International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Hybrid Machine Learning and Regression Approach for Validating a Multi-Dimensional Crime Index in the Context of Crime Against Women

Mr. Siddesh K T [2] K Sinchana [1]

[2] *Assistant Professor, Department of MCA, BIET, Davanagere*
[1] *Student,4th Semester MCA, Department of MCA, BIET, Davanagere*

ABSTRACT

Violence against women has persisted throughout history, manifesting in various forms, including psychological distress, physical harm, and sexual assault. This study presents a crime index based on multiple categories that either directly or indirectly contribute to the emergence of criminal behavior. A composite weighted index was developed, consisting of four sub-indexes that focus on Health, Socioeconomic Status, Education, and the Judiciary. The index's stability and consistency were evaluated through reliability testing.
A hybrid model was formulated by integrating multiple linear regression and robust regression techniques, with random forest and stochastic gradient descent serving as meta-regressors. The models were evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results indicate that the Index exhibits strong reliability, supported by a notably high correlation.

Keywords: *crime index, psychological distress, physical harm, sexual assault, composite weighted index, Health, Socioeconomic Status, Education, Judiciary, reliability testing, baseline models, ensemble models, hybrid model, multiple linear regression, robust regression, random forest.*

I. INTRODUCTION

Violence against women is a pervasive issue that has persisted throughout history, manifesting in various forms, including psychological distress, physical harm, and sexual assault. The complexity of this problem necessitates a comprehensive understanding of the factors that contribute to criminal behavior targeting women. In recent years, there has been a growing recognition of the need for effective tools to assess and address this issue. This study aims to introduce a multi-dimensional crime index specifically designed to evaluate the risk of violence against women, taking into account various socio-economic and environmental factors.
The proposed crime index is a composite weighted index that incorporates four critical sub-indexes: Health, Socioeconomic Status, Education, and Judiciary. By analyzing these dimensions, the index seeks to provide a holistic view of the conditions that foster criminal intentions. The reliability and stability of the index were rigorously tested, ensuring that it can serve as a dependable tool for policymakers and researchers alike.
To validate the effectiveness of the index, a hybrid model was developed, combining multiple linear

regression and robust regression techniques with advanced machine learning methods, including random forest and stochastic gradient descent. This innovative approach allows for a more nuanced understanding of the data, capturing the variance in crime rates with greater accuracy. The models were evaluated using established metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), providing a robust framework for assessing performance.

The findings of this study reveal that women in less developed regions, particularly those facing extreme geographical challenges, are at a heightened risk of victimization. Furthermore, the analysis underscores the significant role that social factors play in the prevalence of violence against women. By identifying these critical elements, the index holds the potential to inform targeted strategies aimed at reducing violence and enhancing the safety of women in various contexts.

In contrast to existing systems that have attempted to quantify crime rates, this study offers a more comprehensive and reliable approach to understanding the dynamics of violence against women. By addressing the limitations of previous models, such as data complexity, availability, and labeling accuracy, the proposed system aims to provide a valuable resource for stakeholders committed to combating this pressing social issue.

II. RELATED WORK

Kierepka introduced an indicator to determine the locations and frequency of crime incidents occurring annually in Wroclaw City. Cartographic studies have been conducted to assess the overall risk of reported offenses, examining the spatial distribution of crime probability in relation to land use conditions and planning directions for Wroclaw. Kwan et al. evaluated the crime index in Hong Kong by comparing the relative severity of fifteen crime categories using Thurstone's scale and calculating the weight of each category. They subsequently developed weighted index of crime intensity based on time series approach

Chaudhari et al. constructed an economic model to explain fluctuations in crime rates through stochastic frontier analysis. Their analysis allows for the identification of the error term in a predictable component under a predetermined set of hypotheses. In the context of a production function, the deterministic term, referred to as technical inefficiency, indicates the gap between the recorded crime rate and the frontier, representing the proportion of unreported offenses.[1]

Nau analyzed the correlation between major crime categories and the accuracy of AGS indexes in determining whether crime rates for 1,069 tracts within the jurisdiction of the Los Angeles Police Department from 2010 to 2014 fell above or below the median and within the highest or lowest quartile. He concluded that the personal crime index serves as a more reliable indicator of urban crime.[2]

Ledingham et al. found that women with disabilities reported a higher prevalence of sexual assault throughout their lives compared to non-disabled women, with those experiencing multiple impairments being at the greatest risk. A significant disparity exists between women with cognitive or multiple impairments and non-disabled women

regarding their exposure to physical or non-physical coercion during their initial sexual experiences.[3]

III. METHODOLOGY

The methodology adopted in this study involves the formulation and validation of a multi-dimensional Crime Index aimed at understanding and predicting crimes against women. This index integrates four key sub-indexes—Health, Socioeconomic Status, Education, and Judiciary—to comprehensively capture the contributing factors. To ensure the index's credibility, both internal and external reliability tests were conducted, and a comparative analysis of machine learning models was performed. The models included baseline techniques, homogeneous ensembles, and a hybrid ensemble. The hybrid model, in particular, combines statistical regression methods with advanced machine learning approaches to enhance predictive accuracy and model robustness.

3.1 Dataset used

The dataset used for developing the Crime Index includes crime-related and socio-economic data collected from official sources in India, particularly focusing on the years 2011 and 2021. The data represents various districts within the state of Rajasthan. These datasets encapsulate a wide range of indicators, including but not limited to health services, educational attainment levels, economic status, and judicial infrastructure. Such a rich dataset allows for the construction of a composite weighted index that reflects the real-world conditions contributing to violence against women.

3.2 Data preprocessing

Before feeding the data into the machine learning models, significant preprocessing steps were undertaken. This included data cleaning to remove missing or inconsistent entries, normalization of different metrics to bring them onto a common scale, and feature transformation to ensure meaningful interpretation of categorical and numerical variables. The preprocessing stage also involved verifying the quality and integrity of the data to prevent biases that could affect model training and predictions. Proper preprocessing ensures that the constructed index and subsequent models reflect reliable and actionable insights.

3.3 Algorithm used

A hybrid modeling approach was employed by combining **Multiple Linear Regression (MLR)** and **Robust Regression** with **Random Forest** and **Stochastic Gradient Descent (SGD)** as meta-regressors. Each algorithm contributed its strengths—linear regression for understanding direct relationships, robust regression for managing outliers, random forest for non-linear interactions, and SGD for scalable learning—thereby enhancing the overall predictive power and resilience of the model.

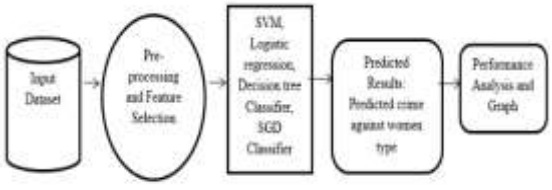

Figure 3.3.1 : System Architecture

3.4 Techniques

The primary technique involved the creation of a **composite weighted index**, integrating data across multiple dimensions to reflect the complexity of crimes against women. To validate the index, both **homogeneous** (using similar algorithms) and **hybrid ensemble techniques** (blending diverse models) were applied. The models were evaluated based on standard error metrics such as **Mean Absolute Error (MAE)**, **Root Mean Square Error (RMSE)**, and **Mean Absolute Percentage Error (MAPE)**.

3.5 Flowchart

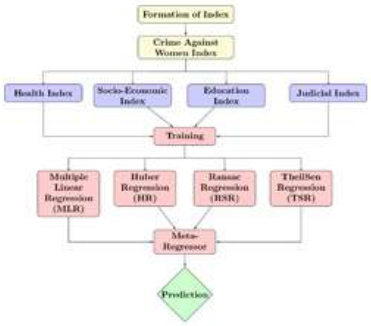

Figure 3.5.1: Flowchart

IV. RESULTS

4.1 Graphs

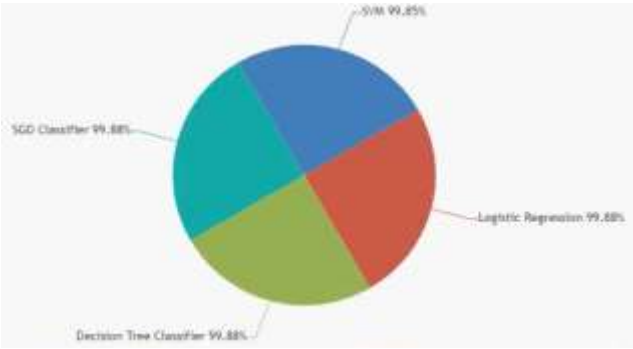

Figure 4.1.1 : Resultant Graph

4.2 Screenshots

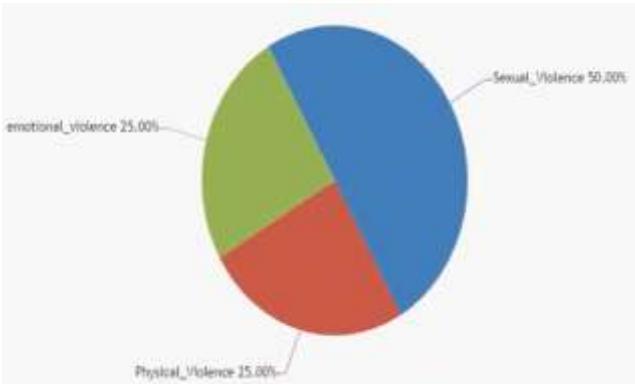

Figure 4.2.1 Crime Against Women Type in pie chart

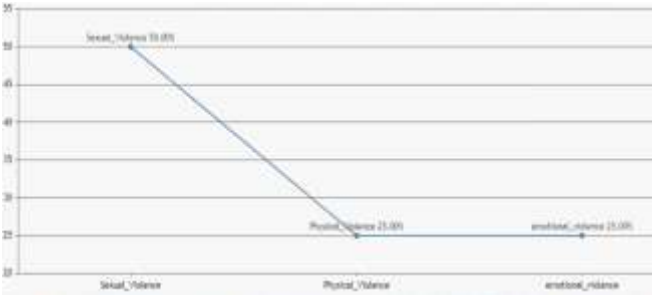

Figure 4.2.2 : Crime Against Women Type in line chart

V. CONCLUSION

This study presents a comprehensive approach to understanding and addressing the issue of violence against women through the development of a multi-dimensional crime index. By integrating various factors such as health, socioeconomic status, education, and judiciary aspects, the proposed index offers a nuanced perspective on the underlying causes of crime against women. The validation of the index through hybrid machine learning techniques demonstrates its reliability and effectiveness in capturing the complexities of crime data. The findings indicate that women in less developed regions, particularly those with extreme geographical conditions, are at a heightened risk of victimization. This highlights the urgent need for

targeted interventions and policies aimed at improving safety and reducing violence against women. The advantages of the proposed system, including the ability to assess the reliability of the index and the use of ensemble hybrid techniques for validation, further enhance its applicability in real-world scenarios. Overall, this research not only contributes to the academic discourse on crime against women but also provides actionable insights for policymakers and stakeholders. By addressing the identified social factors and leveraging the insights from the crime index, there is potential for developing more effective strategies to combat violence against women and promote their safety in society.

VI. REFERENCES

[1] K. Sukhija, S. N. Singh, and J. Kumar, "Spatial visualization approach for detecting criminal hotspots: An analysis of total cognizable crimes in the state of Haryana," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, pp. 1060–1066, May 2017.

[2] A. Anjali and B. R. Kumar, "Spatial analysis of multivariate factors influencing suicide hotspots in urban Tamil Nadu," *J. Affect. Disorders Rep.*, vol. 16, Apr. 2024, Art. no. 100741.

[3] Á. González-Prieto, A. Brú, J. C. Nuño, and J. L. González-Álvarez, "Hybrid machine learning methods for risk assessment in gender-based crime," *Knowl.-Based Syst.*, vol. 260, Jan. 2023, Art. no. 110130.

[4] G. V. Manish, Simran, J. Kumar, and D. K. Choubey, "Identification of hotspot of rape cases in NCT of Delhi:Adata science perspective," in *Proc. Int. Conf. Inf. Syst. Manage. Sci.*, vol. 521. Cham, Switzerland : Springer, 2021, pp. 485–496.

[5] V. Ceccato and A. Loukaitou-Sideris, "Fear of sexual harassment and its impact on safety perceptions in transit environments: A global perspective," *Violence AgainstWomen*, vol. 28, no. 1, pp. 26–48, Jan. 2022.

[6] C. M. Spencer, S. M. Stith, and B. Cafferky, "What puts individuals at risk for physical intimate partner violence perpetration? A meta-analysis examining risk markers for men and women," *Trauma, Violence, Abuse*, vol. 23, no. 1, pp. 36–51, Jan. 2022.

[7] P. K. Saravag and B. R. Kumar, "An application of scan statistics in identification and analysis of hotspot of crime against women in Rajasthan, India," *Appl. Spatial Anal. Policy*, vol. 17, no. 3, pp. 963–982, Sep. 2024.

[8] M. Flood and B. Pease, "Factors influencing attitudes to violence against women," *Trauma, Violence, Abuse*, vol. 10, no. 2, pp. 125–142, Apr. 2009.

[9] S. N. Ogden, M. E. Dichter, and A. R. Bazzi, "Intimate partner violence as a predictor of substance use outcomes among women: A systematic review," *Addictive Behav.*, vol. 127, Apr. 2022, Art. no. 107214.

[10] S. Srivastava, P. Kumar, R. Rashmi, R. Paul, and P. Dhillon, "Does substance use by family members and community affect the substance use among adolescent boys? Evidence from Udaya study, India," *BMC Public Health*, vol. 21, no. 1, pp. 1–10, Dec. 2021.

*****