Predicting Juvenile Crime Severity Using Machine Learning: A Study in Malaysian Rehabilitation Schools

Nurazean Maarop[1], Aslina Mat Asli[2], Ghantan Narayana Samy[3], Roslina Mohammad[4], Wan Rosanisah Wan Mohd[5] and Pritheega Magalingam[6]

[1,3,4,5,6]Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[2]IT Officer, Department of Logistic and Technology, Royal Malaysia Police, Malaysia
[1]nurazean.kl@utm.my mail, [2]aslinaasli@rmp.gov.my, [3]ganthan.kl@utm.my, [4]mroslina.kl@utm.my, [5]wanrosanisah@utm.my, [6]mpritheega.kl@utm.my

Article history

Received:
2 May 2026

Received in revised form:
17 May 2026

Accepted:
30 May 2026

Published online:
15 June 2026

*Corresponding author
nurazean.kl
@utm.my

Abstract

This study aims to classify juvenile crime severity using machine learning techniques based on data collected from 102 juvenile offenders across three rehabilitation schools under the Malaysian Department of Social Welfare. A self-administered survey captured demographic, psychosocial, and behavioral factors, while the target variable was categorized into three levels of crime severity. Three supervised classification models namely Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest were applied to identify predictive patterns, and the Synthetic Minority Oversampling Technique (SMOTE) was employed to address class imbalance. Feature selection and ranking were conducted using Random Forest feature importance, with statistical validation provided through multinomial logistic regression p-values. The results indicate that Random Forest achieved the highest predictive performance, attaining an accuracy of 58.1% before SMOTE and 51.6% after SMOTE, outperforming Logistic Regression (45.2%) and KNN (41.9%). When restricted to the six most important predictors which are Family Background, Emotional Health, Gender, Experience of Premarital Sex, Age Group, and Education Level, the Random Forest model maintained its peak accuracy of 58.1%, indicating that a reduced-feature model can preserve predictive performance while improving interpretability. Within this subset, Family Background and Emotional Health were identified as the most influential features by the Random Forest model. In parallel, multinomial logistic regression analysis identified Gender, Experience of Premarital Sex, and Age Group as statistically significant predictors of crime severity ($p < 0.05$). Overall, the findings suggest that a combination of demographic, behavioral, and psychosocial factors plays a significant role in differentiating juvenile crime severity among Malaysian offenders. This study highlights the potential of AI-driven predictive modelling to support early risk identification, targeted rehabilitation planning, and evidence-based policy development within juvenile justice management.

Keywords: *Machine Learning, Juvenile, Crime Detection, Rehabilitation School, Prediction*

* *Corresponding author. nurazean.kl@utm.my*

1. Introduction

In recent years, Malaysia has witnessed a troubling rise in juvenile offences and disciplinary problems among kids, highlighting the necessity for early identification and predictive intervention measures. Official data reveal that by mid-2025, 367 juvenile crimes were documented at Integrity and Henry Gurney Schools, including drug-related, criminal, and sexual misbehavior instances [1]. From 2023 to June 2025, the Department of Social Welfare recorded 5,633 juvenile cases under the Child Act 2001, of which 3,637 involved individuals aged 16 to 18 years [2]. Alongside criminal crimes, school-related behavioral issues have also escalated. The Malaysian Ministry of Education said that bullying incidences increased from 6,528 in 2023 to 7,681 in 2024, reflecting a 17% rise, with secondary schools comprising over 70% of the recorded instances [3]. The increasing statistics underscore the rising psychological and behavioral difficulties faced by Malaysian teenagers. Thus, implementing data-driven methodologies, including machine learning, may provide more precise forecasting of delinquent behaviors and guide prompt responses in educational and rehabilitative settings.

This research, therefore, aims to leverage machine learning techniques to develop a predictive model for the severity of offenses committed by juveniles in Malaysian rehabilitation schools, building upon existing demographic and perpetrator behavioral data, In particular, this research seeks to: (1) evaluate the predictive accuracy of relevant machine learning models in categorizing juvenile offence severity; (2) ascertain the main predictive factors influencing offence severity; and (3) investigate the degree to which these models can yield actionable insights for rehabilitation strategies.

The findings are anticipated to expand both theoretical and practical knowledge by demonstrating how machine learning-based predictions may improve early diagnosis of at-risk children and guide the development of focused treatments. This study integrates predictive analytics into criminological research, connecting social science inquiry with computational intelligence. These advancements are essential for developing targeted treatments and support systems that promote positive behavioral change and decrease recidivism among juvenile offenders [4].

2. Background

Predictive crime research has developed to tackle urgent public issues by comprehending and forecasting the factors contributing to criminal behavior. Children and adolescents in Malaysia and worldwide may engage in criminal activities owing to a confluence of economic adversity, societal inequities such as child labor and discrimination, familial discord, and psychological or biological influences [5]. Juvenile delinquency is a complicated issue that can be influenced by psychological, social, and environmental factors. This highlights the necessity of performing thorough research to establish evidence-based preventative and intervention measures.

2.1. Juvenile Crime

A review by Ullman et al. [6] examined several manifestations of teenage criminality and violence in persons aged 6 to 18, including aggressiveness, physical violence, antisocial behavior, recidivism, sexual and weapon-related offences, cyber abuse, dating violence, and confinement in secure institutions. The investigation differentiated between adolescence-limited and life-course persistent offending tendencies, utilizing official records, self-reports, and third-party observations. Various criminological theories have attempted to elucidate teenage delinquency, highlighting the significant impact of self-control and social control mechanisms [7]. Juveniles are often characterized as those under 18 years of age who are not considered legal adults, therefore receiving specialized legal treatment, especially regarding criminal culpability [8]. Juvenile crime, or juvenile delinquency, by earlier definition refers to various illegal behaviors perpetrated by adolescents, encompassing both criminal offences and status offenses actions deemed unlawful simply based on the offender's age, such as truancy or curfew breaches [9].

Juvenile schools and rehabilitation centers in Malaysia, including the three to be analyzed in this study, function as essential venues for gathering comprehensive behavioral and contextual data to guide evidence-based treatments [10]. These institutions provide the aggregation of longitudinal data on juvenile offenders, including socio-economic backgrounds, psychological profiles, and reactions to rehabilitation programs [11]. By employing these datasets, machine learning algorithms may discern significant risk and protective characteristics, thereby facilitating the development of tailored preventative intervention techniques [12]. The resultant predictive insights possess practical ramifications for policy development and resource distribution, highlighting initiatives with proven efficacy in mitigating offence severity and recidivism [13]. Moreover, the incorporation of digital mental health technologies in these settings might improve adaptive care via ongoing psychometric assessment, facilitating data-driven treatment strategies that responsively cater to individual requirements [14].

2.2. Machine Learning Application in Crime

This section discusses current research on crime prediction, focusing on methodology and applications pertinent to juvenile delinquency and rehabilitation. The utilization of machine learning in crime analysis has advanced considerably, offering powerful tools to uncover hidden patterns and predict criminal behavior across diverse contexts [15]. Artificial intelligence techniques have been increasingly applied to juvenile delinquency research, where predictive systems support both prevention and rehabilitation strategies. For instance, dense neural networks with long short-term memory architectures have been used to improve the accuracy of offense-type prediction and to guide occupational or rehabilitative interventions [16]. These developments deepen understanding of the multifaceted factors influencing juvenile offence severity, enabling the creation of predictive frameworks that inform targeted and timely intervention strategies.

Ensemble learning methods, such as gradient boosting and random forest algorithms, have demonstrated strong predictive performance across various crime categories, including violent and property crimes [17]. These models typically integrate socioeconomic indicators, demographic characteristics, and historical crime records to reveal complex multivariate relationships [18]. The ability to model such relationships is central to the shift from reactive law enforcement to proactive, data-driven prevention, allowing authorities to anticipate potential offences and allocate resources more effectively [19].

Recent scholarship highlights the transformative potential of machine learning in developing personalized and context-sensitive predictive systems for juvenile crime. By combining diverse datasets including socio-economic indicators, psychological profiles, family history, and behavioral data machine learning models can generate individualized risk assessments and offense severity predictions that move beyond static categorizations [20]. Incorporating spatio-temporal, qualitative, and multimodal data such as social media content, video, or environmental factors further enhances contextual understanding and model precision [21]. Emerging approaches such as reinforcement learning and explainable AI (XAI) also strengthen the interpretability and fairness of predictive outputs, ensuring that ML-driven decisions align with expert reasoning and ethical standards within social and legal contexts.

In Malaysia, the use of predictive analytics in juvenile delinquency research remains nascent, with current studies primarily identifying general risk factors rather than utilizing ML for personalized prediction. Advancing toward localized and culturally adaptive models integrating regional socio-economic, behavioral, and demographic variables could substantially enhance predictive validity and fairness [22]. High-quality, region-specific data are crucial for developing adaptable models that accurately capture Malaysia's socio-cultural diversity. Furthermore, transfer learning and cross-validation techniques can help assess the robustness and generalizability of models trained on international datasets when applied to Malaysian juvenile offender populations. The integration of XAI is equally vital for transparency and stakeholder trust, promoting responsible and interpretable AI use in policy formulation [23].

Logistic Regression is commonly utilized in criminological and behavioral research because of its clarity, interpretability, and suitability for binary classification issues, rendering it effective for discovering statistically significant predictors of juvenile offending. The Random Forest algorithm, which generates numerous decision trees and consolidates their results, improves predictive accuracy and reduces overfitting, providing specific benefits in modelling intricate and non-linear relationships characteristic of socio-behavioral data [24]. K-Nearest Neighbors (KNN), a non-parametric method, is utilized for its capacity to identify localized data structures and adjust to diverse data distributions; however, its efficacy is contingent upon the appropriate selection of the k value and the dimensionality of the feature space. These three models were chosen for their proven efficacy in analogous predictive applications within the criminal justice sector, offering a judicious balance between predictive capability and the necessity for transparency and interpretability in critical decision-making scenarios [25]. Each model was subjected to thorough assessment by a k-fold cross-validation method to guarantee robustness and generalizability.

Overall, machine learning applications in crime prediction signify a paradigm shift from broad statistical profiling to dynamic, individualized, and ethically grounded modeling. These approaches not only improve predictive accuracy but also enhance the capacity of the juvenile justice system to implement evidence-based, culturally sensitive, and preventive interventions aimed at reducing recidivism and promoting long-term rehabilitation outcomes.

2.3. Determinants of Juvenile Crime

The factors influencing adolescent criminality can be examined from multiple perspectives, encompassing both structural and behavioral dimensions of delinquency and drawing on diverse data sources beyond machine learning research. Among these, administrative and registry data including police records, court documents, and national census information are widely utilized in countries such as the United States, India and Sweden, offering objective and comprehensive coverage. Conversely, survey and self-report data obtained from adolescents or families frequently included in research from German and Nigeria offer significant insights on behavioral and contextual dynamics. Certain works further integrate legal and doctrinal resources, including law reviews and court judgements, especially in jurisdictions such as the Indonesia and Turkey, to examine systemic and policy ramifications.

Various predictive factors have consistently appeared as relevant in predicting juvenile criminal risk across different data sets. The most often recognized determinants encompass educational characteristics, parental influences, socioeconomic status, family structure and dynamics, and peer impact [6]. Additional prevalent predictors include school participation, substance use, and the family environment, although factors such as gender disparities, religious or legal considerations, unemployment, and sibling influence are less frequently observed but provide significant contextual nuance [26].

The research examined indicates that the determinants of adolescent delinquency include a complex interplay of familial, socioeconomic, educational, behavioral, and environmental factors. Factors associated to family specifically parental oversight, family composition, quality of relationships, and economic stability were identified as the most often reported predictors [27]. Socioeconomic factors such as poverty, unemployment, and excessive debt were consistently associated with increased risks of teenage criminal behavior [28]. Education serves a crucial preventative function, since school attendance, academic achievement, and educational involvement are negatively correlated with delinquent behavior [27]. Behavioral and peer-related factors, including substance use and peer pressure, manifest as significant situational triggers in many circumstances [28]. Additionally, structural and geographical elements such as urban architecture, street environment, and social infrastructure are recognized as emerging predictors in technologically advanced, data-driven assessments [21]. Community-level resilience, social involvement, and preventative initiatives are acknowledged as protective factors that contribute to the reduction of juvenile criminality in certain areas [29]. Collectively, these characteristics illustrate the multifaceted essence of juvenile delinquency and establish a robust empirical basis for the advancement of machine learning models in this domain.

Table 1 offers a thematic consolidation of the primary predictors discovered in the examined research. The elements are categorized into overarching areas that include demographic parameters, family structure, socioeconomic background, educational environment, peer and community influences, and psychological or behavioral qualities. This classification enhances comprehension of the interconnectedness of these factors and their collective influence on adolescent offending. The summary gives a conceptual basis for the later incorporation of these predictors into the machine learning framework created in this research.

Table 1. Categories of Factors

Source	Category	Factors
[30-31]	Family Factor	Parental supervision, relationship quality, family structure, family size, parental education, and family dysfunction
[28], [32]	Socioeconomic Factors	Family income, poverty, unemployment, economic hardship, parental over-indebtedness, and social disadvantage
[6], [28]	Educational Factors	School attendance, academic performance, education level, and school engagement
[27], [33]	Educational Factors	Drug addiction, substance abuse, peer pressure, aggression, impulsiveness, and sensation seeking
[33-34]	Environmental and Structural Factors	Urban street environment, land use, accessibility, surveillance coverage, and neighborhood safety
[34-35]	Community and Social Factors	Community engagement, youth employment, preventive social programs, and social participation
[34-35]	Sexual and Biological Factors	Sexual drive, early sexual experiences, gender-based behavior, and moral-legal perceptions of sexual offenses

Recent study broadens the understanding of adolescent delinquency by including psychological and sexual-behavioral factors with conventional social and economic variables. Research by Posey et al. [35] highlights the correlation between impulsivity, sensation-seeking, and risk-taking with early participation in delinquent and sexually motivated crimes. Further some studies provide empirical medico-legal data about the influence of biological and sexual impulses, particularly in male adolescents, on the kind and severity of juvenile offences. These results highlight the need for predictive models that include psychological and biological factors, hence improving the ability of machine learning algorithms to encompass the whole profile of delinquency risk across various cultural and behavioral settings.

3. Methodology

This research aims to identify the predictive factors of juvenile crime severity from children aged 12 -18 from three juvenile schools in Malaysia, constituting approximately 80% of the total population, to guarantee interpretability and predictive robustness This study involved 102 juvenile offenders enrolled in these three rehabilitation schools in Malaysia, all operating under the Department of

Social Welfare, Malaysia (Jabatan Kebajikan Masyarakat). Ethical approval was obtained from the relevant authorities and the participating schools.

The researcher utilized the dataset collected through a self-administered survey and registry data in 2019 after obtaining the formal permission by Department of Social Welfare, Malaysia. It captures three primary dimensions: demographic attributes (Gender, Age Group, Education Level, Family Background), psychosocial status (Emotional Health), and behavioral factors (Experience of Premarital Sex). The outcome variable, Crime Severity Category, was classified into three tiers: severe, moderate, and minor. All categorical variables were numerically encoded to facilitate machine learning analysis. Categorical variables were numerically encoded for machine learning classification.

The target variable, Crime Severity Category, was defined as a three-level ordinal variable indicating the severity of offences perpetrated by juvenile respondents. Cases were systematically categorized into: (i) Sever/Serious/Violent Crime (Class 0), encompassing offences that entail physical harm or violent intent (e.g., murder, rape, molestation, robbery, burglary); (ii) Moderate/Organized/Drug-Related Crime (Class 1), involving participation in syndicates, drug activities, and online fraud; and (iii) Minor/Social/Behavioral Misconduct (Class 2), denoting non-violent deviant behaviors such as social misconduct, runaway incidents, extramarital pregnancy, early sexual experience, theft, and physical confrontations. A total of 102 cases were categorized, resulting in a distribution of 45 instances in Class 0, 28 in Class 1, and 29 in Class 2.

In this study, three machine learning algorithms, namely Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest were used to classify offenders according to their crime severity category as shown in Table 2. These three algorithms were identified as being relatively performing for small datasets. Previous studies demonstrate that logistic regression exhibits notable robustness in small-sample scenarios [7], whereas ensemble techniques like Random Forest can yield satisfactory results in small-data environments, albeit with increased caution [36]. Subsequently, the Synthetic Minority Oversampling Technique (SMOTE) was applied to correct class imbalance. The dataset was partitioned into training and testing subsets, utilizing 70% of the data for training and reserving 30% for testing. Every phase of the process encompassing data extraction, cleansing, analysis, and visualization was executed utilizing Python.

This research evaluated three supervised machine learning algorithms Logistic Regression, K-Nearest Neighbours (KNN), and Random Forest for the classification of crime categories using the chosen predictors. Each model underwent training with the training dataset and was subsequently evaluated using the unseen testing dataset. Accuracy scores were calculated for each classifier to evaluate overall predictive performance. Furthermore, classification reports were produced to present class-specific precision, recall, and F1-scores. Confusion matrices were generated for all models and visualized with heatmaps to enhance the interpretation of correct and incorrect classifications across crime categories. The matrices offered a comprehensive depiction of the distribution of prediction errors and the model's tendencies in class differentiation. The performance metrics derived from these models facilitated a comparative assessment of the algorithms regarding predictive accuracy and class-level performance, thus guiding the identification of the most effective method for crime category prediction in this study. Feature selection and ranking were performed with Random Forest feature importance scores, subsequently confirmed by multinomial logistic regression p-values to ascertain the most significant psychological and demographic factors. Hyperparameter optimization for each model utilized grid search with cross-validation, while class weighting was dynamically calculated in Logistic Regression to address imbalance [37]. This thorough and methodical modelling approach guaranteed that the chosen algorithms attained optimal prediction performance while also preserving resilience against the intrinsic constraints of limited and unbalanced datasets [38]. All analyses were performed in Python (Google Colab) utilizing the scikit-learn and stats models' modules.

4. Result

This study identified the target variable as Crime Category based on severity level, with predictor ten features encompassing demographic: School, Gender, Age Group, Race, Education Level, and Premarital Sex Exposure; and behavioral, and psychosocial factors: Risk Factor, Self Esteem Level, Family Background, and Emotional Health. The dataset was analyzed for completeness to confirm the presence of all selected features and the target variable.

4.1. Demographics

The dataset includes several demographic and background factors that reflect personal attributes, social context, and behavioral indications of juvenile respondents. The target variable, Crime Category, was numerically encoded through label encoding to transform categorical class labels into integer values appropriate for machine learning analysis. The Crime Category variable is assessed on an ordinal scale, indicating the escalating severity of offences committed. For analytical reasons, all crimes were categorized into three primary groups, each given a number label for modelling: Severe/Violent Offence (Label = 0), Organized Crime/Drug-Related Offences (Label = 1), Social or Behavioral Misconduct (Classification = 2). Table 2 displays the frequency distribution of instances within these groups. The findings reveal that Serious/Violent Crime (Label 0) constitutes the highest share of instances, with 45 among the 102 incidents (44.1%). This category encompasses severe actions, including violent acts, sexual related crimes, and robbery-related activities. The second-highest category is Social/Behavioral Misconduct (Label 2), including 29 instances (28.4%). Despite their relatively low legal severity, these actions indicate significant psychological and developmental hazards, including social issues, impulsive conduct, or other non-criminal transgressions. Organized/Drug-Related Crime (Label 1) contains 28 instances (27.5%). This category encompasses offences linked to property crimes, theft, and drug-related activities, indicating a moderate degree of criminal engagement.

Table 2. Crime Category

Category	Label	Count
Serious/Violent Crime Serious	0	45

Organized/Drug-Related Crime	1	28
Social/Behavioral misconduct	2	29

The distribution indicates a significant dispersion across the three categories, with a little greater concentration in major or violent offences. This underscores the varied behavioral tendencies among the juvenile population and reinforces the need for tailored intervention strategies based on the severity and nature of the incident.

The other demographic variables are a combination of nominal and ordinal scales, aligning with the requirements of behavioral and social scientific research. This categorization permits precise statistical analysis and proper interpretation of juvenile behavioral patterns. School, Gender, Race, Premarital Sex Exposure, and Risk Influence are assessed as nominal variables, representing categorical data devoid of intrinsic hierarchy. School signifies the rehabilitation facility attended, whilst Gender and Race indicate fundamental demographic characteristics. Premarital Sex Exposure is a dichotomous variable (Yes/No), whereas Risk Influence denotes kinds of influence (peer, family, or other).

Multiple variables are categorized as ordinal according to their intrinsic hierarchy. The Age Group categorizes respondents into early adolescent (ages 12 to 15 years) and late adolescence (16–18 years). Education Level indicates ascending academic achievement (from Lower to Secondary). Self-esteem encompasses varying degrees of personal self-perception (low, moderate, high), while family background adheres to a hierarchical framework indicative of household stability (Stable, Social-Risk, and Economically Challenged). Finally, Emotional Health is classified according to subjective well-being (Poor, Moderate, and Good).

4.2. Machine Learning Modelling Result

The dataset includes several demographic and background factors that reflect personal attributes, social context, and behavioral indications of juvenile respondents. The target variable, Crime Category, was numerically encoded through label encoding to transform categorical class labels into integer values appropriate for machine learning analysis.

The baseline classification results obtained without class balancing reveal moderate predictive performance across all evaluated models. Logistic Regression achieved an overall accuracy of 45.2%, demonstrating stronger discrimination for Severe/Violent Crime but limited sensitivity to minority classes. K-Nearest Neighbors yielded a lower accuracy of 41.9%, with particularly weak recall for Organized/Drug-Related Crime, indicating sensitivity to class imbalance. Among the baseline models, Random Forest achieved the highest performance, with an accuracy of 58.1%. As summarized in Table 3, Random Forest consistently outperformed the other classifiers across both overall accuracy and class-level metrics. The model demonstrated comparatively balanced recall across all three crime categories, particularly improving the identification of Social/Behavioral Misconduct cases. These findings suggest that ensemble-based learning is more effective in capturing the complex, non-linear relationships present in juvenile behavioral data. Emotional Health. The dataset was analyzed for completeness to confirm the presence of all selected features and the target variable.

Table 3. Baseline Model Performance

Model	Accuracy	Macro Precision	Macro Recall	F1 Score
Logistic Regression	45.16	0.43	0.47	0.43
K-Nearest Neighbors	41.94	0.38	0.38	0.37
Random Forest	58.06	0.59	0.62	0.56

4.3. Effect of Class Balancing Using SMOTE

The baseline classification results obtained without class balancing reveal moderate predictive performance across all evaluated models. Following class balancing, changes in model behavior were observed. Table 4 presents the classification performance of the three machine learning models after applying the Synthetic Minority Oversampling Technique (SMOTE). Logistic Regression exhibited improved sensitivity toward minority classes, reflected in an increase in macro recall from 0.47 to 0.54. However, the overall accuracy gain remained modest. K-Nearest Neighbors demonstrated a decline in performance across all evaluation metrics, suggesting that distance-based classifiers are less robust when trained on synthetically oversampled categorical data. In contrast, Random Forest maintained relatively stable performance, achieving the highest post-SMOTE accuracy (51.61%) and balanced macro-averaged metrics, reinforcing its robustness in handling imbalanced juvenile crime datasets. This result reinforces the robustness of ensemble methods in handling imbalanced juvenile crime data.

Table 4. Model performance after SMOTE class balancing

Model	Accuracy	Macro Precision	Macro Recall	F1 Score
Logistic Regression	48.39	0.47	0.54	0.47
K-Nearest Neighbors	35.48	0.34	0.37	0.34
Random Forest	51.61	0.51	0.53	0.49

Macro-averaged precision, recall, and F1-score are reported to mitigate the influence of class imbalance and to ensure equal contribution from all crime categories. These findings indicate that while SMOTE enhances minority-class recall, ensemble-based models remain superior in preserving overall classification stability for small and imbalanced juvenile datasets.

4.4. Feature Importance and Statistical Validation

A reduced-feature Random Forest model using six key predictors achieved an accuracy of 58.1%, matching the best-performing baseline model. Importantly, class-level recall and F1-scores were more evenly distributed across crime categories, indicating improved generalization despite reduced dimensionality. This finding demonstrates that a compact set of psychosocial and demographic predictors is sufficient to retain predictive performance, while enhancing model interpretability and reducing noise. baseline classification results. The Feature importance analysis identified Family Background and Emotional Health as the most influential predictors, followed by Gender and Experience of Premarital Sex as shown in Figure 1. Age Group and Education Level contributed moderately to classification outcomes.

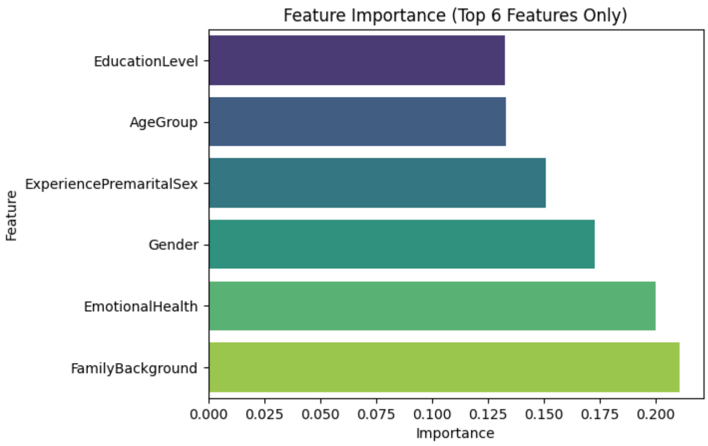

Figure 1. Feature Importance (Random Forest)

Subsequently, further multinomial logistic regression analysis provided inferential support for these findings. Age Group was a significant predictor of Organized/Drug-Related Crime ($p < 0.05$), while Gender and Premarital Sex Exposure were significantly associated with Social/Behavioral Misconduct ($p < 0.05$). These results corroborate the machine learning feature rankings and highlight the importance of psychosocial and developmental factors in juvenile crime severity.

4.5. Descriptive Pattern Consistency

Normalized cross-tabulation analysis revealed descriptive patterns that were consistent with the machine learning and statistical modelling results. Juveniles from disadvantaged family backgrounds exhibited higher proportions of Social/Behavioral Misconduct, while those from more stable family environments were more frequently associated with Severe/Violent Crime categories.

Respondents reporting experience of premarital sex also showed a substantially higher likelihood of being classified under Social/Behavioral Misconduct, supporting its prominence as a predictive behavioral factor

Age-related differences were evident, with older juveniles (16–18 years) demonstrating increased involvement in Organized/Drug-Related Crime, whereas younger offenders were more frequently associated with violent or impulsive offenses. Education level further differentiated crime severity, as juveniles with lower educational attainment were disproportionately represented in minor and behavioral misconduct categories.

Overall, the results consistently indicate that family background, emotional health, and premarital sexual experience are dominant predictors of juvenile crime severity. Random Forest emerged as the most robust classifier across all experimental settings, while class balancing improved minority-class sensitivity without uniformly increasing overall accuracy. Importantly, reduced-feature models retained predictive strength, reinforcing the conclusion that a compact set of psychosocial and demographic variables is sufficient for meaningful juvenile crime severity classification.

5. Discussion

This study examined the applicability of machine learning techniques for classifying juvenile crime severity within Malaysian rehabilitation schools, while simultaneously identifying key demographic, psychosocial, and behavioral predictors. Overall, the findings demonstrate strong convergence with prior criminological and predictive analytics research, while extending existing knowledge through localized, data-driven modeling and multi-method validation.

5.1. Alignment with Prior Empirical Evidence

The prominence of family background as a key predictive factor is highly consistent with prior research across diverse socio-cultural contexts. Studies conducted in India [27, 30], Ukraine [31], Sweden [39], and Russia [40] consistently identify family structure, parental supervision, and socioeconomic stability as central determinants of juvenile offending. The present study corroborates these findings using a machine learning framework, demonstrating that family background not only predicts delinquency risk but also differentiates crime severity categories, particularly social and behavioral misconduct. This extends earlier work by showing how family instability manifests differently across offence types rather than merely increasing overall delinquency risk.

Emotional health emerged as one of the most influential predictors in the Random Forest model, reinforcing psychological perspectives on juvenile crime. This finding aligns closely with the systematic review by Posey et al. [35], which emphasizes impulsivity, emotional dysregulation, and sensation-seeking as core mechanisms underlying adolescent offending. Unlike many prior studies that rely on self-report correlations or qualitative inference, the present study integrates emotional health into a predictive model and validates its influence across both machine learning feature importance and statistical inference, strengthening its empirical relevance.

Age group differences observed in this study further support developmental criminology theories. Older juveniles (16–18 years) exhibited higher likelihoods of involvement in Organized or Drug-Related Crime, while younger adolescents were more frequently associated with impulsive or violent offenses. This pattern aligns with longitudinal findings reported by Ullman et al. [6] and registry-based analyses by Rojas and Bäckman [39], which describe a transition from impulsive to more structured offending with age. The present study extends this trajectory-based understanding by operationalizing age effects within a multi-class predictive framework.

Gender also demonstrated statistically significant effects, particularly in relation to Social/Behavioral Misconduct. This finding aligns with earlier criminological and medico-legal studies documenting gender-based behavioral differences in juvenile offending [5,41]. By validating gender effects through multinomial logistic regression within a Malaysian institutional context, this study contributes localized evidence to a well-established global pattern.

5.2. Methodological Contribution

From a methodological standpoint, this study demonstrates the effectiveness of combining ensemble-based machine learning, class imbalance handling, feature reduction, and statistical inference for juvenile crime classification. Using self-administered survey data capturing demographic, psychosocial, and behavioral factors, the results show that such structured survey-based datasets can be effectively leveraged for machine learning applications when appropriate preprocessing and validation strategies are employed.

Among the evaluated models, Random Forest consistently outperformed Logistic Regression and KNN across baseline, SMOTE-adjusted, and reduced-feature settings, reinforcing the suitability of ensemble learning for modelling complex socio-behavioral data [17-18]. Although the application of SMOTE improved minority-class sensitivity, particularly for Logistic Regression, it did not consistently enhance overall accuracy. This finding aligns with prior studies reporting that synthetic oversampling may yield limited benefits when applied to categorical and psychosocial survey data [28].

Notably, the reduced-feature Random Forest model retained equivalent predictive accuracy while improving interpretability, demonstrating that dimensionality reduction is feasible without compromising performance when key psychosocial predictors are preserved [27]. A further methodological contribution lies in the integration of machine learning feature importance with multinomial logistic regression p-values, enabling both predictive accuracy and inferential transparency. This hybrid validation approach strengthens the robustness of the proposed framework and supports its applicability for evidence-based analysis in juvenile justice research.

5.3. Implication for Policy and Practice

The convergence between this study's findings and international evidence underscores the applicability of machine learning-based predictive tools in juvenile justice systems. For Malaysian rehabilitation schools, the identified predictors family background, emotional health, sexual behavior, age, education level, and gender offer actionable insights for early risk screening, tailored rehabilitation planning, and resource prioritization. The results suggest that integrating AI-driven models into existing assessment frameworks could enhance individualized intervention strategies without replacing professional judgment.

6. Conclusion

This study demonstrates that machine learning techniques, particularly Random Forest, can effectively classify juvenile crime severity using a compact and interpretable set of demographics, psychosocial, and behavioral predictors. By combining predictive modelling with statistical inference, the findings extend existing criminological research through empirical validation within Malaysian rehabilitation institutions, offering context-specific insights that complement international evidence. Family background, emotional health, premarital sexual experience, age group, education level, and gender consistently emerged as salient predictors, reinforcing the central role of social and developmental factors in shaping juvenile offending patterns. These results underscore the potential of AI-driven analytics as a decision-support tool for evidence-based intervention planning, rehabilitation strategies, and juvenile justice policy formulation.

Nevertheless, this study is subject to several limitations. The relatively small sample size, although representative of the institutional population, may constrain generalizability beyond similar settings. The use of self-reported survey data may introduce response bias, particularly for sensitive psychosocial and behavioral variables. In addition, the application of SMOTE yielded mixed performance gains, highlighting its limited effectiveness for categorical psychosocial data, while crime severity was modeled as a nominal multiclass outcome despite its ordinal nature. Future research should leverage larger and longitudinal datasets, explore ordinal and cost-sensitive learning frameworks, and incorporate multi-source administrative and clinical data to enhance model robustness, external validity, and real-world applicability.

Acknowledgments

The authors gratefully acknowledge the financial support provided by Universiti Teknologi Malaysia under the Potential Academic Staff Grant (Q.K130000.2757.03K69). The authors also extend their appreciation to the Department of Social Welfare, Malaysia and the Royal Malaysia Police for their cooperation and support in facilitating this research.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

References

[1] Khoo, G., Yunus, A. and Ibrahim, J. (2025). Over 360 Juvenile Offences Recorded in Integrity, Henry Gurney Schools in First Half of 2025. The Star. Available at: https://www.thestar.com.my/news/nation/2025/08/06/nearly-360-juvenile-offences-recorded-in-intergrity-henry-gurney-schools-in-first-half-2025

[2] Gimino, G., Yunus, A. and Tan, T. (2025). Over 5,000 Juvenile Cases Recorded Since 2023, Says Home Minister. The Star Online. Available at: https://www.thestar.com.my/news/nation/2025/08/20/over-5000-juvenile-cases-recorded-since-2023-says-home-minister

[3] The Straits Times. (2025). Over 7,600 Bullying Cases in Malaysian Schools in 2024, Up 17% From 2023. The Straits Times, 27 August 2025. Available at: https://www.straitstimes.com/asia/se-asia/over-7600-bullying-cases-in-malaysian-schools-in-2024-up-17-malaysian-education-minister

[4] Su, R., John, J. R. and Lin, P. (2023). Machine Learning-Based Prediction for Self-Harm and Suicide Attempts in Adolescents. Psychiatry Research 328 (2023) 115446.

[5] Hazra, D. (2021). Determinants of Juvenile Crime: Evidence from India. International Journal of Social Economics 48:12 (2021) 1740–1767.

[6] Ullman, R., Lereya, S. T., Glendinning, F., Deighton, J., Labno, A., Liverpool, S. and Edbrooke-Childs, J. (2024). Constructs Associated with Youth Crime and Violence Amongst 6–18 Year Olds: A Systematic Review of Systematic Reviews. Aggression and Violent Behavior 75 (2024) 101906.

[7] Garofalo, C. and Sijtsema, J. J. (2022). Clinical Forensic Psychology: Introductory Perspectives on Offending. Cham, Switzerland: Palgrave Macmillan.

[8] Laws of Malaysia. (2006). Child Act 2001 (Act 611). Kuala Lumpur: The Commissioner of Law Revision, Malaysia.

[9] Siegel, L. J. and Welsh, B. C. (2005). Juvenile Delinquency: Theory, Practice, and Law. Belmont, CA: Wadsworth/Thomson Learning.

[10] Amdan, M. A. B., Janius, N., Jasman, M. N. B. and Kasdiah, M. A. H. B. (2024). Advancement of AI-tools in learning for technical vocational education and training (TVET) in Malaysia (empowering students and tutor). International Journal of Science and Research Archive 12:1 (2024) 2061.

[11] Jamaluddin, F., Jamaluddin, A. H., Jamaluddin, F. and Jamaluddin, F. (2025). Malaysia's AI-driven education landscape: Policies, applications, and comparative insights for a digital future. arXiv, arXiv:2509.21858. Available at: https://doi.org/10.48550/arXiv.2509.21858

[12] Na, M., Deli, M. M. and Rauf, U. A. A. (2024). Predicting Mainland Chinese students in Malaysia's AI-based chatbot satisfaction and academic performance: Mediating moderating analysis. Research Square. Available at: https://doi.org/10.21203/rs.3.rs-5322062/v1

[13] Bakar, M. A. A., Ab Ghani, A. T., Abdullah, M. L., Ismail, N. and Ab Aziz, S. (2022). Adaptive Neuro-Fuzzy Inference System (ANFIS) Formulation To Predict Students' Neuroscience Mechanistic: A Concept Of An Intelligent Model To Enhance Mathematics Learning Ability. TEM Journal 11:4 (2022).

[14] Caulley, D., Alemu, Y., Burson, S., Bautista, E. C., Tadesse, G. A., Kottmyer, C. and Sezgin, E. (2023). Objectively quantifying pediatric psychiatric severity using artificial intelligence, voice recognition technology, and universal emotions: Pilot study for artificial intelligence-enabled innovation to address youth mental health crisis. JMIR Research Protocols 12:1 (2023) e51912.

[15] Lu, H., Chen, C., Ma, Y. and Ma, Y. (2025). Lightweight Deep Learning Model for Crime Pattern Recognition Based on Transformer with Simulated Annealing Sparsity and CNN. Scientific Reports 15:1 (2025).

[16] Hou, F. (2022). [Retracted] Echoing Mechanism of Juvenile Delinquency Prevention and Occupational Therapy Education Guidance Based on Artificial Intelligence. Occupational Therapy International, 2022 (1), 9115547.

[17] Lamari, Y., Freskura, B., Abdessamad, A., Eichberg, S. and De Bonviller, S. (2020). Predicting spatial crime occurrences through an efficient ensemble-learning model. ISPRS International Journal of Geo-Information 9:11 (2020) 645.

[18] Sankara, S. and Sugitha, N. (2024). Crime rate analysis and mapping from socio-economic data using deep neural networks. Journal of Computer Science 20:10 (2024) 1203–1213.

[19] Khan, Z., Ali, A., Khan, D. M. and Aldahmani, S. (2024). Regularized ensemble learning for prediction and risk factors assessment of students at risk in the post-COVID era. Scientific Reports 14:1.

[20] Law, J. and Abdullah, A. Y. M. (2024). An Offenders-Offenses Shared Component Spatial Model for Identifying Shared and Specific Hotspots of Offenders and Offenses: A Case Study of Juvenile Delinquents and Violent Crimes in the Greater Toronto Area. Journal of Quantitative Criminology 40:1 (2024) 75–98.

[21] Li, M. and Zhang, Y. (2023). Integrating Social Media Data and Historical Stock Prices for Predictive Analysis: A Reinforcement Learning Approach. International Journal of Advanced Computer Science & Applications 14:12.

[22] Quijano-Sánchez, L., Liberatore, F., Rodríguez-Lorenzo, G., Lillo, R. E. and González-Álvarez, J. L. (2021). A Twist in Intimate Partner Violence Risk Assessment Tools: Gauging the Contribution of Exogenous and Historical Variables. Knowledge-Based Systems 234 (2021) 107586.

[23] Ehsan, H. R. U., Khan, R. A., Yasmeen, R. and Arif, M. (2024). Development and Validation of an AI-Based Model to Predict the Assessment Outcomes of Pre-Clinical MBBS/BDS Students.

[24] Kosgolla, J. V., Smith, D. C., Begum, S. and Reinhart, C. A. (2023). Assessing the Self-Reported Honesty Threshold in Adolescent Epidemiological Research: Comparing Supervised Machine Learning and Inferential Statistical Techniques. BMC Medical Research Methodology 23:1 (2023) 210.

[25] Prasetyo, M. L., Peranginangin, R. A., Martinovic, N., Ichsan, M. and Wicaksono, H. (2025). Artificial Intelligence in Open Innovation Project Management: A Systematic Literature Review on Technologies, Applications, and Integration Requirements. Journal of Open Innovation: Technology, Market, and Complexity 11:1 (2025) 100445.

[26] Karzhaubayeva, L., Otarbayev, G., Issayeva, Z., Milova, Y. and Bissenova, M. (2024). The Role of Traditions and Customs in the Prevention of Juvenile Delinquency in the Republic of Kazakhstan. Scientific Herald of Uzhhorod University, Series "Physics" 55 (2024) 2220–2232.

[27] Abhishek, R. and Balamurugan, J. (2023). An Extensive Study on Drug Abuse Among Medical and Paramedical Learners. Ymer 22:1 (2023) 454–462.

[28] Nesa, M., Shaha, T. R. and Yoon, Y. (2022). Prediction of Juvenile Crime in Bangladesh Due to Drug Addiction Using Machine Learning and Explainable AI Techniques. Journal of Computational Social Science 5:2 (2022) 1467–1487.
[29] Koryakina, Z. I. (2022). Social Factors of Youth Crime Decline in the Republic of Sakha (Yakutia). Sociological Studies 4 (2022) 93–104.
[30] Changalasetty, S. B., Belgacem, B., Badawy, A. S., Ghribi, W., Ahmed, A. M., Bangali, H. and Pemula, R. (2019). Assessing the Relation Between Family Background and Juvenile Delinquency Using Data Mining. Proceedings of the 2019 International Conference on Computer Communication and Informatics (ICCCI), 1–4.
[31] Poltava, K. O., Dubovych, O. V., Serebrennikova, A. V., Sozanskyy, T. I. and Krasnytskyi, I. V. (2020). Juvenile Offenders: Reasons and Characteristics of Criminal Behavior. International Journal of Criminology and Sociology 9 (2020) 1573–1578.
[32] Alati, R. and Ayano, G. (2023). Risk and Protective Factors of Youth Crime: An Umbrella Review. Population Medicine, 5(Suppl.).
[33] Erdmann, A. (2022). The Impact of Peer Groups and Routine Activities on The Victim-Offender Overlap: Evidence From a German Study on Youth Crime. International Criminal Justice Review 32:2 (2022) 178–198.
[34] Wong, S. K. (2017). Divorce, Female and Male Single Parents and Youth Crime: A Social Disorganization Framework. In Grant, S. (Ed.), Divorce: Risk Factors, Patterns and Impact on Children's Well-Being (pp. 1–28). Hauppauge, NY: Nova Science Publishers.
[35] Posey, B. M., Timmer, A. and Ramirez, N. G. (2024). Lessons Learned and Yet to Be Learned From Predictors of Youth Crime Research. Discover Psychology 4 (2024) 152.
[36] Zantvoort, K., Nacke, B., Görlich, D., Hornstein, S., Jacobi, C. and Funk, B. (2024). Estimation of Minimal Data Set Sizes for Machine Learning Predictions in Digital Mental Health Interventions. Digital Medicine 7:1 (2024) 361.
[37] Sundaravadivel, P., Isaac, R. A., Elangovan, D., KrishnaRaj, D., Rahul, V. L. and Raja, R. (2025). RETRACTED ARTICLE: Optimizing Credit Card Fraud Detection with Random Forests and SMOTE. Scientific Reports 15:1 (2025) 17851.
[38] Meegahapola, L., Labhart, F., Phan, T. and Gática-Pérez, D. (2021). Examining the Social Context of Alcohol Drinking in Young Adults with Smartphone Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5:3 (2021) 1.
[39] Rojas, Y. and Bäckman, O. (2024). Parental Over-Indebtedness and Youth Crime in Sweden: A Nationwide Register-Based Study. International Journal of Child, Youth and Family Studies 15:2 (2024) 47–66.
[40] Zhegusov, Y. I. and Koryakina, Z. I. (2022). Social Factors of Youth Crime Decline in the Republic of Sakha (Yakutia). Sociological Studies 4 (2022) 93–104.
[41] Jakhar, J. K., Kaur, A. D., Dagar, T., Dhattarwal, S. K. and Khanagwal, V. P. (2016). A Medico-Legal Study on Juvenile Criminals in State of Haryana. Medico-Legal Update 16:2 (2016) 46.
[42] Salleh, N., Daud, S. M., Sabri, S. F., Ahmad, N. A., Shariff, S.A., Yusof, Y. M. and Adam, M. Z. (2016). Enhancing Temperature Control Method of Thermal Vacuum Chamber for Satellite Testing using Optimization Algorithm: A Review. JurnalTeknologi (Sciences & Engineering) 78: 5–7 (2016) 1–6.