," *Augmented Human Research*, vol. 5, no. 1, Dec. 2020, doi: 10.1007/s41133-020-00032-0.

[9] Priyanka Das and Asit Kumar Das, "Behavioural analysis of crime against women using a graph based clustering approach," *IEEE*, 2017.

[10] Reddy Chandan K and Vinzamuri Bhanukiran, "A Survey of Partitional and Hierarchical Clustering Algorithms," in *Data Clustering Algorithms and applications*, 1st ed., 2014, pp. 1–24.

[11] B. Kaur, L. Ahuja, and V. Kumar, "Factors affecting crime against women using regression and K-means clustering techniques," in *Lecture Notes in Networks and Systems*, vol. 11, Springer, 2018, pp. 149–162. doi: 10.1007/978-981-10-3953-9_15.

[12] Rohit Sharma, "What is Clustering and Different Types of Clustering Methods," https://www.upgrad.com/blog/clustering-and-types-of-clustering-methods/.

[13] E. M. Armstrong, B. Glover Reed, and L. W. Bennett, "How and How Much: Combined Services for Domestic Violence and Substance Abuse," *Violence Against Women*, vol. 25, no. 12, pp. 1450–1470, Oct. 2019, doi: 10.1177/1077801218820201.