PulisAI: Web-based App Crime Analysis for Identifying Hotspots and Crime Patterns in Angeles City

Alberto Lloren
Angeles University Foundation
Angeles City
Pampanga, Philippines

Angelie Merlin
Angeles University Foundation
Angeles City
Pampanga, Philippines

Charles Laxamana
Angeles University Foundation
Angeles City
Pampanga, Philippines

Melissa Pantig
Angeles University Foundation
Angeles City
Pampanga, Philippines

ABSTRACT
Effective policing strategies for mitigating and preventing crimes are currently a significant challenge for law enforcement in the Philippines. While police patrolling helps reduce crime risks through visibility, limitations on resources and technology in police departments heavily hinder the efficiency of traditional crime management. To address operational challenges in resource allocation for crime prevention and response, this study developed an integrated system, "PulisAI", utilizing machine learning for multi-level crime alarm (low, medium, high) prediction for eight high-priority focus crimes. A local crime dataset, alongside publicly available demographic and geographic information, was put into a comprehensive data preprocessing pipeline. Crime incidents were aggregated into unique spatio-temporal blocks, followed by feature engineering to derive the three-tier alarm level target variable and historical crime metrics. A comparative analysis of four machine learning models—Multinomial Logistic Regression, Random Forest, Support Vector Classifier, and XGBoost Classifier—was conducted under imbalanced data and SMOTE-balanced training conditions. The results showed that the XGBoost Classifier, trained on the original imbalanced dataset, is the superior model, achieving an accuracy of 92.69%, a macro-averaged F1-Score of 87.00%, and an Unweighted Average Recall of 82.00%. The selected model was integrated into the PulisAI web application. PulisAI served as a data-driven tool for decision-making for law enforcement in patrol planning, resources allocation, and crime pattern analysis.

General Terms
Algorithms, Machine Learning, Supervised Model, Artificial Intelligence, Decision-making Tool

Keywords
Predictive policing, Machine learning, Crime hotspot, XGBoost, Spatio-temporal analysis, Class imbalance, Crime patterns, Prediction, Police patrolling, Multi-class classification

1. INTRODUCTION
Crime is an unlawful action of a person that negatively affects society, punishable depending on its severity. It can lead to community disturbance and distrust, fear, inequality, poverty, economic loss, and even cost of life. Despite the government's efforts to avert them, violence and terrorism also remain significant concerns. In fact, theft, rape, and physical injury topped the crimes recorded between July 2022 and January 2023 as per Philippine National Police (PNP) [7]. As Ayeng [4] reported, the Philippines was the only Asian country included among the top 50 countries with the highest crime rates within the Southeast Asia region for 2024. Furthermore, Manila was recorded as the city with the highest crime rate among Southeast Asian cities. Thus, this signifies the ongoing challenges the Philippines faces in addressing its widespread occurrence of crimes.

Police patrolling is essential in monitoring and ensuring the safety of the areas where law enforcement is being deployed. However, according to Gom-gom-o [12] and Annang et al. [2], the effectiveness of police patrolling and response is often hampered by significant resource constraints such as inadequate equipment, vehicles, funding, and technology faced by police departments in the Philippines. This context underscores the critical need for strategies that maximize the efficiency of available police resources.

In the field of criminology and policing, the application of machine learning (ML) transitions predictive strategies from reactive policing to proactive. The prevalent use of supervised ML algorithms for spatio-temporal hotspot prediction has been widely confirmed as a method to optimize resource allocation and strategic interventions [6,13]. However, while foundational studies have explored predictive analytics in the Philippines [3,5], there remains a distinct lack of research applying advanced, multi-class predictive policing models within specific local urban environments.

To address this gap, this study developed "PulisAI," an integrated web application utilizing the XGBoost classifier to predict potential crime alarms across a three-tier classification framework (Low, Medium, and High risk). By modelling complex spatio-temporal and demographic data, this research contributes a vital, data-driven policing tool specifically designed for the unique crime landscape of Angeles City, Pampanga, aiding law enforcement to better manage the risk of crime occurrences and hotspots.

2. RELATED LITERATURE
The transition from reactive to proactive law enforcement is increasingly facilitated by data-driven methodologies,

specifically machine learning (ML). Systematic reviews by Jenga et al.[13] and Butt et al. [6] confirm that supervised ML is the primary tool for analyzing spatio-temporal crime patterns, shifting policing strategies toward strategic interventions.

2.1 Machine Learning Approaches in Crime Prediction

The field of criminology frequently utilizes a diverse array of machine learning algorithms, reflecting attempts to find the most effective models for different datasets and prediction goals.

2.1.1 Tree-Based Ensemble Methods

The literature highlights tree-based ensemble methods as particularly powerful due to their ability to model complex, non-linear relationships. Random Forest (RF) stands out as a frequently utilized algorithm for both classification and regression. Yao et al. [18] found that RF significantly improved accuracy by integrating spatial and demographic data, while Wubineh [16] reported an 86.07% accuracy in crime type prediction.

Recent benchmarks, however, suggest that Extreme Gradient Boosting (XGBoost) may offer superior performance for tabular crime data. Safat et al. [15] identified XGBoost as a top-performing architecture, noting its sequential building process—where each new model corrects the errors of its predecessor—makes it exceptionally robust for identifying patterns in high-risk locations [9].

2.1.2 Linear and Boundary-Based Classifiers

Alongside ensemble methods, Support Vector Classifier (SVC) and Multinomial Logistic Regression (MLR) remain prominent in the literature. SVC operates by identifying an optimal hyperplane to separate data into distinct categories—such as three-tier alarm levels—within a high-dimensional space. Conversely, MLR serves as a statistical baseline known for high interpretability, as demonstrated in the crime alarm prediction tasks of Yadhunath et al. [17]. However, literature suggests these linear models often struggle to capture the non-linear interactions inherent in urban crime data—such as the temporal relationship between time of day and suspect density—when compared to more advanced ensemble-based classifiers.

2.2 Factors and Enhancement Techniques

The efficacy of predictive models is contingent upon the quality of input features. Studies by Balahadia et al. [5] and Yunus and Loo [19] highlight that basic spatio-temporal data must be enriched with environmental context—such as proximity to police stations, population density, and land area—to achieve actionable insights. In the Philippine context, Balahadia et al. [5] demonstrated that detailed attributes like exact date, time, and specific location obtained from local police records are vital for model accuracy.

A recurring challenge in criminology datasets is class imbalance, where "High Alarm" events are far less frequent than "Low" ones. Researchers often employ the Synthetic Minority Over-sampling Technique (SMOTE) to address this by generating synthetic minority samples [8, 17]. However, modern algorithms like XGBoost and Random Forest often provide internal weighting mechanisms (class_weight) that can sometimes outperform artificial oversampling in specific localized contexts. Furthermore, hyperparameter optimization using techniques like Grid Search is essential to refine model performance for specific urban environments [1].

2.3 Synthesis and Research Gap

While predictive policing is well-documented globally, there is a distinct lack of research applying advanced multi-class (three-tier) classification models within the specific urban context of the Philippines. Local studies in provinces like Laguna [3,5] have laid the groundwork for predictive analytics, but these often focus on individual crime reports or simpler binary predictions. A localized system tailored for the unique crime landscape, demographic profile, and specific "focus crimes" of Angeles City, Pampanga, remains missing. This study addresses this gap by implementing and evaluating a system that compares XGBoost, Random Forest, SVC, and MLR to provide a validated, data-driven pathway for multi-level crime alarm prediction.

3. METHODOLOGY

3.1 Methodological framework

This study utilized a quantitative, developmental research design guided by the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework to develop a predictive model for crime alarm levels. It started from conducting an interview for alignment with law enforcement operational goals and historical crime data. Following rigorous data cleaning and spatio-temporal feature engineering, four supervised machine learning classifiers were trained and evaluated using metrics robust to class imbalance. Ultimately, the top-performing XGBoost model was deployed into an interactive web application, evaluated by police personnel before deploying into a web application.

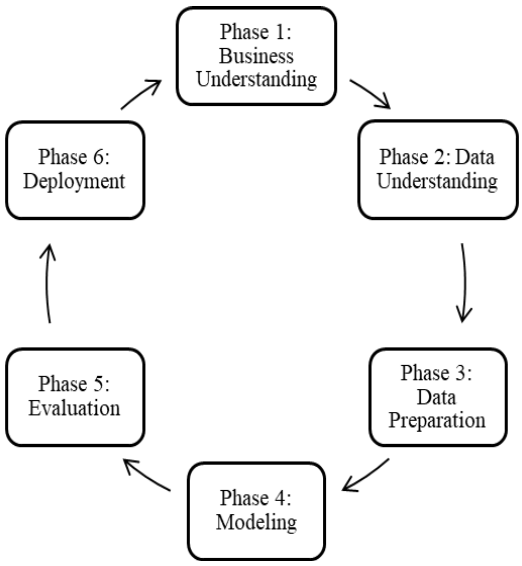

Figure 1. PulisAI CRISP-DM framework

3.2 Sources, Collection, and Preprocessing

3.2.1 Data Collection

The primary data consisted of official anonymized crime incident reports acquired from the Angeles City Police Office (ACPO), covering January 2017 to April 2025. The raw dataset initially comprised 11,896 unique crime incidents, linked to 19,740 suspect records and 7,592 victim records. To provide critical spatial and demographic context, secondary data was gathered from public government portals. Specifically, the population statistics for the 33 barangays of Angeles City were sourced from the Philippine Statistics Authority (Philippine Statistics Authority, 2024) while land area sizes were retrieved from the Department of Interior and Local Government [10]. Geographic coordinates for local police stations were extracted via Google Maps to calculate law enforcement proximity metrics.

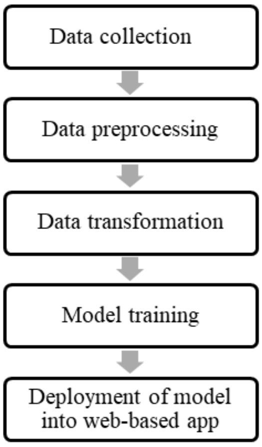

Figure 2. PulisAI Pipeline Architecture

3.2.2 Data Preprocessing

Data preprocessing involved cleaning, imputing, and integrating the raw records into a unified dataset restricted to eight high-priority "focus crimes" specified by the police personnels on Table 1. Raw columns with excessive missing values or irrelevant identifiers were also excluded from the study.

Table 1. Focus Crimes

Focus Crime	Definition
Murder	Includes murder, and parricide
Homicide	All crimes involving homicide
Rape	Includes violation on Anti-Rape Law of 1997 and Republic Act 11648 or rape cases
Physical Injuries	Includes Serious Physical Injuries, Less Serious Physical Injuries, and Slight Physical Injuries
Robbery	Including Robbery with Violence against or intimidation of person (e.g. Robbery with Homicide)
Theft	Including Qualified Theft
Carnapping MC	Incidents of stolen motorcycles
Carnapping MV	Incidents of stolen vehicles

3.2.3 Data Transformation

To avoid data sparsity, the 24-hour day was consolidated into four 6-hour segments (Time_of_Day): Midnight, Morning, Afternoon, and Evening. This prevented the dataset from having unique combinations that may affect the analysis of crime counts per time block. The data was scoped to incidents from 2017–2024 (3,593 incidents) for model training, leaving the 2025 data (85 incidents) as a final, hold-out test set

To create the final dataset, the training data was aggregated by unique spatio-temporal keys—specifically Barangay, Month, Weekday, and Time_of_Day—resulting in 2,571 distinct spatio-temporal blocks. To establish the classification target variable (Alarm_Level), the Crime_Count within each block was categorized into 'Low', 'Medium', and 'High' levels based on the 25th (Q1) and 75th (Q3) percentiles. This thresholding approach was adapted from the methodology of Yadhunath et al. [17] to properly calibrate the risk levels. Following aggregation, the predictive features were reduced to seven key inputs: Population, Area_Sqkm, Avg_Num_Stations_1km, Weekend_Ratio, Avg_Hour, Avg_Victims, and Avg_Suspects.

Table 2. Preprocessed Dataset

Features	Definition
Alarm_Level	The final target variable predicting crime risk
Population	The total population of the barangay
Area_sqkm	The land area of the barangay in square kilometers
Avg_Num_Stations_1km	The number of police stations within a 1km radius of the aggregated block's geographic coordinates.
Weekend_Ratio	The proportion of crimes in the spatio-temporal block that occurred on a weekend
Avg_Hour	The average time of hour (from 0 to 23) of crime occurrences within the specific block.
Avg_Victims	A historical crime attribute about the number of victims within the specific block.
Avg_Suspects	A historical crime attribute to the number of suspects within the specific block.

3.2.4 Model training and evaluation

To evaluate the models, the final aggregated dataset of 2,571 blocks was partitioned into a 75% training set (1,928 samples) and a 25% testing set (643 samples), utilizing a stratified split based on the methods of Yadhunath et al. [17] to preserve class proportions. Recognizing the dataset was heavily skewed toward 'Low' alarm instances, a comparative training approach was established. Models were trained on the original imbalanced data and compared against models trained on a balanced dataset synthetically generated using the Synthetic Minority Over-sampling Technique (SMOTE) (4,251 samples).

Four distinct supervised machine learning classifiers were developed and subjected to hyperparameter tuning via grid and random search with 5-fold cross-validation. Multinomial Logistic Regression (MLR) served as the interpretable statistical baseline. Support Vector Classifier (SVC) was utilized to identify optimal decision boundaries in high-dimensional space. Furthermore, two advanced tree-based ensemble methods—Random Forest (RF) and Extreme Gradient Boosting (XGBoost)—were selected for their robust ability to capture complex, non-linear spatio-temporal interactions.

To ensure optimal model performance and reproducibility, extensive hyperparameter tuning was conducted via grid and random search alongside 5-fold cross-validation. The algorithms were tuned independently for both the original imbalanced training set (1,928 samples) and the SMOTE-oversampled training set (4,251 samples). The resulting optimal configurations utilized for the comparative evaluation are detailed in Table 3.

Table 3. Optimized hyperparameter settings for four models

Model	Non-SMOTE	With SMOTE
XGB	n_estimators: 400, max_depth: 5, learning_rate: 0.05, subsample: 0.9, colsample_bytree: 0.7, gamma: 0, reg_alpha: 0.01, reg_lambda: 1.5	n_estimators: 500, max_depth: 6, learning_rate: 0.05, subsample: 0.8, colsample_bytree: 0.7, gamma: 0.1, reg_alpha: 0, reg_lambda: 1
RF	n_estimators: 500, max_depth: 12, min_samples_split: 10, min_samples_leaf: 4, max_features: 'sqrt'	n_estimators: 300, max_depth: 15, min_samples_split: 3, min_samples_leaf: 4, max_features: 'sqrt'
SVC	C: 50, kernel: 'rbf', gamma: 'scale'	C: 150, kernel: 'rbf', gamma: 'scale'
MLR	C: 0.1, penalty: 'l2', solver: 'lbfgs'	C: 0.01, penalty: 'l2', solver: 'lbfgs'

To avoid the "Accuracy Paradox" inherent in imbalanced datasets, model efficacy was evaluated using Macro-average F1-Score, Precision, Recall, and Unweighted Average Recall (UAR). Additionally, a "Dangerous Misclassification Rate" was calculated to penalize models that erroneously classified 'High' crime alarm regions as 'Low'.

3.2.5 Model Deployment into Web-based Application

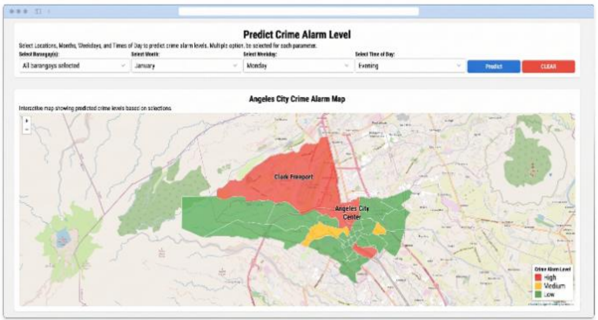

Figure 3: PulisAI Web-app Interface

To transform the theoretical model into an actionable decision support tool for the Angeles City Police Office (ACPO), PulisAI was deployed utilizing a monolithic server-side rendering architecture. The backend, powered by the Flask web framework, manages HTTP routing and data manipulation via Pandas and NumPy, while Joblib loads the serialized XGBoost champion model for high-speed inference. The client-side interface, developed with HTML5, CSS3, JavaScript, and Jinja2 templates, provides three core operational modules: a predictive crime hotspot interface, a comprehensive visualizations dashboard, and an administrative data upload module allowing authorized personnel to input new CSV records for automated model retraining. To facilitate strategic spatial analysis, the system integrates Leaflet.js to render interactive geographic boundary maps. The GeoJSON boundary data for the 33 barangays of Angeles City was sourced from an open-source Philippine geographic repository [11] These maps are dynamically color-coded green, yellow, and red to reflect predicted alarm levels, functioning alongside Plotly.js for detailed historical trend charting.

3.3 Evaluation Metrics

The PulisAI dataset is inherently imbalanced, with 'Low' alarm levels far outnumbering 'Medium' and 'High' levels. This creates an Accuracy Paradox, where a model can achieve high accuracy by simply defaulting to the 'Low' class, rendering it useless for proactive policing. To overcome this, this study adopts the robust evaluation metrics used by Yadhunath et al. [17], which are designed to assess performance on minority classes. The models were evaluated using the following criteria:

Accuracy - the overall percentage of all correct predictions. However, this is not a good indicator of a model's effectiveness due to class imbalance.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Precision - the reliability of the model's positive predictions. It answers the question: "Of all areas the model predicted as 'High' alarm, what percentage were actually 'High' alarms?" High precision minimizes False Positives (FP), ensuring that ACPO resources are not wasted on false alarms.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall (or Sensitivity) - measurement of the model's ability to identify all relevant cases. It answers: "Of all actual 'High' alarm areas, what percentage did the model correctly identify?" High recall minimizes False Negatives (FN), which is critical for ensuring that dangerous hotspots are not missed.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score (Macroaveraged) - the harmonic mean of Precision and Recall, providing a single balanced score for each class. This study uses the macro-average—the unweighted mean of each class's F1-Score. This is crucial as it gives equal importance to the model's performance on the rare 'High' and 'Medium' classes as it does to the common 'Low' class.

$$F1 = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (4)$$

Unweighted Average Recall (UAR) - the simple average of the individual recall scores for each of the three classes. Like the macro-F1, this metric ensures the model is effective at identifying samples from all classes, not just the majority 'Low' class.

$$UAR = \frac{Recall_{Low} + Recall_{Medium} + Recall_{High}}{3} \quad (5)$$

Dangerous Misclassification Rate - Other than standard metrics, the percentage of 'High' alarm regions that are misclassified as 'Low' is also calculated. This is considered the most dangerous failure, as it would deploy minimal resources to a high-risk area. The primary goal of the model is to minimize this to as close to zero as possible.

4. RESULTS AND DISCUSSION

4.1 Correlation Analysis

An analysis of the aggregated 2017-2024 dataset was conducted to identify significant factors correlated with crime incidence. The results revealed that among the selected features, only Population exhibited a moderate positive linear relationship with the target variable, Crime_Count (r = 0.32).

In contrast, all other spatio-temporal and spatial factors—including Area_sqkm (r = 0.09), Avg_Num_Stations_1km (r = 0.04), Weekend_Ratio (r = 0.04), and Avg_Hour (r = 0.04)—yielded correlation coefficients near zero. This indicates a lack of strong linear association between most contextual features and crime volume. This finding is critical; it demonstrates that simple linear models derive minimal predictive power from these inputs, effectively justifying the deployment of complex, tree-based classifiers such as XGBoost and Random Forest to capture the underlying non-linear interactions within the data.

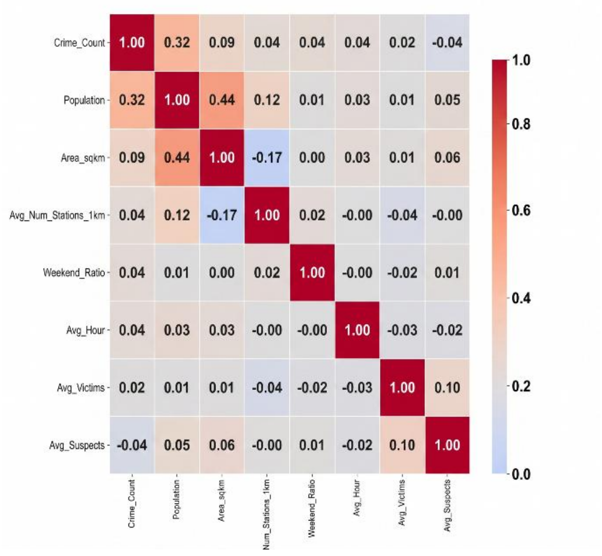

Figure 4. Correlational analysis

4.2 Comparative Model Performance

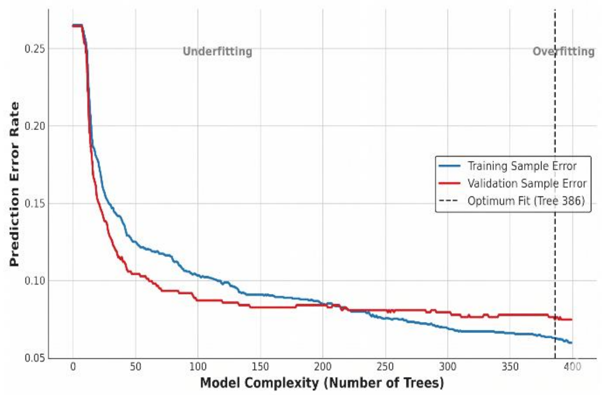

Figure 5. XGBoost Learning Curve

To validate the stability and generalization of the champion model, a learning curve analysis was performed (see Figure 5). The plot illustrates the training and validation error rates as a function of model complexity (number of trees). The model demonstrates a consistent convergence, with the optimum fit achieved at 386 trees. Notably, the narrow gap between the training and validation error rates at the optimum point confirms that the model effectively captures the underlying patterns in Angeles City crime data without suffering from significant overfitting or underfitting. This stability justifies the selection of the non-SMOTE XGBoost configuration for deployment.

Table 4. Comparison of models' performance

Model	Training	Accuracy (%)	Macro F1 score (%)	UAR (%)
XGB	Non-SMOTE	92.69	87.00	82.00
	SMOTE	92.07	85.00	82.00
RF	Non-SMOTE	88.65	79.00	73.00
	SMOTE	86.47	77.00	76.00
SVM	Non-SMOTE	79.47	59.00	55.00
	SMOTE	75.58	58.00	59.00
MLR	Non-SMOTE	52.57	42.00	52.00
	SMOTE	51.01	41.00	51.00

A comparative analysis evaluated the four algorithms under both imbalanced (Non-SMOTE) and oversampled (SMOTE) training conditions. As detailed in Table 4, the evaluation revealed that models trained on the original imbalanced data consistently outperformed their SMOTE-balanced counterparts. The iterations trained with SMOTE exhibited overfitting and poor generalization; notably, applying SMOTE to the SVM and XGBoost models resulted in a significant drop in test accuracy.

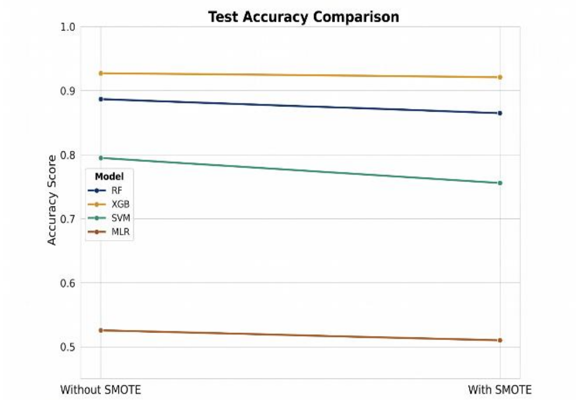

Figure 6. Test Accuracy Comparison across Model Architectures

As shown in Figure 6, the XGBoost Classifier consistently maintains the highest accuracy regardless of the sampling technique used . The visible downward slope for the SVM and Random Forest models when moving from "Without SMOTE" to "With SMOTE" highlights the performance degradation caused by artificial oversampling in this specific urban crime context.

The XGBoost classifier trained on the original imbalanced dataset emerged as the superior model. It achieved the highest overall Test Accuracy (92.69%) and Macro-averaged F1-Score (87%). Crucially, XGBoost yielded the highest Unweighted Average Recall (82%), demonstrating its superior capability to successfully identify the minority 'Medium' and 'High' alarm classes without relying on artificial oversampling. Random Forest provided the second-best performance (88.65% accuracy), while the linear SVM and MLR models failed to adequately capture the data's complexity. Consequently, the non-SMOTE XGBoost model was selected as the engine for the PulisAI application.

4.2.1 Detailed Empirical Analysis

To provide a deeper evaluation of the champion model beyond macro-averages, a class-wise classification report was created for the Non-SMOTE XGBoost model. This detailed breakdown ensures that the high overall accuracy is not merely a result of the majority 'Low' alarm class but reflects robust performance across the 'Medium' and 'High' minority classes.

Table 5. Class-Wise Classification Report for XGBoost (Non-SMOTE)

Alarm Level	Precision	Recall	F1-Score
Low	0.92	1.00	0.96
Medium	0.96	0.76	0.85
High	0.93	0.76	0.84

4.2.2 Analysis of Results

The detailed analysis in Table 5 confirms that the XGBoost model achieves high precision for high-risk hotspots (0.93). This is critical for law enforcement, as it ensures that 93% of the areas identified as "High Alarm" were correctly classified, minimizing the waste of limited police resources on false positives. While the recall for 'Medium' and 'High' classes (0.76) is lower than that of the majority class, this is expected given the inherent imbalance in urban crime data. However, the model significantly outperforms the MLR baseline (UAR: 52%) by capturing non-linear interactions between demographic factors like population and temporal variables like the average hour of occurrence. Notably, the model minimized the most critical failure point: it erroneously predicted only 18.52% (10 out of 54) of actual 'High' alarm test samples as 'Low' risk.

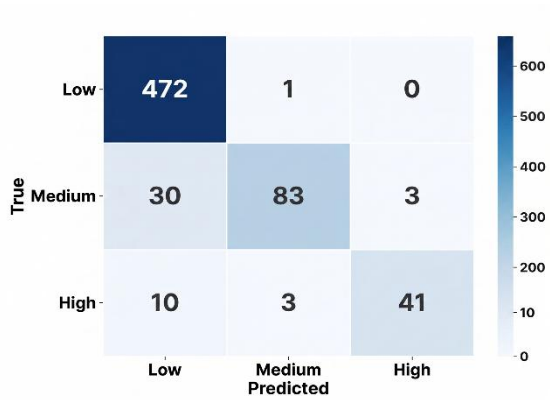

Figure 7. Confusion Matrix of XGBoost Classifier with Class Imbalanced Data

Furthermore, the XGBoost classifier minimized the dangerous misclassification rate, erroneously predicting only 18.52% (10 out of 54) of actual 'High' alarm test samples as 'Low' risk, thereby ensuring that high-risk hotspots are rarely overlooked during patrol planning.

4.2.3 ROC-AUC Performance

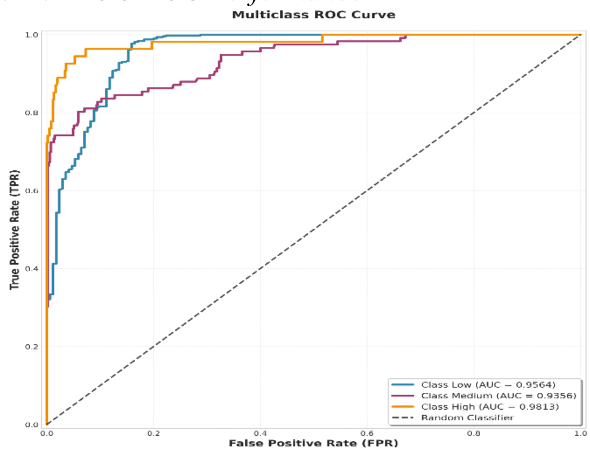

Figure 8. Multiclass Receiver Operating Characteristic (ROC) Curve for PulisAI

The discrimination capability of the XGBoost model across the three alarm levels is further evidenced by the Multiclass ROC Curve (Figure 8). The Area Under the Curve (AUC) scores for all classes are exceptionally high: 0.9564 for 'Low', 0.9356 for 'Medium', and 0.9813 for 'High'. These results indicate that the model possesses a high degree of separability, accurately distinguishing between different risk levels regardless of the classification threshold. The particularly high AUC for the 'High' alarm class (0.9813) is vital for law enforcement, as it ensures reliable identification of the most critical crime hotspots.

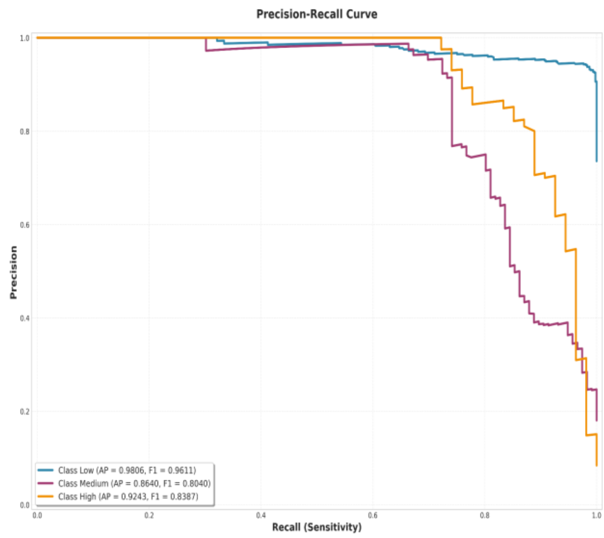

Figure 9. Precision-Recall Curve for Imbalanced Crime Alarm Classification.

Given the inherent class imbalance in the Angeles City crime dataset, where 'Low' alarm instances predominate, the Precision-Recall Curve (Figure 9) provides a more rigorous evaluation of the model's performance on minority classes. The Average Precision (AP) for the 'High' alarm class is 0.9243, with an F1-score of 0.8387. This demonstrates that even without synthetic oversampling (SMOTE), the XGBoost model maintains high precision and recall for the most dangerous hotspots. This balance is crucial for the ACPO, as it minimizes both wasted resources from false positives and the danger of missed detections (false negatives).

4.3 End-user Evaluation Results

To validate the operational viability of the deployed PulisAI web application, a system evaluation was conducted based on the ISO 25010 software quality model. The application was tested by authorized personnel from the Angeles City Police Office (ACPO).

The respondents assessed the system using a 5-point Likert scale across five core software quality characteristics: Functional Suitability, Performance Efficiency, Usability, Reliability, and Compatibility. The quantitative results of the evaluation are summarized in Table 6.

Table 6. ISO 25010 System Evaluation Results

Evaluation Criteria	Average Score
Features and Accuracy (Functional Suitability)	4.66
Speed and Performance (Performance Efficiency)	5.00
Ease of Use (Usability)	5.00
Reliability and Stability (Reliability)	5.00
Working with Other Software (Compatibility)	5.00

As presented in the data, the PulisAI application achieved an overall mean score of 4.93, indicating acceptance by the end-users who are law enforcers. The system achieved perfect scores (5.00) in Performance Efficiency, Usability, Reliability, and Compatibility. This demonstrates that the backend integration of the XGBoost model provided predictions without latency, the interface was intuitive and stable for non-technical officers. Furthermore, the strong rating in Functional Suitability (4.66) confirms that the predictive features and interactive spatial maps directly align with the ACPO's requirements for strategic patrol planning. Ultimately, these results validate that PulisAI successfully transitions the theoretical machine learning model into a highly effective and deployable intelligence tool.

5. CONCLUSION

This study addressed the operational constraints of the Angeles City Police Office (ACPO) by developing "PulisAI," a predictive policing web application. By transforming crime records into spatio-temporal blocks and utilizing a three-tier classification framework, the research demonstrated that local crime patterns correlate with demographic and temporal factors rather than occurring randomly. This establishes a data-driven pathway to transition law enforcement from reactive responses to proactive patrol planning.

A comparative analysis of machine learning algorithms revealed that the XGBoost classifier, trained on the original imbalanced dataset, yielded the highest performance. Achieving test accuracy of 92.69% and a macro F1-score of 87%, the model captured non-linear relationships. The algorithm's internal weighting mechanisms proved more effective at handling class imbalance than artificial oversampling techniques like SMOTE.

However, this study acknowledges limitations stemming from data constraints. First, the limited volume of historical crime records lacked the variance necessary to fully reflect real-world crime occurrences. Consequently, when SMOTE was applied, the models exhibited overfitting and poor generalization. While the non-SMOTE XGBoost model effectively minimized the dangerous misclassification rate, its overall predictive performance remains constrained by the low number of actual 'High' alarm examples. A larger, more comprehensive dataset would inherently improve the base model's sensitivity and reliability across all classes. Furthermore, the system lacks integration with real-time external environmental features, such as weather conditions, traffic patterns, and local events, which often influence crime dynamics. Lastly, the model relies on static demographic features; the population statistics sourced from the Philippine Statistics Authority [14] are updated only every five years. Securing localized, annual demographic records for all barangays would provide the temporal flexibility required to capture dynamic urban changes and support more complex predictive research.

To maximize the system's utility, future research should expand the dataset to include additional geographic areas and socio-economic variables, such as business density. Incorporating real-time data pipelines and external features, like weather or traffic patterns, will transition the tool from a strategic to an operational asset. Finally, ethical auditing and scheduled model retraining are recommended to mitigate historical data bias and adapt to evolving urban crime trends.

6. CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest.

7. INFORMED CONSENT

Informed consent was obtained from the personnel included in the study.

8. ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the faculty of the College of Computer Studies at Angeles University Foundation, for the academic support and resources provided during this study. Special thanks are extended to the Angeles City Police Office for granting access to the anonymized crime incident records and to the authorized personnel who participated in the system evaluation. Their cooperation was essential in validating the operational viability of the PulisAI application.

9. REFERENCES

[1] Abubakar, H., Boukari, S., Gital, A. Y., & Zambuk, F. U. (2024). A hyper-parameter tuned Random Forest algorithm-based on Artificial Bee Colony for improving accuracy, precision and interpretability of crime prediction. Dutse Journal of Pure and Applied Sciences, 10(4a), Article 4a. https://doi.org/10.4314/dujopas.v10i4a.34

[2] Annang, A., Abrokwah, R. K., & Asante, E. O. (2023). Challenges of Effective Policing in Ghana: The Central Regional Command in Perspective. Open Access Library Journal, 10(8), 1–18. https://doi.org/10.4236/oalib.1110506

[3] Asor, J. R. (2020). Implementation of Predictive Crime Analytics in Municipal Crime Management System in Calauan, Laguna, Philippines. International Journal of Advanced Trends in Computer Science and Engineering, 9(1.3), 150–157. https://doi.org/10.30534/ijatcse/2020/2291.32020

[4] Ayeng, R. (2024, September 15). Manila tops cities with highest crime rate in SE Asia-Global poll. Daily Tribune. https://tribune.net.ph/

[5] Balahadia, F. F., Asor, J. R., Catedrilla, G. M. B., Villarica, M., & Cabiente, J. M. (2020). Intelligent Investigation on

Crime Incident Reports in the Province of Laguna through Predictive Model Development. International Journal of Advanced Trends in Computer Science and Engineering, 9(1.3), 139–144. https://doi.org/10.30534/ijatcse/2020/2091.32020

[6] Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A., & Sherazi, H. H. R. (2020). Spatio-Temporal Crime HotSpot Detection and Prediction: A Systematic Literature Review. IEEE Access, 8, 166553–166574. https://doi.org/10.1109/ACCESS.2020.3022808

[7] Casilao, J. (2023, January 9). PNP: Theft, rape, physical injury most prevalent crimes in last 6 months. GMA News Online. https://www.gmanetwork.com/

[8] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

[9] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

[10] Department of the Interior and Local Government. (n.d.). Angeles City Profile, accessed 28 October 2025, https://region3.dilg.gov.ph/

[11] Faeldon, E. (2019). philippines-json-maps [GeoJSON dataset]. GitHub. https://github.com/

[12] Gom-gom-o, J. (2024). CHALLENGES ENCOUNTERED BY THE POLICE OFFICERS IN CRIME PREVENTION IN TABUK CITY, KALINGA. EPRA International Journal of Multidisciplinary Research (IJMR), 10(5), Article 5.

[13] Jenga, K., Catal, C., & Kar, G. (2023). Machine learning in crime prediction. Journal of Ambient Intelligence and Humanized Computing, 14, 2887-2913. https://doi.org/10.1007/s12652-023-04530-y

[14] Philippine Statistics Authority. (2024). 2024 Census of Population (POPCEN): Population Counts Declared Official by the President, accessed 28 October 2025, https://psa.gov.ph/

[15] Safat, W., Asghar, S., & Gilani, S. (2021). Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques. IEEE Access, 1-1. https://doi.org/10.1109/ACCESS.2021.3078117

[16] Wubineh, B. Z. (2024). Crime analysis and prediction using machine-learning approach in the case of Hossana Police Commission. Security Journal, 37(4), 1269–1284. https://doi.org/10.1057/s41284-024-00416-6

[17] Yadhunath, R., Srikanth, S., Sudheer, A., & Palaniswamy, S. (2019). Identification of Criminal Activity Hotspots using Machine Learning to Aid in Effective Utilization of Police Patrolling in Cities with High Crime Rates. 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), 1–6. https://doi.org/10.1109/CSITSS47250.2019.9031057

[18] Yao, S., Wei, M., Yan, L., Wang, C., Dong, X., Liu, F., & Xiong, Y. (2020). Prediction of crime hotspots based on spatial factors of Random Forest. 2020 15th International Conference on Computer Science & Education (ICCSE), 811-815. https://doi.org/10.1109/ICCSE49874.2020.9201899

[19] Yunus, A., & Loo, J. (2024). London street crime analysis and prediction using crowdsourced dataset. Journal of Computational Mathematics and Data Science, 10, 100089. https://doi.org/10.1016/j.jcmds.2023.10008