EXPLORING VIOLENCE AGAINST WOMEN: A DATA MINING APPROACH TO UNCOVERING PATTERNS AND CHARACTERISTICS

Prof. Dikshendra Sarpate[1] and Dr. Manju D. Pawar[2]

[1]Department of Artificial Intelligence and Data Science Engineering, Zeal college of Engineering and Research, Pune, MH, India

ABSTRACT

This paper explores the pervasive issue of Violence Against Women (VAW) using advanced data collection and analysis techniques. We collected extensive textual data from online sources, including news articles, social media, and official reports, and processed it into a structured dataset.
Using text mining and Natural Language Processing (NLP), we analyzed the data to uncover patterns, sentiments, and key themes related to VAW. Techniques like TF-IDF helped identify significant terms, streamlining the analysis. We then trained and evaluated several classifiers, including Naive Bayes, Random Forest, SVM, AdaBoost, and Artificial Neural Networks (ANN), to improve classification accuracy.
Our aim is to use advanced text mining and Natural Language Processing (NLP) techniques to extract actionable insights from the collected data. By applying methods like Term Frequency-Inverse Document Frequency (TF-IDF), we aim to identify and prioritize key textual features related to VAW. This approach enables us to streamline further analysis by concentrating on the most important aspects of the textual data
Our findings suggest that integrating multiple machine learning techniques enhances the understanding of VAW's complex nature. The insights gained can inform targeted interventions and support the development of more effective policies to combat VAW, contributing to a safer and more equitable society for women.

KEYWORDS

NATURAL LANGUAGE PROCESSING, SVM , ANN , TF-IDF

1. INTRODUCTION

Violence Against Women (VAW) is a widespread social issue that exists across most societies, regardless of religion, age, or social class. According to [1], VAW is defined as any act of gender-based violence that causes or is likely to cause physical, sexual, or psychological harm or suffering to women and girls. This includes threats, coercion, or arbitrary deprivation of liberty, whether in public or private life. Similarly, the United Nations defines VAW as any act of gender-based violence that results in physical, sexual, or psychological harm or suffering, including threats, coercion, or deprivation of liberty [2].

Violence Against Women (VAW) is a critical social issue with deep and far-reaching effects on individuals, families, and communities across the globe. Despite increased awareness and ongoing advocacy, VAW remains pervasive, manifesting in various forms such as physical, psychological, and sexual violence.

Over a quarter of women aged 15–49 who have been in a relationship have experienced physical and/or sexual violence by an intimate partner at least once in their lives. The prevalence of lifetime intimate partner violence varies globally,[3] with estimates ranging from 20% in the Western Pacific and 22% in high-income countries and Europe, to 25% in the Americas, 33% in the WHO African region, 31% in the WHO Eastern Mediterranean region, and 33% in the WHO South-East Asia region.

surveys on violence, such as The National Crime Victims Survey (NCVS), often include victims beginning at age 12. In addition, the highest rates of rape and sexual assault are found among women aged 12 to 24 years [4]: females in their teens and 20s are those most likely to be dating, and, therefore, subject to dating violence National Academies of Sciences, Engineering, and Medicine. 1996. Understanding Violence Against Women.

This paper aims to develop a deep understanding of the complex nature of violence against women by collecting and analyzing diverse textual data from sources like news articles, social media, and official reports.

Identification of Patterns and Trends: Through the use of advanced text mining and Natural Language Processing (NLP) techniques, the project seeks to identify patterns, sentiments, and key themes within the data. This analysis will offer valuable insights into the language and discourse surrounding violence against women in various contexts.

Feature Selection and Prioritization: By employing Term Frequency-Inverse Document Frequency (TF-IDF), the project focuses on selecting and prioritizing significant features, terms, and phrases indicative of violence against women. This step is essential for refining the analysis and concentrating on the most relevant aspects of the textual data.

1.1 Web Scraping: Web scraping is the automated technique of extracting data from websites. This process involves using specialized software tools or programming languages to access and retrieve information embedded within the HTML code of web pages. The data collected can range from text and images to links and other content available on the site. Web scraping is frequently employed for purposes such as data analysis, research, and gathering information from various online platforms.

HTML Parsing**:** HTML parsing involves the examination and breakdown of the HTML (Hypertext Markup Language) code that structures a webpage, enabling the extraction of relevant data or information. HTML is the standard language used to create and format webpages, comprising a series of elements represented by tags. Parsing HTML is a crucial step in the web scraping process, as it allows you to navigate through the HTML structure and selectively extract specific content or elements of interest. The first step in HTML parsing is to create a parser object, which is initialized with the HTML content of the webpage and a designated parser type (e.g., 'html.parser'). This parser object provides a structured representation of the HTML document, facilitating easy navigation and data extraction.

1.2 Feature Selection: Feature selection is a process in machine learning and statistics that involves identifying the most important features or variables from a larger set. The goal is to select a subset of features that significantly enhances the predictive performance of a model while minimizing redundancy and reducing noise. This step is essential for improving model efficiency, reducing the risk of overfitting, and gaining a clearer understanding of the key patterns and relationships within the data Feature selection helps reduce the dimensionality of the dataset, making it more manageable and mitigating the curse of dimensionality. By choosing the most relevant features, the process can lead to simpler and more interpretable models that perform better when applied to new, unseen data. This not only enhances model efficiency but also improves the model’s ability to generalize, leading to more accurate predictions and insights.

2. Related work

Violence Against Women (VAW) continues to be a pressing global issue, with recent research underscoring its prevalence, evolving forms, and the complex challenges involved in addressing it. This survey examines the most recent literature (2020-2024) on VAW, focusing on the prevalence of violence, emerging forms, the impact of the COVID-19 pandemic, and current interventions and challenges.

Recent global estimates indicate that approximately 27% of women aged 15-49 have experienced physical and/or sexual intimate partner violence (IPV) during their lifetime [5]. The prevalence is higher in low- and middle-income countries, where social and economic factors exacerbate vulnerability. In some regions, such as Sub-Saharan Africa and South Asia, the rates of IPV exceed 30%. In recent years, there has been growing recognition of new forms of VAW, particularly those facilitated by digital technologies. Cyber violence, including online harassment, cyberstalking, and the non-consensual sharing of intimate images, has seen a significant rise. A 2021 study by the European Institute for Gender Equality (EIGE) found that 1 in 10 women in the EU had experienced cyber harassment since the age of 15. Studies conducted during the pandemic highlight several key trends,

- **Increased Frequency of IPV:** Many women reported an escalation in the frequency and severity of IPV during lockdowns. A study in the United States found a 7.5% increase in IPV-related 911 calls during the early months of the pandemic [6].

- **Barriers to Accessing Support:** The pandemic disrupted access to essential services for VAW survivors, including shelters, counseling, and legal assistance. Many services shifted to online platforms, but access was limited for women in low-resource settings or those with limited digital literacy [7]

There is a growing emphasis on survivor-centered approaches, which prioritize the safety, dignity, and autonomy of survivors in all interventions. These approaches involve survivors in the design and implementation of services and policies, ensuring that their needs and preferences are at the forefront [8] addressing technology with the rise of cyber violence, there is an urgent need to develop legal and technological solutions to protect women online. This includes updating laws to address online harassment and cyberstalking, as well as improving digital literacy and security for women and girls [9].

3 Proposed methodologies for VAW

This paper addresses the challenge of developing a data-driven classification framework for studying Violence Against Women (VAW) using advanced data mining techniques. Traditional methodologies for studying VAW often suffer from limitations such as sample biases, underreporting, and labour-intensive processes, which hinder the accurate and comprehensive understanding of the issue. By leveraging digital technologies and automated classification algorithms, this project aims to improve the accuracy and efficiency of identifying various forms of VAW, including physical, psychological, sexual, and economic abuse. The goal is to enhance our understanding of VAW, support evidence- based interventions, and ultimately contribute to the prevention and eradication of gender-based violence[10][11].

3.1 Data Collection:

As the initial process of the proposed strategy, The UN Secretary-General’s Database on Violence Against Women is a dataset introduced by the United Nations General Assembly in 2009, now managed by UN Women. It contains comprehensive global data on the extent, nature, and consequences of various forms of violence against women, as well as the effectiveness of policies and programs aimed at eliminating such violence. The dataset, which is 1073 KB in size, includes eight columns and 12,601 rows, with key variables such as RecordID, Country, Gender, Demographics Question, Demographics, Response, Question, Survey Year, and Value.

The dataset covers a wide geographical scope, providing a global perspective on violence against women. To prepare the data for analysis, several text preprocessing techniques are employed. First, common stop words are filtered out from the Question column using the NLTK library, which helps to reduce irrelevant data and enhance the significance of the text. The text is then normalized by converting it to lowercase to ensure consistency across various columns, including Country, Education, Marital Status, Employment, and Residence. To ensure the text is clean and uniform, special characters and irrelevant content are removed using regular expressions. A dictionary of tokens is maintained to support further analysis. These preprocessing steps ensure that the dataset is thoroughly cleaned, standardized, and ready for subsequent modelling and analysis, providing a solid foundation for understanding and addressing violence against women on a global scale.

Feature extraction is a crucial step in preparing data for classification tasks. In our study, we explored two prominent feature extraction methods: Chi-squared (chi2) and Term Frequency-Inverse Document Frequency (TF-IDF). The goal was to identify which method provides the optimal features for classifying instances of violence against women and to apply the most effective classification algorithm based on this selection

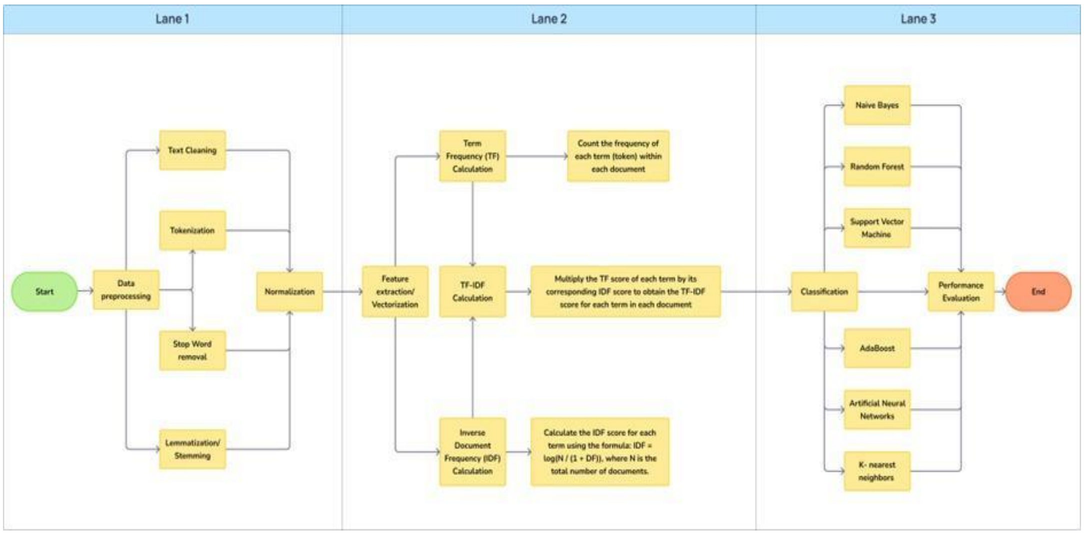

Lane 1
Lane 2
Lane 3
Start
Data preprocessing
Text Cleaning
Tokenization
Stop Word removal
Lemmatization/ Stemming
Normalization
Feature extraction/ Vectorization
Term Frequency (TF) Calculation
Count the frequency of each term (token) within each document
TF-IDF Calculation
Multiply the TF score of each term by its corresponding IDF score to obtain the TF-IDF score for each term in each document
Inverse Document Frequency (IDF) Calculation
Calculate the IDF score for each term using the formula: IDF = log(N / (1 + DF)), where N is the total number of documents.
Classification
Naive Bayes
Random Forest
Support Vector Machine
AdaBoost
Artificial Neural Networks
K- nearest neighbors
Performance Evaluation
End

Figure 1. Flowchart that outlines a machine learning text classification process

Figure.1 appears to be a flowchart that outlines a machine learning text classification process divided into three lanes: **Lane 1: Data Preprocessing**

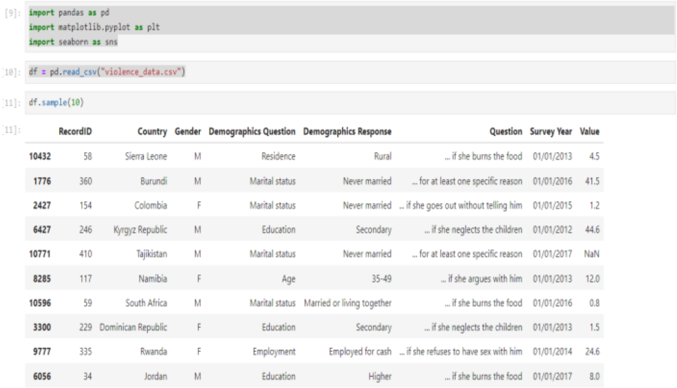
1. **Start:** The process begins with data preprocessing.
2. **Text Cleaning:** The text is cleaned to remove unwanted characters, punctuation, etc.
3. **Tokenization:** The text is split into individual tokens (words or phrases).
4. **Normalization:** The text is standardized, usually involving lowercasing, expanding contractions, etc.
5. **Stop Word Removal:** Common words that do not contribute much meaning (like "the," "and") are removed.
6. **Lemmatization/Stemming:** The words are reduced to their base or root form.

Lane 2: Feature Selection/Vectorization shown in Figure 2 (a) and (b) respectively,

1. **Term Frequency (TF) Calculation:** The frequency of each term in each document is counted.
2. **Inverse Document Frequency (IDF) Calculation:** The IDF score is calculated for each term across all documents.
3. **TF-IDF Calculation:** The TF score is multiplied by the corresponding IDF score to calculate the TF-IDF score for each term in each document.

Lane 3: Classification and Performance Evaluation

1. **Classification:** Multiple classification algorithms are applied to the vectorized data:
   - Naive Bayes
   - Random Forest
   - Support Vector Machine
   - AdaBoost
   - Artificial Neural Networks
   - K-nearest neighbors
2. **Performance Evaluation:** The models are evaluated to determine their performance.
3. **End:** The process concludes.



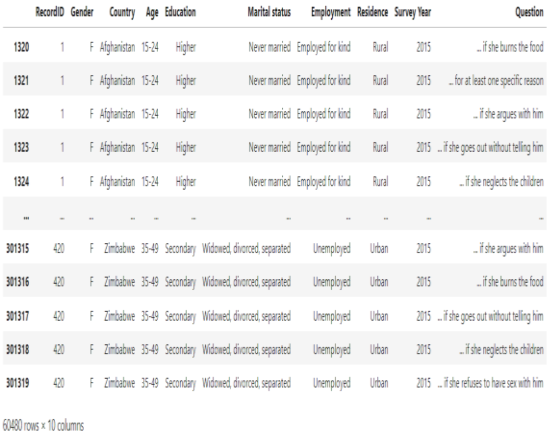
```
[9]: import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns

[10]: df = pd.read_csv("violence_data.csv")

[11]: df.sample(10)
```

[11]:

	RecordID	Country	Gender	Demographics Question	Demographics Response	Question	Survey Year	Value
10432	58	Sierra Leone	M	Residence	Rural	... if she burns the food	01/01/2013	4.5
1776	360	Burundi	M	Marital status	Never married	... for at least one specific reason	01/01/2016	41.5
2427	154	Colombia	F	Marital status	Never married	... if she goes out without telling him	01/01/2015	1.2
6427	246	Kyrgyz Republic	M	Education	Secondary	... if she neglects the children	01/01/2012	44.6
10771	410	Tajikistan	M	Marital status	Never married	... for at least one specific reason	01/01/2017	NaN
8285	117	Namibia	F	Age	35-49	... if she argues with him	01/01/2013	12.0
10596	59	South Africa	M	Marital status	Married or living together	... if she burns the food	01/01/2016	0.8
3300	229	Dominican Republic	F	Education	Secondary	... if she neglects the children	01/01/2013	1.5
9777	335	Rwanda	F	Employment	Employed for cash	... if she refuses to have sex with him	01/01/2014	24.6
6056	34	Jordan	M	Education	Higher	... if she burns the food	01/01/2017	8.0



	RecordID	Gender	Country	Age	Education	Marital status	Employment	Residence	Survey Year	Question
1320	1	F	Afghanistan	15-24	Higher	Never married	Employed for kind	Rural	2015	... if she burns the food
1321	1	F	Afghanistan	15-24	Higher	Never married	Employed for kind	Rural	2015	... for at least one specific reason
1322	1	F	Afghanistan	15-24	Higher	Never married	Employed for kind	Rural	2015	... if she argues with him
1323	1	F	Afghanistan	15-24	Higher	Never married	Employed for kind	Rural	2015	... if she goes out without telling him
1324	1	F	Afghanistan	15-24	Higher	Never married	Employed for kind	Rural	2015	... if she neglects the children
...	...	...	...	...	...	...	...	...	...	...
301315	420	F	Zimbabwe	35-49	Secondary	Widowed, divorced, separated	Unemployed	Urban	2015	... if she argues with him
301316	420	F	Zimbabwe	35-49	Secondary	Widowed, divorced, separated	Unemployed	Urban	2015	... if she burns the food
301317	420	F	Zimbabwe	35-49	Secondary	Widowed, divorced, separated	Unemployed	Urban	2015	... if she goes out without telling him
301318	420	F	Zimbabwe	35-49	Secondary	Widowed, divorced, separated	Unemployed	Urban	2015	... if she neglects the children
301319	420	F	Zimbabwe	35-49	Secondary	Widowed, divorced, separated	Unemployed	Urban	2015	... if she refuses to have sex with him

60480 rows × 10 columns

Figure 2 (a) Before Feature Engineering **(b) After Feature Engineering**

3.2 AI for Data Mining

Artificial Intelligence (AI) greatly enhances data mining by enabling automated pattern recognition and analysis. Machine learning algorithms, such as clustering and classification, help identify important patterns in large datasets, making tasks like customer segmentation and trend prediction easier. AI-powered Natural Language Processing (NLP) tools can extract valuable insights from unstructured text, supporting applications like sentiment analysis and topic modeling. Additionally, AI is crucial in anomaly detection, identifying irregularities that could indicate fraud or unexpected events. Recommender systems use AI to analyze user behavior and provide personalized suggestions, improving user experiences in areas like e-commerce and content consumption. Techniques like deep learning, ensemble learning, and

AutoML further boost the accuracy and efficiency of data mining, leading to better knowledge discovery and decision-making. AI-driven tools also enhance interactive data exploration, allowing users to effortlessly explore complex datasets and uncover patterns through user-friendly interfaces. The integration of AI and data mining not only automates repetitive tasks but also enables the creation of advanced models that can manage complex data structures, driving innovation and efficiency across various industries.

3.3 Data Mining

Text mining, or text analytics, involves extracting valuable insights from large amounts of unstructured text data, such as documents, articles, and social media posts. This process uses data mining techniques to uncover patterns, relationships, and trends within the text that might not be easily detected through manual analysis. A key part of text mining is the application of natural language processing (NLP) techniques, which help preprocess and analyze the text by extracting entities, sentiments, topics, and relationships. Machine learning algorithms, including classification, clustering, and topic modelling, are then used to categorize documents, determine sentiment, and identify underlying themes. Text mining has valuable applications, such as analysing customer feedback through sentiment analysis, organizing content through topic modelling, and improving information retrieval for more efficient searching through large document collections. By integrating data mining techniques into text mining, we can better manage and understand vast amounts of textual data, leading to the development of intelligent systems that can extract meaningful insights from the growing body of text available

3.4 Text Mining

Text mining for datasets related to violence against women utilizes natural language processing (NLP) and data mining techniques to extract valuable insights from textual data like news articles, social media posts, and reports. By analyzing this unstructured text, researchers can uncover key themes, sentiment, and contextual information, revealing patterns and trends associated with gender-based violence. This approach aids in identifying recurring issues, predicting future patterns, and assessing the impact of interventions, thereby providing actionable insights to complement traditional research methods and empower efforts to address and prevent violence against women.

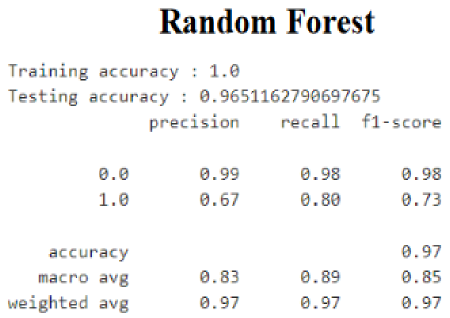
In our investigation into violence against women, we adopt a multifaceted approach that integrates various classification techniques to analyze and understand the data. By employing algorithms such as Naive Bayes, Random Forest, Support Vector Machine (SVM), AdaBoost, and Artificial Neural Networks (ANN), we aim to accurately classify instances of violence and identify underlying patterns. These methodologies enable us to discern complex relationships within the data, providing a comprehensive framework for understanding and addressing gender-based violence.

4.Results and discussion

In evaluating the performance of various classification techniques on our dataset related to violence against women, we observed figure 2 is a notable difference across models. The Gaussian Naive Bayes algorithm achieved an accuracy of 65.98% but struggled with class imbalance, predominantly predicting only one class and failing to identify true positives for other classes. The Support Vector Classifier (SVC) showed strong performance with a training accuracy of 97.52% and a testing accuracy of 94.19%, though its precision and recall for the minority class were lower, indicating a need to address class imbalance. The Random Forest classifier excelled with a perfect training accuracy and high testing accuracy of 96.51%, but its perfect training accuracy suggests potential overfitting.

The AdaBoost classifier, using 50 estimators, achieved an accuracy of approximately 65.98%. However, the confusion matrix revealed a critical limitation: the model exclusively predicted class 1 for all instances, mirroring issues found in the Naive Bayes, SVM, and XGBoost classifiers. This led to zero correct predictions for classes 0 and 2, underscoring a significant problem with class imbalance. Although the overall accuracy might seem acceptable, the classifier's inability to correctly classify instances from classes 0 and 2 indicates that the model is not effectively capturing the distinctions among the classes. To enhance the model's performance, it is crucial to address this imbalance by adjusting class weights, oversampling the minority classes, or fine-tuning parameters to better reflect the distribution of the data.The AdaBoost classifier demonstrated robust performance with a training accuracy of 97.96% and a testing accuracy of 97.67%, showing high precision and recall for the majority class and decent metrics for the minority class. The Decision Tree classifier achieved perfect training accuracy but lower testing accuracy of 83.72%, with strong performance for the majority class but poor results for the minority class, indicating potential overfitting and class imbalance issues. Finally, the Logistic Regression model performed excellently with a training accuracy of 96.65% and a testing accuracy of 97.67%, maintaining high precision and recall for both classes..

The experimental setup for analyzing violence against women (VAW) involves several key components and configurations designed to ensure accurate and reliable results. The setup includes the computational environment, data sources, preprocessing steps, feature extraction methods, classification models, and performance evaluation techniques. Each aspect is meticulously configured to facilitate the comprehensive analysis of the collected data.

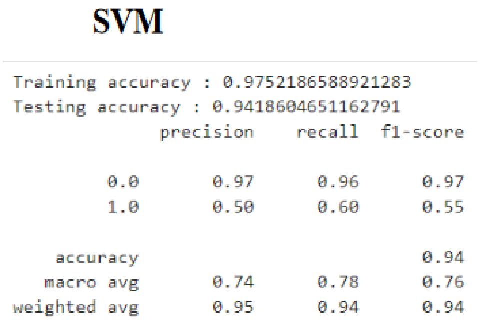


Random Forest

```
Training accuracy : 1.0
Testing accuracy : 0.9651162790697675
              precision    recall  f1-score

         0.0       0.99      0.98      0.98
         1.0       0.67      0.80      0.73

    accuracy                           0.97
   macro avg       0.83      0.89      0.85
weighted avg       0.97      0.97      0.97
```

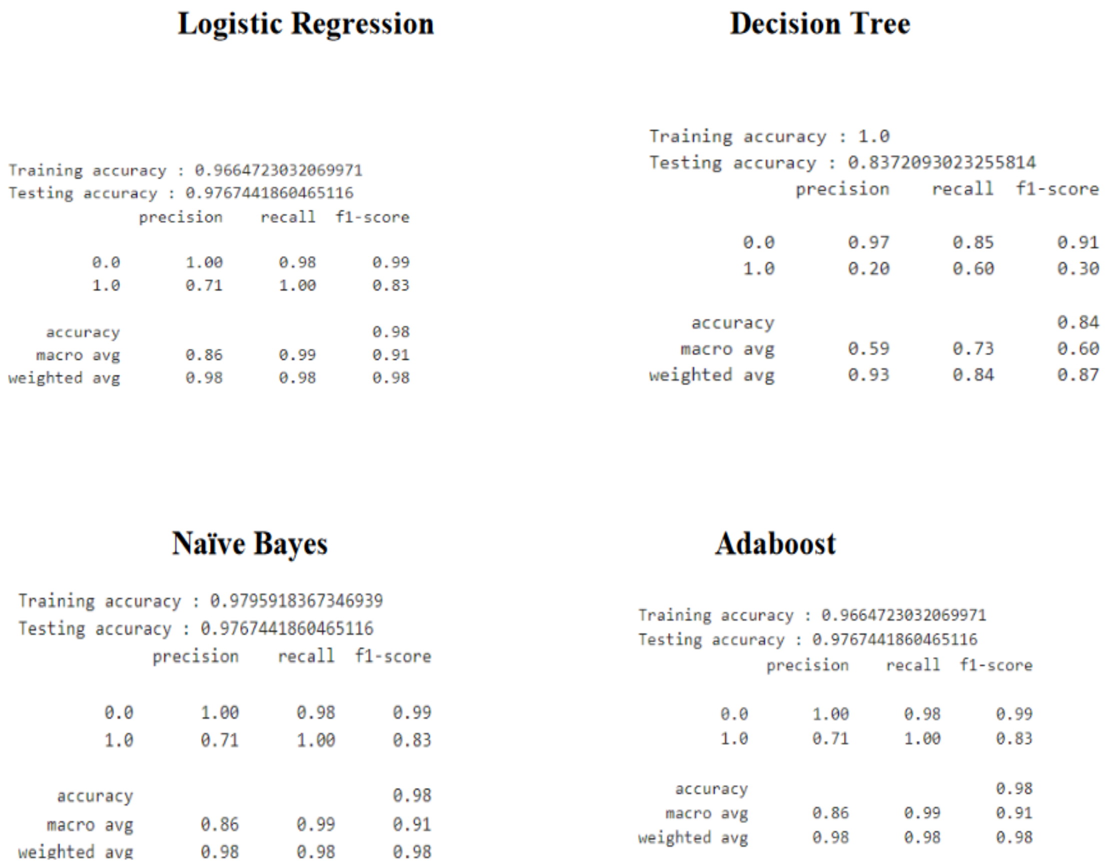


SVM

```
Training accuracy : 0.9752186588921283
Testing accuracy : 0.9418604651162791
              precision    recall  f1-score

         0.0       0.97      0.96      0.97
         1.0       0.50      0.60      0.55

    accuracy                           0.94
   macro avg       0.74      0.78      0.76
weighted avg       0.95      0.94      0.94
```



Logistic Regression

```
Training accuracy : 0.9664723032069971
Testing accuracy : 0.9767441860465116
              precision    recall  f1-score

         0.0       1.00      0.98      0.99
         1.0       0.71      1.00      0.83

    accuracy                           0.98
   macro avg       0.86      0.99      0.91
weighted avg       0.98      0.98      0.98
```

Decision Tree

```
Training accuracy : 1.0
Testing accuracy : 0.8372093023255814
              precision    recall  f1-score

         0.0       0.97      0.85      0.91
         1.0       0.20      0.60      0.30

    accuracy                           0.84
   macro avg       0.59      0.73      0.60
weighted avg       0.93      0.84      0.87
```

Naïve Bayes

```
Training accuracy : 0.9795918367346939
Testing accuracy : 0.9767441860465116
              precision    recall  f1-score

         0.0       1.00      0.98      0.99
         1.0       0.71      1.00      0.83

    accuracy                           0.98
   macro avg       0.86      0.99      0.91
weighted avg       0.98      0.98      0.98
```

Adaboost

```
Training accuracy : 0.9664723032069971
Testing accuracy : 0.9767441860465116
              precision    recall  f1-score

         0.0       1.00      0.98      0.99
         1.0       0.71      1.00      0.83

    accuracy                           0.98
   macro avg       0.86      0.99      0.91
weighted avg       0.98      0.98      0.98
```

Figure 2 Performance of various classification techniques

5. CONCLUSION

Our study underscores the effectiveness of integrating advanced classification techniques—Naive Bayes, Random Forest, Support Vector Machine (SVM), AdaBoost, and Artificial Neural Networks (ANN) to analyze violence against women comprehensively.

Among these, the Naive Bayes algorithm proves particularly notable, achieving a high testing accuracy of approximately 97.67%. This performance highlights its strength in managing complex datasets with high dimensionality. Naive Bayes is adept at capturing intricate patterns and relationships within the data, leveraging probabilistic models to make accurate predictions. Its effectiveness is further attributed to its ability to handle high-dimensional feature spaces and its resilience in maintaining performance despite variations in data complexity.

Furthermore, the Random Forest classifier, with its ensemble approach, demonstrates robustness by aggregating the predictions of multiple decision trees. This method not only enhances accuracy but also mitigates overfitting a common issue in machine learning models—thereby improving generalization. This approach is effective in managing diverse and high-dimensional data, offering a balanced performance across different classes.

The Support Vector Machine (SVM) and AdaBoost classifiers also contribute significantly, with SVM achieving high training accuracy and AdaBoost demonstrating strong performance with high precision and recall metrics. These models are proficient in handling complex classification tasks, providing valuable insights into the patterns of violence against women.

Overall, these advanced classification techniques collectively offer a comprehensive framework for understanding and addressing violence against women, with each model contributing unique strengths to the analysis.

REFERENCES

[1] United Nation-Women. (2020) “Intimate partner violence in five Caricom countries: Findings from national prevalence surveys on violence against women”, (May). file:///C:/Users/inbalh/Downloads/20201009CARICOMResearchBrief5.pdf

[2] Assembly, U.G.: Declaration on the elimination of violence against women. UN General Assembly (1993)

[3] World Health Organization Reports on Violence Against Women" - WHO publications provide comprehensive global data on VAW.

[4] Bachman; L E Saltzman , (1995) “Violence Against Women: Estimates From the Redesigned Survey” NCJ Number 154348, Bureau of Justice Statistics (BJS) Address 810 Seventh Street NW, Washington, DC 20531, United States.

[5] World Health Organization (WHO). (2021). Violence Against Women Prevalence Estimates, 2018.

[6] Leslie, E., & Wilson, R. (2020). “Sheltering in place and domestic violence: Evidence from calls for service during COVID-1”9. Journal of Public Economics, 189, 104241

[7] Roesch, E., Amin, A., Gupta, J., & García-Moreno, C. (2020). “Violence against women during COVID-19 pandemic restrictions”. BMJ Global Health, 5(5), e002719

[8] Cattaneo, L. B., & Goodman, L. A. (2021). Empowerment-based domestic violence intervention: A new paradigm. Springer

[9] Henry, N., & Powell, A. (2022). Technology-Facilitated Violence Against Women: Recent Developments and Policy Responses. *Journal of Gender Studies*, 31(5), 673-691.

[10] Australian Institute of Health and Welfare. Family,Domestic and SexualViolence in Australia. (2018).Available online at:https://www.aihw.gov.au/reports/domestic-violence/family-domestic-sexual-violence- in-australia-2018/contents/summary (accessed September, 2021)

[11] Australian Institute of Health and Welfare. Family,Domestic and SexualViolence in Australia: Continuing the National Story. (2019). Availableonline at: https://www.aihw.gov.au/reports/domestic-violence/family-domestic-sexual- violence-australia-2019/contents/summary (accessedSeptember, 2021).

[12] Spasic I, Nenadic G.(2020) “ Clinical text data in machine learning: systematic review” .JMIR Med Inf. 8:e17984. doi: 10.2196/1798.

[13] Ananyan S. Crime pattern analysis through text mining. In: AMCIS 2004Proceedings: 236. New York, NY (2004).

[14] Haleem MS, Han L, Harding PJ, Ellison M.(2019) "An automated text miningapproach for classifying mental-ill health incidents from police incidentlogs for data-driven intelligence". IN: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). Bari: IEEE

[15] Victor B, Perron BE, Sokol R, Fedina L, Ryan JP.(2020) "Automated identification of domestic violence in written child welfare records: leveraging text miningand machine learning to enhance social work research and evaluation" .Soc SocWork Rese.12. doi: 10.1086/712734

[16] Arystianis G, Adily A, Schofield PW, Greenberg D, Jorm L, Nenadic G.(2019) " Automated analysis of domestic violence police reports to explore abuse types and victim injurie"s. J Med 21:e13067. doi:10.2196./13067

Authors

Mr. Dikshendra Daulat Sarpate Pursuing PhD from the VTU Belgavi,India, Currently, he is an HOD,AI&DS Department in the Zeal college of engineering and research, Pune. His area of interests includes Data Science, Artificial Intelligence and Machine Learning and NLP, Generative AI.

Dr. Manju D.Pawar obtained her PhD from the Dr. Babasaheb Amdedkar Marathwada University, Chh. Sambhajinagar, Maharashtra, India. Currently, She is an Associate Professor in the Zeal college of engineering and research, Pune. Her area of interests includes Signal and Image, Artificial Intelligence and Machine Learning and Cloud computing. etc.