Received: 26 September 2025 | Revised: 3 March 2026 | Accepted: 24 March 2026 | Published online: 14 April 2026

RESEARCH ARTICLE

Archives of Advanced Engineering Science
2026, Vol. 00(00) 1–12
DOI: 10.47852/bonviewAAES62027763

BON VIEW PUBLISHING

Predictive Modeling of Public Sentiment Toward MONUSCO's Mandate: A Machine Learning Approach

Héritier Nsenge Mpia[1,*] and Moïse Kambale Kasambya[2]

[1] *Faculty of Sciences and Technology, Université de l'Assomption au Congo, Democratic Republic of the Congo*

[2] *Department of Business and IT, Université de l'Assomption au Congo, Democratic Republic of the Congo*

Abstract: In eastern Democratic Republic of the Congo, persistent insecurity and protracted armed conflict have intensified public dissatisfaction with the United Nations Organization Stabilization Mission in the Democratic Republic of the Congo (MONUSCO). In Butembo, debates over renewing MONUSCO's mandate have become increasingly contentious, yet empirical evidence on the determinants of public opposition remains limited. This study addresses this gap by using machine learning to predict refusal to extend MONUSCO's mandate from primary survey data. A cross-sectional quantitative survey was conducted among 5518 respondents in Butembo using a structured questionnaire comprising 28 items on sociodemographic characteristics and perceptions of MONUSCO's performance and local security conditions. After excluding the outcome variable, exploratory factor analysis identified latent structures in the data. Fifteen variables with factor loadings above 0.40 were retained and grouped into two dimensions: sociodemographic factors and self-conviction factors. These variables were then used to train and test three classifiers: decision tree, logistic regression, and support vector machine (SVM). SVM yielded the best predictive performance, achieving 93.1% accuracy on the held-out test set, compared with 90.8% for logistic regression and 89.6% for decision tree. The best-performing model was further deployed in a Flask-based web prototype for real-time prediction. Overall, the study demonstrates the value of combining latent variable modeling and machine learning to analyze public opinion in conflict-affected settings.

Keywords: machine learning algorithms, contextual factors, sociodemographic factor, self-conviction factor, prediction of social facts

1. Introduction

The population of Butembo, located in North Kivu in the Democratic Republic of Congo (DRC), has lived for many years in a climate marked by recurring insecurity and deep social instability. Over time, and particularly over the last two decades, local confidence in the United Nations Organization Stabilization Mission in the Democratic Republic of the Congo (MONUSCO) has gradually weakened. Many residents question its effectiveness, often drawing on their own experiences of violence and the continued presence of armed groups. Butembo—an important commercial hub with a rapidly growing population—offers a revealing illustration of the tensions that can arise between local communities and international peacekeeping forces [1]. Understanding how people form their opinions about MONUSCO's mandate is therefore essential both for local authorities and for broader peacekeeping strategies.

Earlier research on MONUSCO has mostly relied on qualitative or descriptive approaches. These studies underline recurring criticisms such as perceived passivity, operational delays, and a limited impact on restoring security. However, they often stop short of quantifying or modeling the factors that shape public dissatisfaction [2]. In parallel, other work in conflict-affected regions has shown how machine learning (ML) techniques can help forecast political instability or outbreaks of violence. These studies demonstrate the potential of data-driven approaches in complex social contexts. Yet, to the best of our knowledge, such methods have not been applied to examine public dissent toward peacekeeping missions in the DRC [3]. This gap motivates the present study, which applies ML models to primary data collected from 5518 respondents to determine whether individuals support or oppose the extension of MONUSCO's mandate [4].

The conceptual basis of this research draws from theories of public trust, institutional legitimacy, and political behavior [5, 6]. Public sentiment—understood here as the shared attitudes or convictions of a population—plays a decisive role in whether peacekeeping efforts are accepted or rejected. In the case of MONUSCO [7], people's views are shaped not only by their personal experiences with insecurity but also by broader historical, social, and political influences, including perceptions of fairness, competence, and transparency [8]. Studies on peacekeeping

*Corresponding author: Nsenge Mpia Héritier, Faculty of Sciences and Technology, Université de l'Assomption au Congo, Democratic Republic of the Congo. Email: nsengempia@uaconline.edu.cd

generally suggest that when any of these dimensions is perceived as weak or inconsistent, trust tends to erode.

To move beyond purely qualitative descriptions, this study integrates two categories of variables: sociodemographic characteristics (such as age, gender, marital status, and education level) and self-conviction factors (including views about MONUSCO's interests, its mission, and its effectiveness) [9]. Using exploratory factor analysis (EFA) [10], the authors extracted two latent constructs representing these dimensions, following established psychometric standards [11]. These constructs were then used as predictors in ML models aimed at classifying respondents according to their willingness or refusal to extend the mandate [2].

ML techniques [12, 13] have previously shown strong performance in sociopolitical research, particularly when analyzing complex behavioral patterns. Algorithms such as support vector machines (SVM) [14], logistic regression [15, 16], and decision trees [17] are widely used because they can handle structured datasets and detect subtle relationships that are not always visible through traditional statistical techniques. Prior studies show that SVMs often perform particularly well in classification tasks involving multidimensional data. Building on these insights, the present study evaluates the three algorithms to determine which best predicts refusal to extend MONUSCO's mandate [18]. Studies such as Zhang et al. [19] and Muzalia Kihangu et al. [9] have documented local sentiments regarding international interventions, emphasizing the role of perceived passivity and vested interests in shaping public opinion. These works suggest that trust in peacekeeping missions depends not only on their operational success but also on their transparency and accountability.

By combining contextual understanding, latent variable extraction, and predictive modeling, this study offers a quantitative perspective on public perceptions in a conflict-affected environment. It complements existing literature by providing measurable evidence of the factors associated with dissatisfaction and by showing how ML-based approaches can support more informed and responsive decision-making in peacekeeping operations [20–23].

2. Methods and Materials

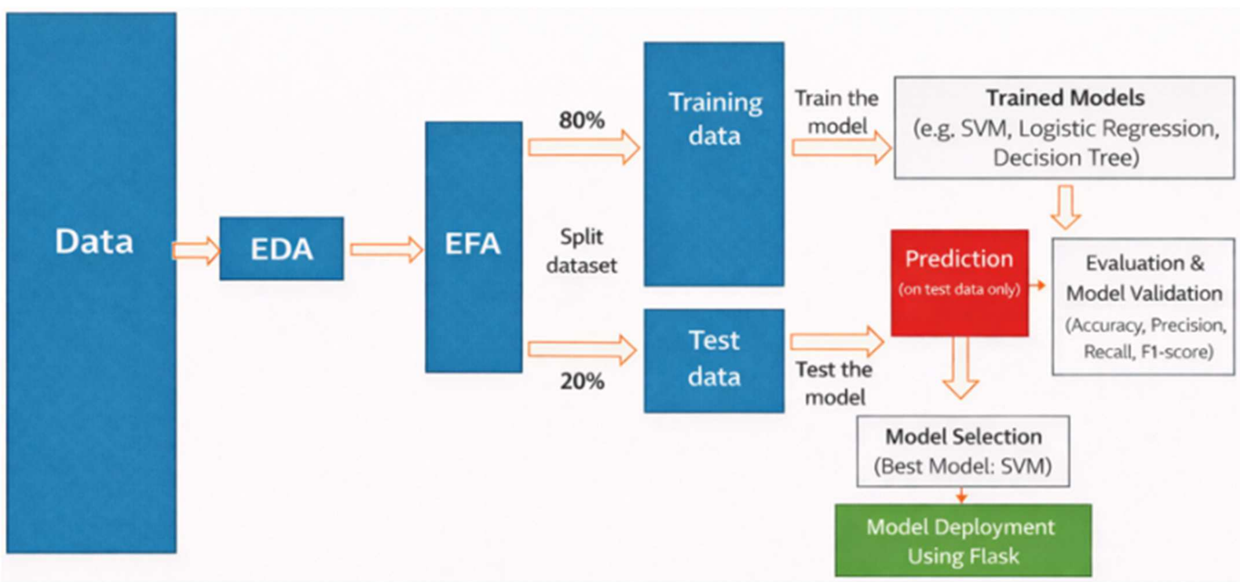
2.1. Research design

In this research, the authors used the quantitative method, which is based on numerical variables that correspond to quantities because it designates several things, particularly data or techniques for collecting and processing these data [24]. A quantitative study is used to prove or demonstrate facts by quantifying a phenomenon. The results are often expressed in the form of figures (statistics). This method can, for example, be carried out using a survey (answers to one question) or a questionnaire (answers to several questions). The results of a quantitative study are expressed in figures and can be used to calculate averages, count the frequency of a given response, and divide data into percentages. The results of quantitative studies are most often found in the form of statistical tables or graphs [25]. The study used primary data. The primary data used were original and unique and were collected directly by the researcher from a single source according to his or her needs. Secondary data are readily available but are not pure because they have undergone extensive statistical processing. The quantitative data collected were used to develop the ML models. The stages of the overall investigation of this research are summarized and visualized in Figure 1.

The pipeline includes data collection, exploratory data analysis (EDA), data preprocessing, EFA, ML model training and evaluation, and deployment using Flask. EFA was estimated on the training data only, and the derived factors were used as inputs for both training and test sets. The final selected model, trained on the training data and evaluated on the test set, was deployed using Flask.

To obtain the raw data, the authors used a survey questionnaire. This enabled them to collect data from the respondents. EDA was conducted as the second phase to examine the dataset's distribution, identify patterns, and detect anomalies using visualizations and summary statistics. EDA served to inform the researchers about potential data inconsistencies and the relationships among variables. Visualizations and statistical summaries

Figure 1
Workflow of the study



were used to identify patterns, outliers, and correlations between variables. This step was crucial for determining how to clean and transform the data for subsequent analysis. Unlike data preprocessing, which prepares data for modeling, EDA is a distinct process aimed at understanding and investigating the dataset before feature selection and modeling [26]. Following this, the data preprocessing phase was conducted. This phase included cleaning the raw survey data by addressing missing values and encoding categorical variables into numeric formats. These preprocessing steps were necessary to ensure that the dataset was compatible with ML algorithms and ready for model training [26]. In this process, missing data (<2%) were handled via median imputation to preserve the distributional characteristics of continuous variables. All numerical features were standardized using z-score normalization to ensure comparability across scales. A fixed randomization seed (42) was applied during the 80/20 train-test split to enhance reproducibility of model evaluation results.

To achieve the first objective of this study, the authors conducted an EFA. EFA was conducted to identify the underlying contextual factors that predict the refusal to extend MONUSCO's mandate. This step aligns with the study's objective of determining significant features for the predictive model, as illustrated in Figure 1. EFA is a statistical method used to uncover latent structures within datasets by grouping variables based on their correlations [27]. EFA was preferred over principal component analysis (PCA) as it reveals latent constructs underlying observed correlations, consistent with psychometric and behavioral modeling standards [28]. In contrast, PCA merely reduces dimensionality without explicitly modeling latent structures or measurement errors, making EFA more suitable for theory-driven social science research.

By using EFA, the authors followed these steps: (1) Data Adequacy Assessment. The Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity were performed to assess the suitability of the dataset for EFA. (2) Factor Extraction. The study employed the "minimum residual" method for factor extraction, and factors were rotated using the "varimax" method to maximize interpretability [10]. (3) Factor Interpretation. Sixteen variables with factor loadings greater than 0.4 were grouped into two latent factors: (i) sociodemographic factors (age, gender, marital status, place of birth, education level, profession, security in the neighborhood, and political affiliation) and (ii) self-conviction factors (perceptions regarding MONUSCO's interests, its mission contract, local security institutions, and beliefs about whether insecurity will end if MONUSCO leaves). EFA was crucial in identifying features that reflected both contextual and respondent-driven factors. These latent variables provided a compact, meaningful representation of the original dataset, which was essential for reducing dimensionality and improving model performance. By identifying these two key factors, the study ensured that only the most relevant predictors were used in training the ML models [13].

The study employed three ML algorithms: SVM, logistic regression, and decision tree. These algorithms were chosen due to their complementary strengths in handling classification problems and their established use in similar studies. SVM is a robust supervised learning algorithm, widely used for binary classification tasks, as it finds an optimal hyperplane that separates classes effectively, even in high-dimensional spaces. Its ability to use the kernel trick allows it to handle complex, nonlinear decision boundaries. For the SVM classifier, probabilistic outputs were enabled using the probability = True option in scikit-learn, which applies Platt scaling to transform decision scores into calibrated probability estimates based on the training data [13]. Previous studies, such as those by Musumba et al. [4], have demonstrated SVM's superior accuracy in predictive modeling for sociopolitical outcomes. Logistic regression was selected for its simplicity, interpretability, and strong performance on linearly separable data. It is particularly effective for problems involving binary outcomes, as in this study, where the target variable is whether respondents support the extension of the MONUSCO mandate. It has been a foundational algorithm in conflict prediction due to its capability to model probabilities and provide insights into feature importance. Decision trees were included for their intuitive interpretability and ability to handle both categorical and numerical data. Decision trees are particularly suitable for understanding the hierarchical importance of contextual factors influencing public opinion. Studies such as Mutsotsya et al. [13] have highlighted the effectiveness of decision trees in exploratory research involving diverse contextual features. By using these three algorithms, the study ensured a robust comparison across models with distinct approaches to classification. The selected algorithms have been validated in previous research for their predictive accuracy and relevance to social and political data.

Those models were evaluated using four metrics to assess their final performance and compare them with each other. With the large number of metrics available, it is not always easy to choose the one that best suits the use case. The final result can significantly differ depending on the metric used to optimize and select the models [29]. For this study, the models were evaluated for accuracy, precision, recall, and F-score. The evaluation is used to test the final performance of the algorithm and is carried out on the test set. The precision is the accuracy. The percentage of tuples that the classifier labels as positive is actually positive. Recall is complete. The percentage of positive tuples for which the classifier was labeled as positive was also recorded.

The accuracy, which is the percentage of tuples in the test set that are correctly classified, is calculated using the following formula. The harmonic mean of precision and recall is used to obtain the F-score. Therefore, the F-score will only take a high value if both of its components are high. When the number of classes is not balanced, the F-score often reflects the precision [30].

For better model validation, the dataset was divided into a training set and a test set. Specifically, the authors allocated 80% of the data to the training set and 20% to the test set. In addition, we went on to train our model. This procedure enabled us to select four algorithms from which we selected the most suitable algorithm. The model training stage was followed by the evaluation stage to validate the model that performed best.

2.2. Target population

The population of a study is nothing more than the set of elements of an observed phenomenon that have the same properties and are of the same nature. In other words, it is the statistical universe to which the researcher questions himself to gather the necessary information. The target population is the entire population that the results of the survey should represent. This may be the case for an entire country or a single region. The set of members of a specific group on which the results will be applicable is known [10]. The target population for this study was the entire population of the city of Butembo, located in the Province of North Kivu, DRC.

2.3. Sampling procedure

A simple random sampling approach was employed to collect primary data from the population of Butembo, DRC. Under this approach, each eligible individual had an equal probability of being selected, thereby minimizing selection bias and supporting the generalizability of the findings [31]. This method is appropriate for large and heterogeneous populations where no prior information on subgroup proportions is required.

Given that the total population size was not precisely known in advance, Cochran's formula was used to estimate the minimum required sample size at a 95% confidence level and a 5% margin of error:

$$n0 = \frac{Z^2 \times p \times (1-p)}{e^2}$$

where Z represents the Z-score corresponding to the desired confidence level (1.96 for 95%), p is the estimated population proportion (set at 0.5 to maximize variance), and e denotes the margin of error (0.05). Based on this calculation, the minimum sample size was estimated at 384 respondents. The minimum sample size was therefore computed as follows:

$$n0 = \frac{(1,96)^2 \times (0,5) \times (1-0,5)}{(0,05)^2}$$

$$n0 = \frac{3,8416 \times 0,5 \times 0,5}{0,0025} = 384,16$$

To enhance the robustness and precision of the analysis, data were ultimately collected from 5518 respondents, substantially exceeding the minimum requirement. Although sociodemographic characteristics such as age, gender, education level, and profession were monitored to ensure diversity in the sample, no formal stratification procedure was applied during respondent selection. This approach ensured broad population coverage while maintaining the randomness of the sampling process [25].

2.4. Research instrument

To collect the data, the authors used Google Forms, developed by Google, to create an online quiz, form, or survey to gather information or conduct a survey or poll. The questionnaire for this research took into account demographic characteristics, questions related to the respondent's sociopolitical context, and questions related to the respondent's views on the MONUSCO mandate in the DRC. The survey consisted of 28 closed questions, including 6 binary questions (which offered only 2 response options) and 22 Likert-type questions. The dependent variable for this study was "MONUSCO must necessarily leave the region: Yes or No." For the independent variables, 5 questions were considered dyadic, 22 questions were given a 5-point Likert scale, and 1 question was given a 6-point Likert scale.

2.5. Validity and reliability of the instrument

After finalizing the survey questionnaire, it was reviewed by subject-matter experts to assess its content validity, clarity, and relevance to the study objectives, as well as to minimize potential respondent fatigue. Reliability analysis was conducted on a subsample of 1103 respondents, corresponding to approximately 20% of the full dataset, which is a commonly adopted proportion for internal consistency assessment.

From a theoretical perspective, reliability was examined in line with classical internal consistency frameworks, including Cronbach's reliability theory and McDonald's conceptual approach to scale consistency. In practice, internal consistency was quantified using Cronbach's alpha, which remains the most widely reported and interpretable reliability coefficient in EFA. Cronbach's alpha was computed using Python based on the standard formulation:

$$\alpha_{st} = \frac{N \times \bar{r}}{1 + (N-1) \times \bar{r}}$$

where N denotes the number of items and $\bar{r}$ represents the average inter-item correlation [27].

Following the EFA, two latent factors demonstrated satisfactory internal consistency, with Cronbach's alpha values of 0.82 and 0.80, respectively (Table 1). These values exceed commonly accepted thresholds for exploratory research and confirm the reliability of the extracted factors. The two factors jointly explained 76.27% of the total variance, supporting their retention for subsequent ML modeling [10].

2.6. Data processing and analysis

The data collected were first translated into English, given that the language spoken in the DRC was French. All French responses were double-translated by bilingual experts and validated using a back-translation procedure to ensure linguistic consistency and semantic equivalence [32]. This multistep verification process ensured that key sociopolitical expressions and cultural nuances were preserved across translations, reducing the risk of semantic bias in the ML inputs. These data, validated and standardized, were then processed using the Python function map( ) and saved in CSV format for subsequent analysis. Although the minimum required sample size was satisfied, the authors chose to retain the entire dataset (N = 5518) in order to preserve population diversity and the underlying multivariate structure of the data. The proportion of missing values was very low (below 2%) and showed no systematic pattern across variables. For this reason, median imputation was applied as a cautious and pragmatic approach, allowing the authors to limit unnecessary data loss while preserving the distributional characteristics and correlation patterns required for EFA and subsequent ML modeling [10]. Pandas is the library that helped to import it into the Jupyter Notebook. Pandas is a Python library used to manipulate and analyze tabular data and time series [26]. On the

Table 1
Extracted retained factors

Factor	Eigenvalue	Explained variance	% of variance	Cumulative variance	Cronbach's alpha
Factor 1	3.664383	2.950035	39.25	39.25	0.82
Factor 2	3.272724	2.347413	37.02	76.27	0.80

Note: Extraction Method: "Minimum residual"; Rotation Method: "Varimax".

other hand, for activities related to the development of ML models, the authors used the Python library sklearn. This library is more useful and robust for ML in Python. It provides a selection of efficient tools for ML and statistical modeling, including classification, regression, clustering, and dimensionality reduction via a Python consistency interface. The library, which is largely written in Python, is built on NumPy, SciPy, and Matplotlib [33].

The suitability of the data for EFA was assessed using the KMO measure of sampling adequacy and Bartlett's test of sphericity. Bartlett's test evaluates whether the correlation matrix significantly differs from an identity matrix. The null hypothesis assumes that the variables are uncorrelated; rejection of this hypothesis indicates that sufficient inter-variable relationships exist to justify the application of factor analysis [27]. In this study, sampling adequacy was confirmed (KMO = 0.78), and Bartlett's test was statistically significant (χ^2 = 3567.2, p < 0.001), meeting conventional thresholds for factor analysis validity. The EFA was conducted using Python libraries such as FactorAnalyzer and Statsmodels [27].

After model training, the best-performing SVM model was serialized using the pickle module, allowing it to be stored and reloaded efficiently. This serialized model was then integrated into a Flask-based web application, which serves as an interface enabling users to input new data and receive real-time predictions of public sentiment. Flask was employed to manage HTTP requests, handle user inputs, and deploy the predictive model as an accessible web service [13]. After training the developed models with Python, the authors deployed the outperformed model in web technology using the Flask framework.

Flask is a small, lightweight Python web framework that provides useful tools and features that make it easy to construct web applications in Python. It offers developers a degree of flexibility and is a more accessible framework for new developers, as you can quickly build a web application using a single Python file. The Flask is REST-based. Flasks are also extensible and do not force a particular directory structure or require complicated standard code before starting [13].

The potential for overfitting is a critical concern in ML models, particularly when analyzing sociopolitical data. To address this issue, the dataset was first split into training (80%) and test (20%) subsets using a fixed random seed to ensure reproducibility. The held-out test set was used exclusively to report final model performance and to evaluate generalization to unseen data [26]. Model complexity was controlled through regularization, particularly in the SVM classifier, and dimensionality reduction was achieved by retaining only variables with significant factor loadings from the EFA [27]. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics computed on the test set. In addition, a 10-fold cross-validation procedure was conducted on the training data to assess model stability and robustness. Cross-validation results are reported as mean accuracy and standard deviation and are used solely as a consistency check rather than as final performance estimates. The cross-validation outcomes were consistent with the test-set results, confirming the stability and generalizability of the selected models [34]. The consistency between test-set performance and cross-validation results further suggests that the observed performance is not attributable to overfitting.

2.7. Addressing potential biases in data collection

The primary data for this study were collected using a structured questionnaire administered to 5518 respondents from Butembo. While care was taken to design an unbiased and comprehensive survey, potential biases in the data collection process must be acknowledged:

1) Selection bias. The study utilized a random sampling method to minimize selection bias. However, focusing solely on respondents from Butembo may limit the generalizability of the findings to other regions of the DRC. Future research should include other geographic areas to enhance the generalizability of the results.
2) Response bias. Respondents' self-reported data may have been influenced by social desirability or fear of expressing dissenting opinions in a politically sensitive context. Such response biases are common in conflict-prone regions and can skew survey results. To mitigate this, the questionnaire was anonymized, and neutral language was used to reduce the risk of biased responses for sensitive research settings.
3) Sampling frame limitations. The lack of precise demographic data for Butembo's population made it challenging to ensure the sample was fully representative of all subgroups. Using Cochran's formula for sample size determination helped ensure an adequate sample size. However, oversampling certain demographics may have inadvertently skewed the dataset. Future studies should aim to refine sampling techniques and include stratified sampling where population data are available.
4) Questionnaire design. Closed-ended questions, while useful for quantitative analysis, may have constrained respondents' ability to fully express their views. Incorporating open-ended questions or mixed-methods approaches could provide richer insights into public sentiment [24]. Furthermore, cultural and linguistic considerations must be taken into account to ensure questions are understood consistently by respondents.

2.8. Criteria for selecting variables

The variables included in this study were selected through a combination of theoretical considerations and empirical validation. The initial pool of 28 variables was identified based on an extensive review of the literature on public sentiment and peacekeeping missions, focusing on factors frequently associated with public opinion in conflict zones [18]. Sociodemographic variables, such as age, gender, education level, and marital status, were included due to their demonstrated influence on public sentiment in similar studies [4]. Self-conviction factors, such as perceptions of MONUSCO's performance, vested interests, and passivity, were selected for their relevance to the research objectives and alignment with prior findings on public dissatisfaction with peacekeeping missions [9].

To refine this selection, an EFA was conducted, guided by statistical criteria. The KMO test value and Bartlett's test of sphericity result confirmed the dataset's adequacy for factor analysis. Variables with factor loadings greater than 0.4 were retained, as this threshold is commonly used to identify statistically significant relationships in social science research [27]. The final set of 15 variables grouped into two factors, which are the sociodemographic and the self-conviction, was chosen for their high correlations with latent constructs and their relevance to the study's objectives. This rigorous selection process ensured the inclusion of variables that were both theoretically grounded and empirically validated, providing a robust foundation for the ML models [10].

3. Results and Discussion

3.1. Demographic information

After the data preprocessing phase, especially the EDA stage, several responses or sociodemographic elements of the respondents in this research were visualized. The authors collected sociodemographic information on place of birth and place of growth. They therefore constructed a profile of the respondents according to the method of data collection chosen, using graphs to explain our responses. Figure 2 shows that many of the respondents in this study were born in the town of Butembo (52.2%), the town of Beni (25.4%), the area around Butembo (11.5%), the area around Beni (7.2%), or elsewhere (3.7%). The demographic factor in this research also consisted of determining the growth rate (the environment in which people grew up). The answers obtained show that the inhabitants of Butembo represent 67.3%, those of Beni represent 16.9%, those of Beni represent 6.1%, those of Butembo represent 8%, and those of other provinces represent 1.6%. These two demographic variables played an important role in this research, as they helped the authors to conduct their research in society.

3.2. Research results

3.2.1. Relationship between features and the target variable

To examine the relationship between the predictor variables and the target variable ("refusal to extend MONUSCO's mandate"), pairwise correlation coefficients were computed. For clarity, Figure 3 presents pairwise correlations for a representative subset of key predictors, rather than the full set of variables retained after EFA. Overall, all associations between individual predictors and the target variable were weak in absolute terms ($|r| < 0.2$), indicating the absence of strong linear relationships.

Among sociodemographic variables, age ($r \approx 0.05$), education level ($r \approx -0.08$), and gender ($r \approx 0.18$) showed weak correlations with mandate refusal. These coefficients should therefore not be interpreted as meaningful standalone effects, but rather as relative differences among predictors. For instance, gender exhibited a comparatively higher correlation than age or education, although all associations remain weak.

Similarly, self-conviction variables displayed weak correlations with the target variable. Trust in nongovernmental organization (NGO) showed a negative association ($r \approx -0.18$), while perceptions of MONUSCO's vested interests ($r \approx 0.12$) and the belief that peace would be restored after MONUSCO's departure ($r \approx 0.06$) were weakly and positively correlated with mandate refusal. Other contextual variables, including neighborhood, place of birth, and political engagement, exhibited negligible correlations ($|r| < 0.10$).

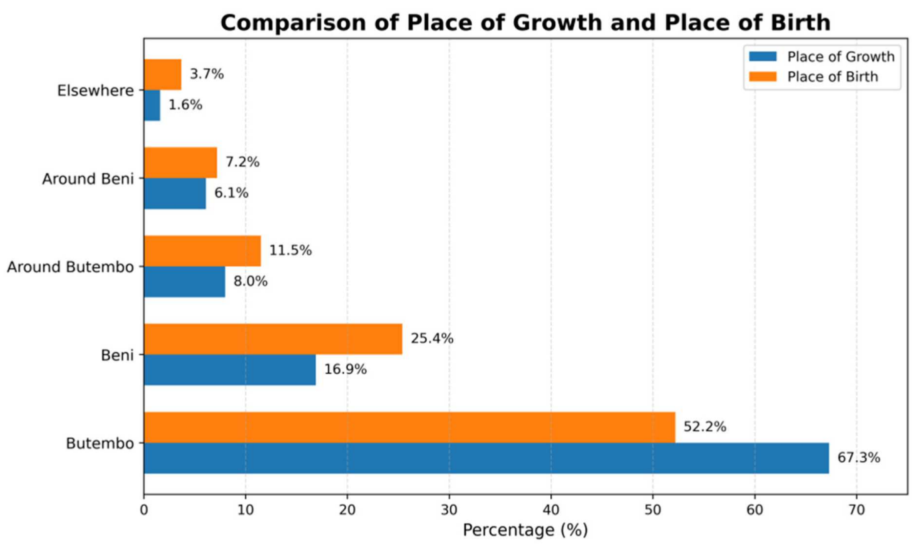
Taken together, these results indicate that no single predictor demonstrates a strong linear association with refusal of MONUSCO's mandate. Instead, the correlations are best interpreted comparatively, highlighting relative differences across variables rather than absolute effects. This finding supports the use of multivariate ML models to capture nonlinear interactions and combined effects that cannot be inferred from pairwise correlations alone.

3.2.2. Results for achieving the first research objective

To achieve the first objective of this study, the authors had to first execute the KMO and Bartlett tests to determine the feasibility of the EFA study and whether the available data were suitable for conducting EFA. As a result, the overall KMO value for the data was 0.78, which was satisfactory. This figure suggested that the researchers could carry out their intended factor analysis. The Bartlett test was significant at p = 0.0. The authors were able to reject the null hypothesis that the matrix was identical as a result of this. After the EFA phase, two factors were found to be reliable, as indicated by Cronbach's alpha values of 0.82 and 0.80, as shown in Table 1. The total cumulative variance percentage was 76.27%, which is a decisive criterion in deciding the number of factors to retain [10].

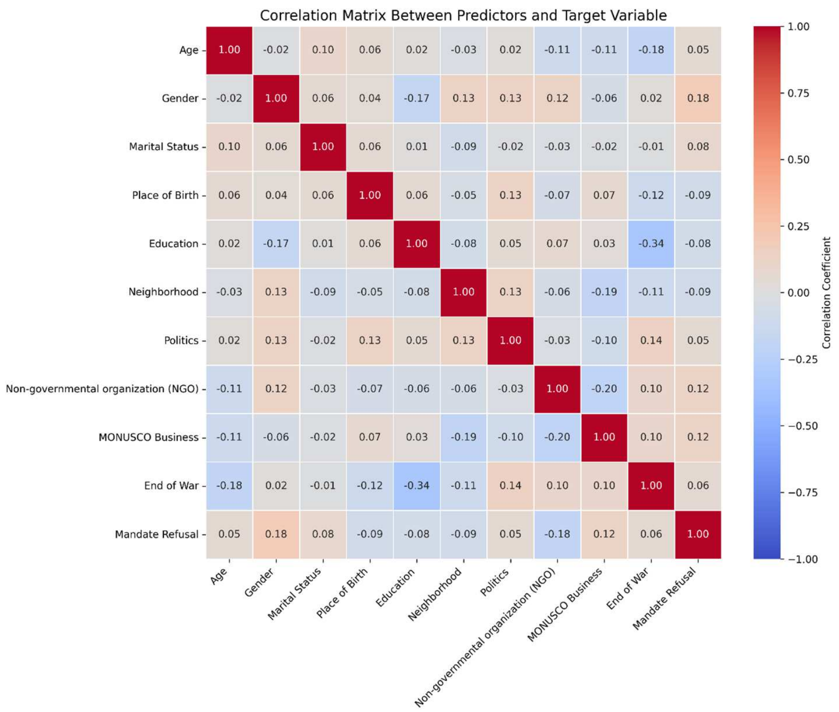
The study followed the commonly accepted rule of thumb by retaining only variables with factor loadings greater than 0.4 [10]. The questionnaire initially comprised 28 items, including one target variable measuring refusal to extend MONUSCO's mandate. This target variable was excluded prior to the EFA, resulting in 27 observed variables used for factor extraction. Based

Figure 2
Distribution of respondents by place of birth and place of growth



Data source: Primary survey (N = 5518).

Figure 3
Correlation matrix between predictor variables and target variables (“Refusal to extend MONUSCO’s mandate”). Color intensity represents Pearson correlation values



Source: Authors’ computation in Python (sklearn, seaborn).

on the EFA results, 15 variables loaded onto two latent factors and were retained as contextual predictors for subsequent ML modeling. The corresponding factor loadings for these 15 variables are reported in column 4 of Table 2.

The EFA yielded two latent factors with satisfactory reliability: Factor 1 (sociodemographic) (α = 0.82) and Factor 2 (self-conviction) (α = 0.80). Fifteen variables with loadings $\geq$ 0.4 were retained (see Table 2). These numeric results fulfill the first research objective by identifying statistically supported constructs underlying public sentiment toward MONUSCO’s mandate.

The emphasis on sociodemographic and self-conviction factors derives from these results, as both groups of variables showed the strongest factor loadings. This confirms that demographic attributes and individual perceptions jointly influence public attitudes toward MONUSCO’s mandate.

3.2.3. Results for achieving the second research objective

The authors chose the SVM, decision tree, and logistic regression algorithms for this study. A data sample of 5,518 records was used to predict the refusal to extend MONUSCO’s mandate to the DRC. The following findings were obtained using the performance measures shown in Table 3. Table 3 shows that the SVM outperformed the other methods, achieving an accuracy of 93.1% on the held-out test set. According to the analysis of the decision tree algorithm, the accuracy was 89.6%. Logistic regression had an accuracy of 90.8%.

To comprehensively evaluate the performance of the proposed models, a comparative analysis was conducted among the three algorithms utilized. These models were selected due to their diverse methodological approaches and relevance in sociopolitical studies [10]. The evaluation metrics (accuracy, precision, recall, and F1-score) were applied consistently to ensure a reliable comparison. The SVM model outperformed the other two, achieving the highest accuracy of 93.1%, as well as superior precision, recall, and F1-score [14]. Logistic regression followed with an accuracy of 90.8%, demonstrating its capability to handle linearly separable data effectively. The decision tree model, while intuitive and interpretable, achieved an accuracy of 89.6%, reflecting its limitations in handling complex, nonlinear relationships. This comparison underscores the robustness of the SVM model in high-dimensional feature spaces and its suitability for the prediction task within this sociopolitical context.

The bar graph in Figure 4 illustrates the performance of SVM, logistic regression, and decision tree models across accuracy, recall, precision, and F-score metrics. SVM outperforms the other models in all metrics, achieving the highest accuracy (93.1%), recall (99%), precision (99%), and F-score (99%), demonstrating its robustness and reliability. Logistic regression and decision tree exhibit comparable performance, with accuracy around 90%, and slightly lower recall, precision, and F-score values (approximately 91% and 89%). These results highlight SVM as the most effective model for this dataset, minimizing errors and providing superior predictive capability.

Knowing that the second objective of this study was to determine which ML algorithm best predicted the refusal of MONUSCO’s mandate extension, the results of the comparison in Table 3 showed that the SVM algorithm was the best compared to the others. This led the authors to conclude that the SVM algorithm best predicted this social reality.

3.2.4. Results for achieving the third research objective

In this section, the results relating to the third objective of this study are presented. This consisted of deploying the SVM model, which was found to perform best in Flask technology. To illustrate the model’s predictive functionality in a transparent and replicable manner, a data-driven example is presented

Table 2
Correlations between items and retained factors

Factor	Related items	Variable	Factor load
1	Your age	AGE	0.742393
	Your gender	GENDER	0.611087
	Marital status	MS	0.657237
	Place of birth	POB	0.530255
	Education level	EDUCATION	0.569666
	Profession	PROFESSION	0.651442
	The neighborhood where you live is secure	NEIGHBOR	0.698374
	Are you a member of a political party?	POLITICS	0.628374
2	Can a nongovernmental organization provide security where MONUSCO has failed?	NGO	0.699274
	The perpetual kidnappings in areas under MONUSCO control are proof of its passivity	KIDNAPPING	0.798374
	Cutting off civilians' roads in areas under MONUSCO control is proof of MONUSCO's passivity	ROADCUTOFF	0.688354
	MONUSCO has vested interests in eastern DRC	UNBUSINESS	0.658201
	MONUSCO and the Congolese government have a vested interest in the insecurity in the east of the country	INTEREST	0.675896
	The MONUSCO mission contract must be revised if it is to remain in the region	REVISIONCON-TRACT	0.798362
	Insecurity will end if MONUSCO leaves the DRC	ENDWAR	0.824513

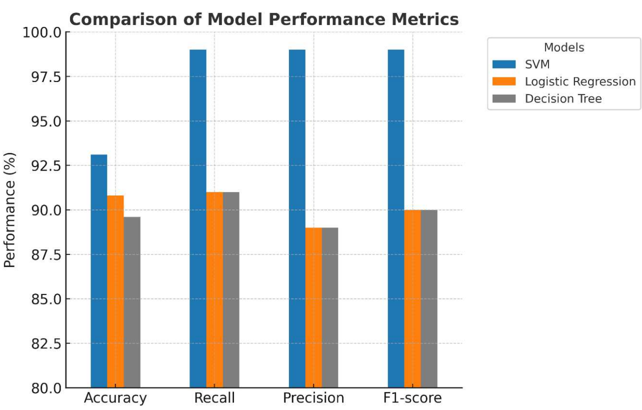
Note: AGE = Age; GENDER = Gender; MS = Marital Status; POB = Place of Birth; EDUCATION = Education Level; PROFESSION = Occupation; NEIGHBOR = Neighborhood Security; POLITICS = Political Affiliation; NGO = Trust in Nongovernmental Organization,; KIDNAPPING = Perceived MONUSCO Passivity through Kidnappings; ROADCUTOFF = Perceived MONUSCO Passivity through Road Closures; UNBUSINESS = MONUSCO's Business Interests; INTEREST = MONUSCO–Government Vested Interest; REVISIONCONTRACT = Need for Mandate Revision; ENDWAR = Belief that War Ends after Withdrawal.

Table 3
Test-set performance of ML models predicting refusal to extend MONUSCO's mandate

No.	Models	Accuracy (%) ± SD	Recall ± SD	Precision ± SD	F-score ± SD
1	SVM	93.1 ± 1.2	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01
2	Logistic regression	90.8 ± 1.5	0.91 ± 0.02	0.89 ± 0.02	0.9 ± 0.02
3	Decision tree	89.6 ± 1.7	0.91 ± 0.03	0.89 ± 0.03	0.9 ± 0.03

Source: Authors' computation using Python.
Note: Values represent performance on the held-out test set. Mean ± standard deviation from 10-fold cross-validation conducted on the training data are reported separately to assess model stability.

Figure 4
Comparison of SVM, logistic regression, and decision tree models across accuracy, precision, recall, and F1-score metrics



Source: Authors' analysis in Python.

instead of a narrative use case. Table 4 summarizes anonymized sample inputs and the corresponding predicted probabilities of mandate refusal generated by the SVM classifier. The predicted probabilities reported in Table 4 correspond to calibrated probability estimates produced by the SVM model using Platt scaling.

The above tabular presentation conveys the operational logic of the predictive model: respondents combining low institutional trust and high perceptions of MONUSCO's vested interests exhibit the highest predicted probability of rejecting the mandate. Conversely, individuals with greater trust in domestic institutions and limited attribution of profit motives to MONUSCO show a substantially lower likelihood of opposition.

3.2.5. Model interpretability and feature importance

To enhance transparency and interpretability, a SHapley Additive exPlanations (SHAP) analysis was conducted to assess

Table 4
Example of model input variables and predicted probabilities of mandate refusal

Respondent	Age	Education level	Trust in non-governmental organization	Belief in MONUSCO's business interests	Expectation of end of war	Predicted probability of mandate refusal
R-1032	42	Secondary	2 (Low)	4 (High)	3 (Moderate)	0.84
R-2175	29	University	5 (High)	1 (Low)	1 (Low)	0.22
R-3509	37	College	3 (Moderate)	3 (Moderate)	4 (High)	0.68
R-4582	54	Primary	1 (Very Low)	5 (Very high)	5 (Very high)	0.93

the marginal contribution of each feature to the SVM model's predictions. The results indicate that trust in nongovernment organization (NGO), expectations regarding the end of insecurity (ENDWAR), and education level (EDUCATION) were the most influential predictors of refusal to extend MONUSCO's mandate, as reflected by their higher SHAP values. These findings suggest that perception-based variables play a more prominent role than basic sociodemographic characteristics, such as age or gender, in shaping public opinion.

Other influential predictors included perceptions of MONUSCO's vested interests (UNBUSINESS), which further emphasize the importance of trust, institutional performance, and perceived mission effectiveness in explaining opposition to the mandate. In contrast, structural demographic attributes such as marital status or place of birth exhibited comparatively low SHAP importance, indicating a limited contribution to the model's predictive outcomes.

3.3. Discussion of results

The findings of this study provide insights into the factors associated with refusal to extend MONUSCO's mandate in Butembo, DRC. By applying ML techniques, the analysis explored how sociodemographic characteristics and self-conviction factors relate to public attitudes toward the mission. This section situates the results within the existing literature, discusses their implications, and outlines limitations and threats to validity. Previous studies have documented persistent skepticism regarding MONUSCO's effectiveness, a perception that is reflected in the present analysis, where variables linked to perceived passivity of the mission, such as kidnappings and road blockages, emerged as relevant predictors in the modeling stage.

From a methodological perspective, the EFA confirmed the suitability of the dataset for latent structure extraction (KMO = 0.78; Bartlett's test $p < 0.001$). Two internally consistent constructs were identified, with Cronbach's alpha values of 0.82 and 0.80, jointly explaining 76.27% of the total variance. These results support the retention of sociodemographic and self-conviction dimensions as meaningful representations of the underlying data structure for subsequent modeling.

The correlation analysis presented in Figure 3 showed that associations between individual predictors and mandate refusal were weak in absolute terms. Sociodemographic variables such as gender ($r = 0.18$), age ($r = 0.05$), and education level ($r = -0.08$) exhibited only limited linear relationships with the target variable. Similarly, perceptions related to MONUSCO's business interests ($r = 0.12$) and trust in an NGO ($r = -0.18$) were weakly correlated with mandate refusal. These coefficients are best interpreted comparatively, highlighting relative differences across variables rather than meaningful standalone effects.

Despite these weak pairwise correlations ($|r| < 0.2$), the SVM model achieved strong predictive performance compared to previous research [2, 4], with an accuracy of 93.1%. This result indicates that the model was able to capture nonlinear and multidimensional interactions among variables that are not apparent in bivariate analyses [35]. The performance gap between SVM, logistic regression (90.8%), and decision tree (89.6%) was modest, suggesting consistency across modeling approaches while confirming the robustness of the selected framework. The high predictive accuracy, combined with cross-validation results, indicates that the model generalizes well beyond the training data and is suitable for applied contexts such as monitoring public sentiment.

Importantly, the predictive contribution of variables such as perceptions of MONUSCO's passivity, beliefs regarding the end of insecurity, and education level should be understood within this multivariate context. While these features display weak linear associations individually, their combined influence within the SVM model demonstrates that public attitudes toward the mandate are shaped by interacting perceptions rather than isolated factors. This finding aligns with prior sociopolitical research showing that complex belief systems often emerge through interdependent influences rather than single determinants.

The results are also consistent with established theories of public trust and institutional legitimacy. In peacekeeping contexts, perceptions of effectiveness, fairness, and responsiveness play a central role in shaping public acceptance. In this study, negative evaluations of MONUSCO's role and greater confidence in NGOs were associated with lower support for mandate extension, reflecting broader dynamics of legitimacy and trust observed in fragile and conflict-affected settings.

The empirical findings support the assumptions outlined in the introduction. The EFA results indicated that sociodemographic variables (age, gender, education level, and marital status) formed a coherent latent construct, while self-conviction variables related to NGO trust and mission performance constituted a second factor. Within the SVM model, variables associated with perceptions of MONUSCO's role and expectations regarding insecurity contributed most strongly to prediction, suggesting that personal convictions and evaluations of institutional performance play an important role when considered jointly.

Finally, the findings should be interpreted in light of the specific sociopolitical context of Butembo, a region characterized by prolonged insecurity and complex governance challenges. Public attitudes toward MONUSCO are shaped not only by individual experiences but also by broader historical and political narratives [36, 37]. While the model identifies relevant predictors of mandate refusal, future research could extend this work by incorporating additional contextual variables and testing the framework in other regions to assess external validity.

Beyond its methodological contribution, the proposed predictive framework offers practical implications for peacekeeping and policy analysis. By identifying patterns associated with opposition to MONUSCO's mandate, the model may support targeted communication strategies and context-sensitive interventions aimed at improving institutional legitimacy and public engagement. These results illustrate how predictive analytics can complement traditional qualitative approaches, offering data-driven insights to inform decision-making in fragile contexts.

4. Conclusions, Contributions, and Recommendations

This study provides a quantitative and predictive analysis of public opinion on the MONUSCO mandate in Butembo, DRC, contributing to the understanding of factors influencing public sentiment on peacekeeping efforts. Guided by two assumptions, the study achieved its objectives and offered actionable insights. First, the research assumed that sociodemographic characteristics significantly shape public opinions. This assumption was validated, as key predictors, such as age, marital status, and education level, emerged as influential variables. Second, the study assumed that public perceptions of MONUSCO's operational performance and intentions, particularly perceived passivity and vested interests, drive dissatisfaction. The findings supported this, emphasizing the importance of trust and accountability in peacekeeping operations. The sensitivity analyses collectively confirm the robustness of the EFA results. The stability of the identified sociodemographic and self-conviction predictors across various conditions validates the reliability of the study's conclusions. This methodological rigor enhances confidence in the applicability of the findings to broader contexts and supports their use in predictive modeling frameworks.

The practical implications of these findings are manifold. The deployed predictive model based on the SVM algorithm serves as a robust tool for real-time sentiment analysis. This application enables policymakers and stakeholders to dynamically assess public opinion, design targeted interventions, and address critical grievances such as perceived inefficiencies and vested interests. Such data-driven strategies can enhance transparency, accountability, and trust, ultimately fostering greater acceptance of peacekeeping operations. Additionally, the methodology and findings of this study offer scalability, providing a framework that can be adapted to analyze public sentiment in other conflict-prone regions. Beyond predictive analytics, the integration of AI in peacekeeping logistics, early warning, and civilian protection offers constructive alternatives to its militarized applications in weapon systems, aligning with ethical AI deployment frameworks.

Beyond its specific application to the MONUSCO case, the proposed framework can be adapted to other regions and peacekeeping contexts by retraining the model with localized datasets. The Flask-based predictive system can also be scaled by incorporating new variables such as media narratives, social media sentiment, and local governance indicators. Furthermore, integration into institutional dashboards or early-warning systems would allow decision-makers in international organizations to monitor public sentiment in real time and anticipate legitimacy risks. This adaptability positions the framework as a transferable and data-driven decision-support tool for evidence-based peacekeeping and conflict prevention across diverse contexts. The predictive model offers clear policy and operational insights. MONUSCO and the Congolese government can use the results to monitor public acceptance of the mandate across regions. The model identifies key drivers of opposition, such as perceived passivity and low NGO trust. These insights can guide targeted communication and engagement with local communities. In addition, integrating the model into an early-warning system would help detect emerging dissatisfaction and support timely, data-driven policy responses.

Looking ahead, future research could explore the development of a recommender system that builds upon the predictive insights from this study. Such a system could offer personalized policy recommendations based on sociodemographic characteristics and perceptions of MONUSCO's performance, enabling stakeholders to tailor interventions more effectively to specific communities. Although the present analysis focuses exclusively on Butembo, the proposed predictive framework can be retrained with data from other regions of the DRC to enhance its generalization and policy relevance. Future extensions could apply the same methodology to provinces such as Beni, Goma, or Bukavu, using transfer learning or incremental retraining to capture regional variations in public perception. Such adaptations would support more inclusive, data-driven peacekeeping strategies and strengthen the external validity of the model. Future research could also integrate real-time or social media data to capture dynamic shifts in public sentiment. Combining survey-based predictors with online discourse would allow continuous model updating and richer contextual interpretation. Such hybrid data approaches are increasingly central to computational peace research, enabling early detection of emerging tensions and improving the responsiveness of peacekeeping strategies.

Finally, applying data science in fragile and politically sensitive contexts requires ethical caution. Misinterpretation or misuse of predictive outputs could reinforce biases or stigmatize certain communities. The study therefore stresses the importance of transparency, contextual understanding, and responsible data handling when deploying ML models for policy insights [38, 39]. Ensuring that algorithms are retrained with diverse and representative data is essential to avoid reinforcing existing inequalities and to promote ethical, inclusive, and trustworthy applications of AI in peacekeeping and governance.

Acknowledgments

The authors thank all respondents from Butembo who participated in the survey and the colleagues who assisted with questionnaire review and linguistic validation of the translated responses.

Ethical Statement

This research involved human participants and was conducted in accordance with applicable ethical standards. Ethical approval was granted by Université de l'Assomption au Congo under reference number UAC/Bbo.082024. The study was based on voluntary participation in a questionnaire survey. Informed consent was obtained from all participants prior to data collection, and responses were anonymized and analyzed in aggregate form.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/staniherstaniher/monusco-dataset-butembo.

Author Contribution Statement

Nsenge Mpia Héritier: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Kambale Kasambya Moïse:** Methodology, Software, Validation, Formal analysis, Investigation, Resources, Supervision.

References

[1] Kasolene, P. P. (2024). Les groupes de pression actifs en Ville de Butembo en quête d'identité. *Parcours et Initiatives: Revue interdisciplinaire du Graben (PIRIG)*, (26), 93–110. https://doi.org/10.57988/crig-2486

[2] Murphy, M., Sharpe, E., & Huang, K. (2024). The promise of machine learning in violent conflict forecasting. *Data & Policy*, *6*, e35. https://doi.org/10.1017/dap.2024.27

[3] Bazzi, S., Blair, R. A., Blattman, C., Dube, O., Gudgeon, M., & Peck, R. (2022). The promise and pitfalls of conflict prediction: evidence from Colombia and Indonesia. *Review of Economics and Statistics*, *104*(4), 764–779. https://doi.org/10.1162/rest_a_01016

[4] Musumba, M., Fatema, N., & Kibriya, S. (2021). Prevention is better than cure: Machine learning approach to conflict prediction in sub-Saharan Africa. *Sustainability*, *13*(13). https://doi.org/10.3390/su13137366

[5] Rydén, O., de Fine Licht, K., Rönnerstrand, B., Harring, N., Brülde, B., & Jagers, S. C. (2024). Exploring the measurement of political trust: A multilevel observational analysis of six Swedish public agencies. *Social Sciences & Humanities Open*, *10*, 100885. https://doi.org/10.2139/ssrn.4316518

[6] Devine, D. (2025). Political trust and redistribution preferences. *Journal of European Public Policy*, *32*(10), 2439–2462. https://doi.org/10.1080/13501763.2024.2413194

[7] Ohnishi, K. (2024). Compellence by denial against armed groups: UN peacekeeping in Ituri, the Democratic Republic of the Congo. *European Journal of International Security*, *9*(4), 631–649. https://doi.org/10.1017/eis.2024.26

[8] Huh, C. U., & Park, H. W. (2024). Setting the public sentiment: Examining the relationship between social media and news sentiments. *Systems*, *12*(3), 105. https://doi.org/10.3390/systems12030105

[9] Muzalia Kihangu, G., Bahati, A., Batumike, E., & Bisimwa, S. (2022). Ni paix, ni guerre? La persistance des conflits et de l'insécurité dans le territoire de Kalehe au Sud-Kivu.

[10] Héritier Nsenge, M., Lucy Waruguru, M., & Simon Nyaga, M. (2023). Exploratory factor analysis of Congolese information technology graduates' employability: Towards sustainable employment. *Sage Open*, *13*(4), 21582440231210109. https://doi.org/10.1177/21582440231210109

[11] Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

[12] Pinto, A. S., Abreu, A., Costa, E., & Paiva, J. (2023). How machine learning (ML) is transforming higher education: A systematic literature review. *Journal of Information Systems Engineering & Management*, *8*(2): 21168.

[13] Mutsotsya, S. P., Mpia, H. N., Nzanzu, M. M., Baelani, I. N., & Kasolene, M. K. (2024). Prédiction des notes finales des étudiants en fin du premier cycle: Utilisation de Data Mining éducatif. *Etincelle*, *25*(2). https://www.ishango-uac.net/articledetail.php?id=10.61532/rime252114#

[14] Blanco, V., Japón, A., & Puerto, J. (2022). A mathematical programming approach to SVM-based classification with label noise. *Computers & Industrial Engineering*, *172*, 108611. https://doi.org/10.1016/j.cie.2022.108611

[15] Starbuck, C. (2023). *Logistic regression. In: The fundamentals of people analytics*. Cham: Springer. https://doi.org/10.1007/978-3-031-28674-2_12

[16] Rymarczyk, T., Kozłowski, E., Kłosowski, G., & Niderla, K. (2019). Logistic regression for machine learning in process tomography. *Sensors*, *19*(15), 3400. https://doi.org/10.3390/s19153400

[17] Kulikov, A., Loskutov, A., Bezdushniy, D., & Petrov, I. (2023). Decision tree models and machine learning algorithms in the fault recognition on power lines with branches. *Energies*, *16*(14), 5563. https://doi.org/10.3390/en16145563

[18] Murphy, R. (2016). UN peacekeeping in the Democratic Republic of the Congo and the protection of civilians. *Journal of Conflict and Security Law*, *21*(2), 209–246. https://doi.org/10.1093/jcsl/krv030

[19] Zhang, T., Qi, X., He, Q., Hee, J., Takesue, R., Yan, Y., & Tang, K. (2021). The effects of conflicts and self-reported insecurity on maternal healthcare utilisation and children health outcomes in the Democratic Republic of Congo (DRC). In *Healthcare*, *9*(7), 842. https://doi.org/10.3390/healthcare9070842

[20] Wadhwani, G. K., Varshney, P. K., Gupta, A., & Kumar, S. (2023). Sentiment analysis and comprehensive evaluation of supervised machine learning models using Twitter data on Russia–Ukraine war. *SN Computer Science*, *4*(4), 346. https://doi.org/10.1007/s42979-023-01790-5

[21] Aslan, S. (2023). A deep learning-based sentiment analysis approach (MF-CNN-BILSTM) and topic modeling of tweets related to the Ukraine–Russia conflict. *Applied Soft Computing*, *143*, 110404. https://doi.org/10.1016/j.asoc.2023.110404

[22] Ouyang, B., Song, Y., Li, Y., Wu, F., Yu, H., Wang, Y., . . . , & Bauchy, M. (2021). Using machine learning to predict concrete's strength: Learning from small datasets. *Engineering Research Express*, *3*(1), 015022. https://doi.org/10.1088/2631-8695/abe344

[23] Ab Talib, M. S., & Zulfakar, M. H. (2024). Sustainable halal food supply chain management in a small rentier halal market. *Arab Gulf Journal of Scientific Research*, *42*(3), 449–463. https://doi.org/10.1108/AGJSR-11-2022-0251

[24] Pilcher, N., & Cortazzi, M. (2024). 'Qualitative'and' quantitative'methods and approaches across subject fields: Implications for research values, assumptions, and practices. *Quality & Quantity*, *58*(3), 2357–2387. https://doi.org/10.1007/s11135-023-01734-4

[25] Lo, F. Y., Rey-Martí, A., & Botella-Carrubi, D. (2020). Research methods in business: Quantitative and qualitative comparative analysis. *Journal of Business Research*, *115*, 221–224. https://doi.org/10.1016/j.jbusres.2020.05.003

[26] Young, B., Ingwersen, W. W., Bergmann, M., Hernandez-Betancur, J. D., Ghosh, T., Bell, E., & Cashman, S. (2022).

A system for standardizing and combining us environmental protection agency emissions and waste inventory data. *Applied Sciences*, *12*(7), 3447. https://doi.org/10.3390/app12073447

[27] Persson, I., & Khojasteh, J. (2021). Python packages for exploratory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(6), 983–988. https://doi.org/10.1080/10705511.2021.1910037

[28] Alavi, M., Visentin, D.C., Thapa, D.K., Hunt, G.E., Watson, R, & Cleary, M. (2020)). Exploratory factor analysis and principal component analysis in clinical studies: Which one should you use? *Journal Advanced Nursing*, *76*(8), 1886–1889. https://doi.org/10.1111/jan.14377

[29] Orozco-Arias, S., Piña, J. S., Tabares-Soto, R., Castillo-Ossa, L. F., Guyot, R., & Isaza, G. (2020). Measuring performance metrics of machine learning algorithms for detecting and classifying transposable elements. *Processes*, *8*(6), 638. https://doi.org/10.3390/pr8060638

[30] Jierula, A., Wang, S., Oh, T. M., & Wang, P. (2021). Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data. *Applied Sciences*, *11*(5), 2314. https://doi.org/10.3390/app11052314

[31] Hossan, D., Dato'Mansor, Z., & Jaharuddin, N. S. (2023). Research population and sampling in quantitative study. *International Journal of Business and Technopreneurship (IJBT)*, *13*(3), 209–222. https://doi.org/10.58915/ijbt.v13i3.263

[32] Kowal, M. (2024). Translation practices in cross-cultural social research and guidelines for the most popular approach: Back-translation. *Anthropological Review*, *87*(3), 19–32. https://doi.org/10.18778/1898-6773.87.3.02

[33] Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, *11*(4), 193. https://doi.org/10.3390/info11040193

[34] Lumumba, V. W., Kiprotich, D., Lemasulani Mpaine, M., Grace Makena, N., & Daniel Kavita, M. (2024). Comparative analysis of cross-validation techniques: LOOCV, K-folds cross-validation, and repeated K-folds cross-validation in machine learning models. *American Journal of Theoretical and Applied Statistics*, *13*(5). https://doi.org/10.11648/j.ajtas.20241305.13

[35] Kyriazos, T., & Poga, M. (2024). Application of machine learning models in social sciences: Managing nonlinear relationships. *Encyclopedia*, *4*(4), 1790–1805. https://doi.org/10.3390/encyclopedia4040118

[36] Beetham, D. (2013). *The legitimation of power*. UK: Bloomsbury Publishing. https://doi.org/10.1007/978-1-349-21599-7

[37] Von Billerbeck, S. B. K. (2016). *Whose peace?: Local ownership and United Nations peacekeeping*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198755708.001.0001

[38] Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.8cd550d1

[39] Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys (CSUR)*, *55*(3), 1–44. https://doi.org/10.1145/3494672

How to Cite: Mpia, H. N., & Kasambya, M. K. (2026). Predictive Modeling of Public Sentiment Toward MONUSCO's Mandate: A Machine Learning Approach. *Archives of Advanced Engineering Science*. https://doi.org/10.47852/bonviewAAES62027763