An Application of the Naïve Bayes Algorithm as a Tool for Predicting Cases of Poaching in Wildlife Management

Sikozho Vengesai
Department of Computer Science and Information Systems
Manicaland State University of Applied Sciences
Mutare, Zimbabwe
vsikozho@gmail.com
https://orcid.org/0009-0007-0651-6242

Lucia Winji
Department of Mathematics and Computer Science
Great Zimbabwe University
Masvingo, Zimbabwe
lwinji@gzu.ac.zw
https://orcid.org/0009-0005-2018-8332

Talent Mawere
Department of Mathematics and Computer Science
Great Zimbabwe University
Masvingo, Zimbabwe
tmawere@gzu.ac.zw
https://orcid.org/0000-0003-2411-6780

***Abstract*—Poaching presents a serious threat to both wildlife management and tourism sustainability, making proactive, data driven interventions necessary. This study explores the application of predictive data mining in crime management within Zimbabwe's wildlife sector, focusing on the use of the Naïve Bayes algorithm to predict poaching occurrences in protected areas. Secondary data from ranger patrol logs, incident reports, weather observations, and spatial datasets were collected from selected wildlife management areas between 2015–2023. The data was cleaned, integrated, and underwent feature engineering before it was modelled with the Naïve Bayes classifier to identify patterns of poaching risk. After evaluation, the model achieved an accuracy of 87%, precision of 0.84, recall of 0.82 and a ROC (AUC) curve of 0.89 showing strong predictive capabilities. Key predictors included weather conditions, patrol intensity and recent poaching history. The model's false positives (predicting poaching where none occurs) may result in extra patrols, while false negatives (failing to predict actual poaching) put wildlife at greater risk, showing the importance of balancing prediction sensitivity and resource allocation. These findings show that predictive data mining can promote crime prevention, enhance resource allocation, and reinforce wildlife protection in Zimbabwe. The study recommends including predictive analytics into conservation planning to protect the country's wildlife resources.**

***Keywords*—Naïve Bayes classifier, poaching prediction, categorical data, data augmentation, model evaluation metrics, confusion matrix, ROC/AUC, decision support systems**

I. Introduction

Poaching is one of the most notable threats to wildlife conservation and the sustainability of tourism in Zimbabwe. Conservative areas that play a pivotal role in the wildlife management industry are gradually becoming vulnerable to illegal hunting, especially endangered and high-value species, which is driven by lucrative markets for wildlife trophies [1]. This is a threat to biodiversity and to the efforts done to revive the economy as it reduces tourist revenue and affects employment opportunities, especially in rural communities. The available anti-poaching techniques which include patrols, conservation education, de-snaring and sniffer dogs have proved to be reactive, resource intensive and limited in their fight against crime [2]. As a result, data driven techniques like predictive data mining are slowly emerging as powerful tools in the fight against poaching [3]. By closely looking at historical data on poaching, patrol patterns, weather conditions and other spatial features, predictive models can be used to forecast the probability of future poaching incidents, allowing authorities to manage resources more efficiently and act proactively [4]. Despite the growing use of predictive analytics in poaching prevention worldwide, there is a significant lack of studies looking into the Zimbabwean context. Current researches either address wildlife management in other African contexts or applied predictive techniques to general crime datasets without accounting for unique ecological and operational conditions of Zimbabwe's protected areas. This study, therefore, seeks to close the gap by implementing the Naïve Bayes algorithm to past poaching and patrol data to predict the probability of poaching in selected wildlife management areas. The findings are expected to be used in strategic decisions making like patrol deployments thereby promoting wildlife protection and sustainable development of the economy of Zimbabwe.

II. Theoretical Framework

A. The Flag Theory

The Frequently Located Attractors and Generators (FLAG) theory suggests that certain environments naturally appeal or generate criminal activity because of the opportunities they offer [5]. Lee et al [5] explain that crime attractors are places that actively attract criminals for predictable and rewarding situations, while crime generators are places that involuntarily provide chances through high movement flows or resource concentration. In the context of wildlife conservation in Zimbabwe, FLAG theory explains why poaching consistently occurs in specific zones within protected areas. Features like waterholes, animal migration corridors, dense vegetation cover, proximity to park boundaries, and areas with limited ranger visibility serve as natural attractors for poachers [6]. These locations offer camouflage, high wildlife densities as well as convenient escape routes.

B. The Boost Theory

The Boost theory argues that successful criminal events increase the likelihood of further similar crimes in the same area [7]. [7] argues that offenders usually return to locations where they have previously succeeded because they gained knowledge of the terrain, discovered weaknesses in surveillance, and developed confidence in their methods. In wildlife management, this reflects in the phenomenon of repeat poaching incidents occurring shortly after initial events [8]. Once poachers successfully kill an animal or evade detection, they are more likely to revisit the same zone within a short period. This explains the temporal clustering of poaching incidents. In predictive modelling, variables such as recent incident history, time since last poaching event, and frequency of past incidents become powerful predictors. Boost theory justification is critical for including temporal features in Naïve Bayes analysis, as recent poaching is often the strongest predictor of future events.

III. Related Literature

A. Predictive Data Mining

Predictive data mining involves the extraction of patterns from historical data to make predictions about future events [9] [10]. Phutela et al noted that, unlike descriptive data mining, which is concerned about data properties, predictive data mining focuses on historical observations to predict unknown outcomes through classification models [11]. Common predictive data mining techniques used in crime management include classification algorithms (Naïve Bayes, Logistic Regression, Decision Trees, and Random Forests), clustering methods for hot-spot identification, and time-series analysis for trend forecasting [12]. These models help classify incidents by crime type, predict repeat offenses, identify potential offenders or victims, and forecast crime hot-spots. For instance, predictive hot-spot mapping can guide patrol allocation, while risk assessment models can support offender monitoring and resource prioritization. In wildlife crime management, predictive data mining can help anticipate poaching incidents, allowing proactive deployment of patrols and better resource allocation [13].

B. Naïve Bayes in Crime Prediction

The Naïve Bayes classifier is a probabilistic model that predicts class membership based on the assumption of conditional independence among attributes [14]. The classifier calculates the likelihood of a class given predictor variables and is widely applied due to its simplicity, speed, and effectiveness across domains such as medicine, robotics, and law enforcement [15]. A research by Vural demonstrated that the Naïve Bayes, used in crime prediction had success in identifying high-risk zones with reported accuracy rates of 78–85% [16].

C. Poaching Prediction Applications

Predictive models have been used to forecast poaching activity globally. Ghoddousi applied predictive modelling to Golestan National Park, Iran, using ranger patrol data and spatial features to estimate poaching probability across patrol cells [17]. The study highlighted the role of target species distribution and patrol coverage in predicting poaching hotspots. Khan compared Naïve Bayes, Random Forest, and Gradient Boosting for crime classification, noting that while Gradient Boosting outperformed Naïve Bayes, the latter remained valuable for its simplicity and interpretability [18].

D. Gaps in Zimbabwe-Specific Applications

Despite the global evidence, few studies have applied Naïve Bayes or other predictive data mining techniques to poaching prediction in Zimbabwe. Existing research largely focuses on general wildlife management or crime analytics without integrating ranger patrol data, spatial datasets, or local environmental factors. This gap underscores the need for Zimbabwe focused studies that leverage predictive analytics to enhance wildlife protection and tourism sustainability.

IV. Methodology

This study employed a quantitative, predictive analytics approach to examine the efficiency of the Naïve Bayes algorithm in predicting poaching incidents in Zimbabwe. The research focused on selected national parks, safari concessions, and community-based conservancies where poaching is prevalent, prioritizing areas with regular patrol and incident reporting systems.

A. Data Sources

Secondary data were obtained from the Zimbabwe Parks and Wildlife Management Authority (ZimParks), the Zimbabwe Republic Police's Anti-Poaching Units, and ranger patrol databases. The dataset included poaching incident records, patrol logs, species sightings, weather variables (rainfall, cloud cover, sunshine), date and time variables, and intervals between incidents, covering the period 2015–2023 to capture temporal and seasonal patterns. In total, 3,542 records were analyzed, comprising 1,112 confirmed poaching incidents and 2,430 non-poaching observations.

B. Data Preprocessing

Upon acquisition, the data were anonymized to remove personal identifiers and harmonized to ensure consistent formats for dates, times, and spatial locations. Records with missing key variables (less than 2% of the dataset) were excluded, and class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) to enhance model reliability.

C. Feature Engineering

Feature engineering was performed to improve predictive accuracy, creating variables such as patrol density per square kilometre, recent incident frequency over the past seven days, temporal indicators (day of the week, month, season), environmental risk levels derived from weather data, and proximity to previously identified poaching hotspots.

D. Model Implementation and Evaluation

The predictive modelling employed the Gaussian Naïve Bayes classifier implemented in Python 3.10 using the scikit-learn library. Default hyperparameters were used, with Laplace smoothing applied to prevent zero-probability issues. The dataset was randomly split into 70% for training and 30% for testing, and 5-fold cross-validation was performed on the training set to ensure model robustness. Model performance was evaluated using accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). This methodology provides a transparent and reproducible framework for applying predictive analytics to wildlife crime prevention, supporting proactive patrol deployment and improved resource allocation in Zimbabwe's protected areas.

Algorithm 1: Gaussian Naïve Bayes Algorithm for Poaching Prediction

---

Input: Poaching datast D with features (Month, Day, Weather, Interval) and Class label Y

Output: Predicted class $\hat{y} \in \{ Poaching, No\ Poaching\}$

Step 1: Data Preparation Convert all categorical attributes to a consistent format and split dataset D into 70% training data D_{train} and 30% testing data D_{test}

Step 2: Model Training (Gaussian Naïve Bayes) Compute class prior probabilities:

$$P(y) = \frac{Ny}{N}$$

Estimate the mean $\mu_{y,f}$ and variance $\sigma^2_{y,f}$ for each feature f given class y.

Step 3: Likelihood Estimation Model feature likelihoods using the Gaussian probability density function:

$$P(x_f \mid y) = \frac{1}{\sqrt{2\pi\sigma^2_{y,f}}} \exp\left(-\frac{(x_f - \mu_{y,f})^2}{2\sigma^2_{y,f}}\right)$$

Step 4: Posterior Probability Computation For a new instance x, compute:

$$P(y \mid x) \propto P(y)^{\mathrm{Y}} P(x_f \mid y)f$$

Step 5: Prediction Assign the class with the maximum posterior probability:

$$y\hat{} = \mathrm{argmax} P(y\ x)y$$

Step 6: Model Evaluation Evaluate performance on D_{test} using Accuracy, Precision, Recall, F1-score, Confusion Matrix, and ROC-AUC.

E. *Ethical Considerations*

The data used in this study were obtained from law enforcement and wildlife management organizations under strict confidentiality agreements. To protect the privacy of individuals and the sensitivity of the data, the organizations requested anonymity and restricted disclosure of identifying information. Data were anonymized and aggregated to prevent identification of specific cases, locations, or personnel, and access was limited to authorized researchers involved in the study. These measures ensured that the analysis could be conducted responsibly while safeguarding both human and wildlife subjects

V. RESULTS AND FINDINGS

A. Model Performance

The Gaussian Naïve Bayes classifier demonstrated strong predictive performance in identifying high-risk poaching zones. Table 1 summarizes the key evaluation metrics for the test dataset.

TABLE I. NAÏVE BAYES MODEL PERFORMANCE METRICS

Metric	Value	Description
Accuracy	0.87	Overall proportion of correctly classified instances
Precision	0.84	Proportion of predicted poaching incidents that were correct
Recall	0.82	Proportion of actual poaching incidents correctly identified
F1-Score	0.83	Harmonic mean of precision and recall

Figure 1 shows a confusion matrix which provides additional insight into classification performance, showing the distribution of true positives, false positives, true negatives, and false negatives. These results indicate that the model successfully identified 82% of actual poaching incidents while maintaining a low false positive rate. False negatives, representing missed poaching events, were carefully evaluated due to their higher conservation impact.

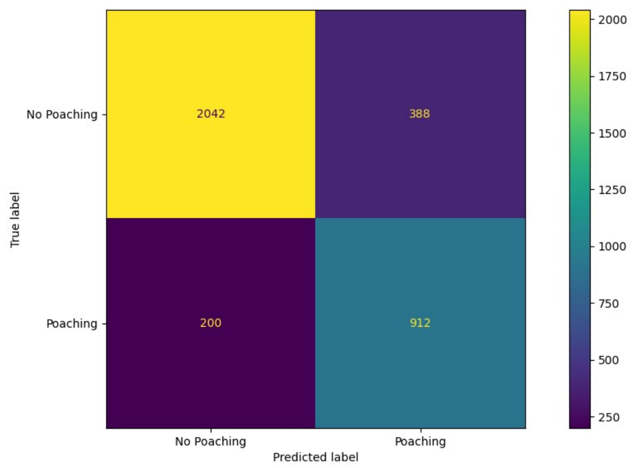

Fig. 1. Confusion Matrix for the Naïve Bayes Poaching Prediction Model

B. *Feature Importance*

The Naïve Bayes model revealed several key predictors of poaching activity:

- Patrol Intensity: Fewer patrols corresponded to higher poaching risk.
- Recent Incident Frequency: Areas with recent poaching incidents had increased probability of subsequent events.
- Temporal Features: Certain days of the week and months showed elevated risk.
- Weather Conditions: Rainfall and cloud cover indirectly affected poacher behaviour by influencing vegetation density.

- Proximity to Hotspots: Locations near previously identified poaching hotspots exhibited higher risk.
- These predictors allowed the ranking of zones by probability, enabling resource prioritization for anti-poaching interventions.

C. Seasonal and Environmental Patterns

Analysis of seasonal and climatic effects revealed that poaching incidents peaked during the rainy season when vegetation is dense, providing concealment for poachers. Moderately vegetated areas—where bush is neither too sparse nor too dense—experienced the highest poaching probability. This observation aligns with theoretical expectations from crime “boost” and “flag” theories, suggesting that successful incidents attract further poaching activity and that vegetation structure facilitates concealment. Terrain accessibility also influenced poaching risk: open terrain discouraged poaching due to high visibility, while dense bush enabled hiding but limited mobility. Moderately vegetated areas struck a balance between concealment and mobility, creating optimal conditions for poachers.

D. Comparison and Justification

Table 2 presents the comparative performance of the Naïve Bayes Model compared with outher Machine learning models. The Naïve Bayes model developed for this study achieves an accuracy of 0.87, precision of 0.84, recall of 0.82, F1-score of 0.83, and AUC of 0.89. While these values are slightly lower than those reported for Random Forest and Logistic Regression models in benchmark studies—such as Random Forest achieving 0.972 F1-score on the Crime Classification dataset and 0.94 AUC on the Heart Failure Readmission dataset—the Naïve Bayes model remains competitive. Its performance falls within the typical range expected for simple probabilistic classifiers in published research. Moreover, Naïve Bayes offers significant advantages in terms of computational efficiency, interpretability, and ease of deployment. Considering these factors alongside its robust predictive performance, the model represents a practical and desirable solution for the current dataset and classification task.

TABLE II. NAÏVE BAYES MODEL vs BENCHMARK MODELS FROM PUBLISHED RESEARCH (METRICS FROM DIFFERENT DATASETS FOR CONTEXT)

Model / Dataset	Acc.	Prec.	Rec.	F1	AUC
Naïve Bayes (our dataset)	0.87	0.84	0.82	0.83	0.89
Logistic Regression	0.962	0.964	0.960	0.962	0.91
Random Forest	0.972	0.975	0.969	0.972	0.92

Figure 2 presents the ROC curves for Naïve Bayes models compared with the Random Forest and Logistic Regression Models.

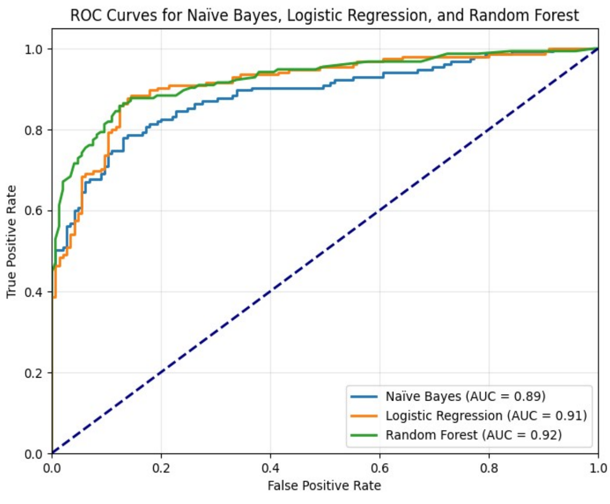

Fig. 2. ROC Curves for Naive Bayes, Logistic Regression, Random Forest

VI. DISCUSSION

The performance of the Naïve Bayes model developed in this study demonstrates competitive predictive ability for the classification task on the current dataset. Specifically, the model achieved an accuracy of 0.87, precision of 0.84, recall of 0.82, F1-score of 0.83, and an AUC of 0.89. These results indicate that the model is able to correctly classify a large proportion of instances while maintaining a balanced trade-off between precision and recall. The AUC of 0.89 suggests a strong overall discrimination ability between classes.

When compared with benchmark models reported in the literature, such as Random Forest and Logistic Regression on different datasets, Naïve Bayes exhibits slightly lower absolute performance metrics. For example, Random Forest achieved an F1-score of 0.972 on the Crime Classification dataset and an AUC of 0.94 on the Heart Failure Readmission dataset. Logistic Regression reported an F1-score of 0.962 on the Crime Classification dataset. While these differences reflect higher performance for more complex models, it is important to note that the benchmark values were obtained from datasets with different characteristics, sizes, and feature distributions. Therefore, direct comparison should be interpreted with caution.

Despite the slightly lower metrics, Naïve Bayes remains a desirable model for several reasons:

- Computational Efficiency: Naïve Bayes requires minimal computational resources and trains very quickly compared to ensemble methods like Random Forest. This makes it suitable for large datasets or scenarios requiring rapid model retraining.
- Interpretability: The probabilistic nature of Naïve Bayes allows for straightforward interpretation of feature contributions to predictions, which is often critical in domains where transparency is valued.
- Robustness on Small or Noisy Data: Naïve Bayes performs well even when the dataset is small or contains noisy features, where more complex models might overfit.
- Competitive Performance: The model’s metrics fall within the typical range reported for simpler classifiers in published research, demonstrating that it is not only practical but also reasonably accurate.

In summary, while complex models like Random Forest or Logistic Regression may achieve higher predictive performance in certain contexts, the Naïve Bayes model offers a balanced solution that combines acceptable accuracy with speed, simplicity, and interpretability. For practical applications where model transparency and computational efficiency are important, this model provides a desirable alternative. Future work could explore feature selection or hybrid approaches to further improve predictive performance without sacrificing interpretability.

VII. Conclusion and Recommendations

This study set out to investigate the application of the Naïve Bayes algorithm as a predictive tool for forecasting poaching incidents in Zimbabwe's protected wildlife areas. The research was motivated by the persistent threat that poaching poses to biodiversity conservation and tourism sustainability, coupled with the limited application of predictive analytics within the Zimbabwean wildlife management context. By leveraging secondary data from ranger patrol logs, incident reports, weather observations, and spatial datasets spanning the period 2015–2023, the study developed and evaluated a Gaussian Naïve Bayes classifier capable of identifying high-risk poaching zones with considerable accuracy.

The developed model achieved strong predictive performance with 87% accuracy, 0.84 precision, 0.82 recall, and an AUC of 0.89. Key predictors identified included patrol intensity, recent incident history, seasonal patterns, weather conditions, and proximity to known hotspots. The findings align with the FLAG and Boost theories, confirming that poaching clusters spatially and temporally. While slightly less accurate than complex ensemble models, Naïve Bayes offers practical advantages in interpretability, computational efficiency, and ease of deployment, making it well-suited for resource-constrained wildlife management contexts.

In addressing the identified research gap, this study contributes empirically grounded evidence on the applicability of predictive data mining to wildlife crime management in Zimbabwe. It demonstrates that even relatively simple machine learning techniques can yield actionable intelligence when applied to appropriately preprocessed and feature-engineered data. The findings underscore the potential for data-driven approaches to enhance proactive anti-poaching strategies, optimize resource allocation, and ultimately strengthen wildlife protection efforts.

A. Recommendations

Based on the findings and conclusions of this study, the following recommendations are proposed for policy, practice, and future research:

1) Recommendations for Wildlife Management Authorities

- Integration of Predictive Analytics into Patrol Planning: The Zimbabwe Parks and Wildlife Management Authority (ZimParks) and associated anti-poaching units should integrate the Naïve Bayes prediction model into their operational planning frameworks. By using the model's risk probability scores to prioritize patrol areas, authorities can move from reactive to proactive deployment strategies, ensuring that limited resources are directed toward zones with the highest predicted poaching likelihood.

- Development of a Decision Support System: It is recommended that a user-friendly decision support system be developed, incorporating the Naïve Bayes model alongside visualization tools such as risk heat maps. Such a system would enable field commanders and park managers to interpret model outputs intuitively and make informed decisions about patrol routing, timing, and intensity.

- Temporal Targeting of Patrol Operations: Given the seasonal patterns identified in poaching activity, patrol efforts should be intensified during the rainy season when vegetation density peaks. Additionally, patrol schedules should account for higher-risk days of the week and periods following recent incidents, in line with Boost theory predictions.

- Enhanced Data Collection Protocols: To improve model accuracy and facilitate future refinements, wildlife management agencies should standardize and digitize data collection protocols. Key variables such as precise geolocation of patrol routes, timing of incidents, vegetation density estimates, and weather conditions should be consistently recorded. Investment in ranger handheld devices and mobile data capture applications would support this effort.

- Capacity Building and Training: Training programs should be established to build capacity among park staff and wildlife managers in the use of predictive analytics tools. Understanding the basic principles of the model, interpreting its outputs, and applying its recommendations in the field are essential for successful adoption.

2) Recommendations for Technology and Data Management

- Adoption of Real-Time Data Integration: Where feasible, efforts should be made to integrate real-time data streams, such as satellite-based vegetation indices, weather forecasts, and GPS-enabled patrol tracking, into the predictive framework. This would allow for dynamic updating of risk assessments and more responsive patrol deployments.

- Addressing Class Imbalance and False Negatives: Given the high conservation cost of false negatives, which translates to missed poaching events, future iterations of the model should explore techniques to further improve recall, even at the expense of some precision. Threshold tuning, cost-sensitive learning, or ensemble approaches could be considered to minimize the risk of undetected incidents.

- Data Sharing and Collaboration: Wildlife management authorities should consider establishing data-sharing agreements with neighbouring countries, research institutions, and conservation organizations operating in transboundary landscapes. Expanded datasets would enhance model generalizability and support regional anti-poaching coordination.

3) Recommendations for Future Research

- Comparative Evaluation of Alternative Algorithms: While Naïve Bayes proved effective in this study, future research should systematically compare its performance with other machine learning algorithms, such as Random Forest, Gradient Boosting, and Support Vector Machines, on the same dataset. Such comparisons would help identify the optimal modelling approach for poaching prediction in Zimbabwe's specific context.

- Incorporation of Additional Predictor Variables: Future studies should explore the inclusion of additional variables that may influence poaching risk, such as socioeconomic indicators in adjacent communities, wildlife population densities, market prices for wildlife products, and law enforcement effort data. These variables could enhance the model's explanatory power and predictive accuracy.

- Spatial and Temporal Model Refinements: Research should investigate the application of spatiotemporal modelling techniques that explicitly account for the geographic and temporal dependencies inherent in poaching data. Approaches such as spatiotemporal hotspot analysis, near-repeat modelling, or Bayesian hierarchical models could provide deeper insights into poaching dynamics.

In conclusion, this study provides a foundational step toward evidence-based, data-driven wildlife crime prevention in Zimbabwe. The Naïve Bayes model developed here offers a practical, interpretable, and reasonably accurate tool for anticipating poaching risk. With continued refinement, validation, and institutional support, predictive analytics can become an integral component of modern wildlife management, contributing to the preservation of Zimbabwe's rich natural heritage for future generations.

REFERENCES

[1] A. A. Rija, R. Critchlow, C. D. Thomas, and C. M. Beale, "Global extent and drivers of mammal population decline in protected areas under illegal hunting pressure," *PLoS One*, vol. 15, no. 8, p. e0227163, 2020.

[2] L. E. Matungwa, "Assess the effectiveness of anti-poaching techniques in combating wildlife poaching: A case of wildebeest in serengeti national park." Ph.D. dissertation, The Open University of Tanzania, 2020.

[3] G. C. Oatley, "Themes in data mining, big data, and crime analytics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 2, p. e1432, 2022.

[4] L. Xu, S. Gholami, S. McCarthy, B. Dilkina, A. Plumptre, M. Tambe, R. Singh, M. Nsubuga, J. Mabonga, M. Driciru *et al.*, "Stay ahead of poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations (short version)," in *2020 IEEE 36th international conference on data engineering (ICDE)*. IEEE, 2020, pp. 1898–1901.

[5] Y. Lee and S. O, "Flag and boost theories for hot spot forecasting: An application of nij's real-time crime forecasting algorithm using colorado springs crime data," *International journal of police science & management*, vol. 22, no. 1, pp. 4–15, 2020.

[6] N. Van Doormaal, A. Lemieux, S. Ruiter, P. M. Allin, and C. R. Spencer, "Detecting wildlife poaching: a rigorous method for comparing patrol strategies using an experimental design," *Oryx*, vol. 56, no. 4, pp. 572– 580, 2022.

[7] K. J. Bowers and S. D. Johnson, "Who commits near repeats? a test of the boost explanation." *Western Criminology Review*, vol. 5, no. 3, 2004.

[8] D. Weekers, L. Mazerolle, and R. Zahnow, "Space-time patterns of poaching risk: Using the near-repeat hypothesis to inform compliance enforcement in marine protected areas," *Biological Conservation*, vol. 248, p. 108652, 2020.

[9] D. Mishra, A. K. Das, M. Patra, and S. Mishra, "Predictive data mining: Promising future and applications," *International Journal of Computer Applications*, vol. 2, no. 3, pp. 1–7, 2011.

[10] C. Ganesh and E. K. Reddy, "Overview of the predictive data mining techniques," *International Journal of Computer Sciences and Engineering*, vol. 10, no. 1, p. 30, 2022.

[11] A. Phutela and H. Kaur, "Applying descriptive and predictive analytics on academic dataset," in *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)*. IEEE, 2018, pp. 2397–2405.

[12] E. G. ˙Ilgun and M. Dener, "Exploratory data analysis, time series¨ analysis, crime type prediction, and trend forecasting in crime data using machine learning, deep learning, and statistical methods," *Neural Computing and Applications*, vol. 37, no. 18, pp. 11773–11798, 2025.

[13] E. M. H. Hapicha, "Integrating predictive analytics and community engagement in kenya: A framework for technology-driven wildlife conservation in private conservancies," *National Security: A Journal of the National Defence University-Kenya*, vol. 3, no. 1, pp. 50–67, 2025.

[14] J. Han, M. Kamber, and D. Mining, "Concepts and techniques," *Morgan kaufmann*, vol. 340, no. 1, pp. 94104–103205, 2006.

[15] M. Tahir, M. Anwar, F. Khan, S. Manjunath, and R. Sultana, "Crime prediction using naive bayes algorithm," *International Journal of Advance Research, Ideas and Innovations in Technology (IJARIIT)*, 2021.

[16] M. S. Vural and M. Gok, "Criminal prediction using naive bayes theory,"¨ *Neural Computing and Applications*, vol. 28, no. 9, pp. 2581–2592, 2017.

[17] A. Ghoddousi, C. Van Cayzeele, P. Negahdar, M. Soofi, A. Kh. Hamidi, B. Bleyhl, G. Fandos, I. Khorozyan, M. Waltert, and T. Kuemmerle, "Understanding spatial patterns of poaching pressure using ranger logbook data to optimize future patrolling strategies," *Ecological Applications*, vol. 32, no. 5, p. e2601, 2022.

[18] M. Khan, A. Ali, and Y. Alharbi, "Predicting and preventing crime: a crime prediction model using san francisco crime data by classification techniques," *Complexity*, vol. 2022, no. 1, p. 4830411, 2022.