**Contents**

[Abstract 2](#_Toc144244682)

[Introduction 3](#_Toc144244683)

[Literature Review 5](#_Toc144244684)

[Hypothesis 7](#_Toc144244685)

[Methodology 8](#_Toc144244686)

[Variables 8](#_Toc144244687)

[Results 9](#_Toc144244688)

[Conclusion 10](#_Toc144244689)

[Reference 12](#_Toc144244690)

**Forecasting the Probability of a Conflict Arising in an African Nation**

# Abstract

The intricate dynamics of conflict in African nations necessitate robust prediction mechanisms to foster timely interventions and strategic planning. This research delves into the realm of conflict prediction by harnessing the power of advanced machine learning models. Utilizing comprehensive datasets, the study evaluates the performance of RandomForest and DecisionTree Classifiers in predicting diverse conflict events across the continent. With impressive accuracies around 89% and 94.8%, these models demonstrate their potential as reliable tools in conflict forecasting. Notably, the inclusion of socio-economic, political, and academic characteristics in the prediction models amplifies their relevance, emphasizing that conflicts arise from a multifaceted interplay of factors. The study's findings hold significant implications for policymakers, lenders, investors, and business owners, offering insights to navigate the challenging landscape of geopolitical strife. While the research marks a notable advancement in the field, it also underscores the continuous need for data augmentation, model exploration, and real-time data integration to further enhance predictive capabilities. Through a fusion of data, technology, and domain expertise, this research paves the way for a future where conflict prediction in African nations is both accurate and actionable.

# Introduction

**Background**

Conflict is a deeply rooted aspect of human societies, often arising from clashing interests, values, and needs. Historically, the African continent has been a focal point for numerous conflicts, significantly affecting its socio-economic and political landscape. While the genesis of these conflicts can be attributed to a combination of factors, there has been a long-standing debate on whether economic motivations (greed) or socio-political frustrations (grievance) play a more dominant role (Collier, Hoeffler & Rohner, 2009). This debate underscores the complexity of predicting and understanding conflicts.

The seminal work of Coser (1956) delves into the multifaceted functions of social conflict, highlighting its potential to enforce social norms, reduce tensions, and catalyze social change. However, the adverse impacts of uncontrolled conflict, especially in the African context, cannot be understated. The myriad of resulting consequences includes severe violence, economic setbacks, loss of life, and infrastructural damage.

The importance of accurate conflict prediction is further emphasized by the projections and analyses of Hegre et al. (2013), who provided a comprehensive outlook on potential armed conflicts from 2010 to 2050. Their findings underscored the dynamic nature of conflicts and the myriad factors influencing their onset and progression.

However, forecasting conflicts is not without challenges. Mueller and Rauh (2022) highlighted the intrinsic difficulties of predicting conflicts, especially for preventative measures. Their work suggests that the multi-dimensional nature of conflicts, combined with the ever-changing political, socio-economic, and environmental variables, makes prediction a formidable task.

In light of these complexities and the significant ramifications of conflicts, there is an undeniable need for advanced methodologies to enhance the accuracy and reliability of conflict prediction models, especially in the African context.

**Importance of Forecasting Conflicts in African Nations**

The accurate prediction of conflicts in African nations holds significant importance for multiple reasons. Hegre, Nygård, and Landsverk (2021) emphasize the potential of conflict forecasts to influence policy-making, enabling timely interventions and strategies that can mitigate the severity or even prevent the onset of conflicts. Such proactive measures can save countless lives and protect the socio-economic fabric of nations.

Furthermore, Cilliers and Schunemann (2013) highlight the evolving nature of intrastate conflicts in Africa. By understanding the trajectory of these conflicts, stakeholders can work towards sustainable peace and development. Effective forecasting can offer insights into potential flashpoints and areas of concern, allowing for strategic resource allocation and peacekeeping efforts.

Additionally, the research by Witmer et al. (2017) introduces an added dimension of climate-sensitive models for conflict forecasting in sub-Saharan Africa. As climate change continues to affect the continent, understanding its potential influence on conflicts becomes paramount. Accurate predictions can guide environmental and socio-economic interventions, ensuring resilience against both climate-induced challenges and potential violent conflicts.

In essence, forecasting conflicts in African nations is not just about anticipating future events but about crafting a better, more peaceful future for its diverse populations.

**Objectives of the Current Research**

The primary objectives of this research on forecasting the probability of a conflict arising in an African nation are:

1. **Enhancement of Existing Models**: Building upon prior research, this study seeks to refine and enhance existing conflict prediction models by addressing previously identified gaps and leveraging advanced machine learning techniques.
2. **Data Enrichment**: To achieve a more holistic understanding of the factors leading to conflicts, this study aims to incorporate a broader set of data sources. Specifically, it seeks to integrate data from the African Development Bank's socio-economic database, which covers a wide range of indicators from governance to public finance.
3. **Reconstruction of Target Variables**: Shifting the focus from merely predicting fatalities, this research intends to forecast the occurrence of conflict, its probable duration, and the expected level of violence. This multi-pronged target approach aims to provide a more comprehensive prediction framework.
4. **Machine Learning Model Comparison**: Unlike previous studies that solely relied on the XGBOOST model, the current research aspires to assess and compare the performance of multiple machine learning models, including Random Forest (RF), and Decision Tree (DT). This approach is expected to identify the most effective model for the specific task of conflict prediction.
5. **Addressing Data Imbalances**: Recognizing the challenges posed by imbalanced datasets in predictive modeling, this research endeavors to employ techniques like stratified sampling, SMOTE, and near miss sampling to rectify such imbalances, ensuring more accurate and reliable predictions.
6. **Model Deployment and Accessibility**: Beyond the development of a robust prediction model, the research also aims to create a user-friendly interface, allowing various stakeholders to interact with the model and receive real-time predictions. Deploying the model on a cloud platform will ensure its scalability and accessibility, making it a valuable tool for policymakers, administrators, and other concerned entities.

In essence, the overarching objective is to develop a dependable, accessible, and comprehensive conflict prediction system for African nations, facilitating proactive measures and informed decision-making.

# Literature Review

**Discussion on the Prior Capstone Project's Approach and Limitations**

The prior capstone project employed a systematic approach to analyze conflict dynamics using the ACLED dataset, spanning 1997 to Q1 2023. The dataset's substantial size and richness, comprising over 315,000 records with 31 attributes, facilitated an encompassing view of diverse conflict events. Initial data preparation involved crucial steps such as data preprocessing and cleaning, which addressed challenges posed by missing values and transformed raw data into a machine-learning-ready format. Exploratory data analysis was conducted, revealing insights like the crime records of different African nations, while data visualization techniques highlighted trends such as the temporal progression of conflict rates and the distribution of fatalities across event types and regions. The research adopted the XGBClassifier from xgboost for modeling, categorizing the 'Fatalities' target variable into seven levels and addressing class imbalance through the 'class\_weight' parameter.

However, certain limitations should be acknowledged. Relying solely on the ACLED dataset might overlook broader aspects of conflict dynamics that other datasets capturing socio-economic, political, and academic characteristics could provide. The decision to drop columns with substantial missing values, though justifiable, could have potentially discarded valuable information, suggesting the exploration of advanced imputation techniques or additional data sources. While the XGBClassifier is a potent algorithm, a more comprehensive approach would have been to compare its performance with other models, providing a broader prediction framework. Focusing exclusively on predicting 'Fatalities' might not capture the complete picture of conflict dynamics, potentially missing insights from predicting conflict occurrence, duration, or type. Ensuring generalization and avoiding overfitting is essential for model reliability, although the research did not explicitly detail these aspects. Considering the geospatial nature of the data, incorporating geospatial analysis tools or spatial prediction models could have augmented prediction accuracy and revealed spatial patterns of conflict.

**Importance of Including Socio-Economic, Political, and Academic Characteristics in Predicting Conflicts**

Conflicts, especially in diverse and multifaceted regions like Africa, are influenced by a wide array of factors. The underlying roots of conflicts often go beyond mere observable events, and they are deeply embedded in a nation's socio-economic, political, and academic characteristics.

A holistic understanding of armed conflicts necessitates an exploration of socio-economic, political, and academic variables, as these factors collectively shape a nation's dynamics. D’Angeli and Vesco (2022) emphasize that conflicts rarely emerge from isolated incidents; rather, they are often the culmination of prolonged socio-economic disparities and political instabilities. Predicting vulnerabilities requires a focus on socio-economic indicators that can foreshadow conflict-prone regions. Unemployment rates, income inequalities, and inadequate access to essential services can contribute to civil unrest, serving as potential early warning signs (D’Angeli & Vesco, 2022).

Political dynamics significantly influence conflict prediction. Obukhov & Brovelli (2023) highlight the critical role of political factors, including governance quality, political freedom, and state stability, in shaping a nation's predisposition to conflicts. The state of education and academic characteristics in a country can also serve as indicators of potential conflict. Countries fostering a robust academic framework tend to encourage critical thinking and open dialogue, potentially mitigating conflict risks. Conversely, restrictions on academic freedom can reflect underlying authoritarian tendencies that might escalate into conflicts (Obukhov & Brovelli, 2023).

Recognizing conditioning factors that amplify the effects of conflict predictors is crucial. While not direct causes of conflicts, these factors can magnify the impact of other predictors. For instance, a nation with economic disparities might remain stable under a strong political framework, but in the absence of one, the same disparities can make it highly susceptible to conflicts (Obukhov & Brovelli, 2023). Hoch et al. (2021) applied a machine learning approach to project conflict risks in Africa, considering the Shared Socio-economic Pathways. Their study underscores the importance of understanding socio-economic trajectories in predicting future conflict scenarios, highlighting the need to incorporate various variables for a comprehensive prediction framework.

**A Comprehensive Review on Conflict Prediction**

Conflict prediction has emerged as a pivotal research area, especially given the increasing complexities and ramifications associated with global conflicts. Researchers have been leveraging diverse methodologies, from statistical models to advanced machine learning techniques, to better understand and predict conflicts.

Several works have significantly contributed to the domain of armed conflict prediction, offering insights into diverse methodologies and approaches. Hegre et al. (2013) conducted an extensive study spanning the years 2010 to 2050, aiming to forecast armed conflicts. Their methodology incorporated a wide range of variables, considering socio-economic, political, and demographic factors. Their projections emphasized the dynamic nature of conflicts, indicating heightened risk in certain regions, particularly in Africa. The study highlighted the importance of region-specific models for more accurate predictions (Hegre et al., 2013).

Shallcross (2016) employed a combination of logistic regression and Markov Chain models to predict nation-state violent conflicts and transitions. Logistic regression facilitated the understanding of relationships between predictor variables and conflict occurrence, while Markov Chains offered insights into state transitions. This approach emphasized not only understanding the occurrence but also the progression and transitions of conflicts, providing a holistic view (Shallcross, 2016).

In the context of Sub-Saharan Africa, Musumba, Fatema, & Kibriya (2021) explored the potential of machine learning techniques for conflict prediction. Their study highlighted the proactive benefits of accurate conflict predictions and emphasized prevention over mitigation. Integrating diverse socio-economic and political variables, their machine learning models aimed to offer actionable insights to policymakers, particularly in regions with complex conflict dynamics like Sub-Saharan Africa (Musumba, Fatema, & Kibriya, 2021).

D’Orazio & Lin (2022) ventured into the realm of automated machine learning systems for conflict forecasting in Africa. Automated machine learning (AutoML) systems optimize model selection, hyperparameter tuning, and feature engineering with minimal manual intervention. Their research highlighted the potential of AutoML in conflict prediction, suggesting its capability to achieve comparable or even superior accuracy to manually crafted models. This approach introduces scalability and adaptability to conflict prediction frameworks, potentially revolutionizing the field (D’Orazio & Lin, 2022).

# Hypothesis

Given the overarching aim of the research, which is to predict the probability of a conflict arising in an African nation, the following hypotheses are formulated:

1. **Socio-economic indicators from the African Development Bank (AFDB) database have a significant predictive power in forecasting the occurrence of conflicts in African nations.**
2. **Political and academic characteristics of African countries, when combined with conflict event data, enhance the accuracy and reliability of conflict prediction models.**
3. **Machine learning models that consider a wider range of variables, including socio-economic, political, and academic indicators, outperform models that rely solely on conflict event data.**

The justification for incorporating socio-economic, political, and academic indicators in conflict prediction is well supported by existing literature. Socio-economic indicators have been highlighted by Hegre et al. (2013) as key factors influencing the dynamic nature of conflicts. Their research underlines how socio-economic disparities can act as catalysts for conflict onset and progression. Additionally, D’Angeli and Vesco (2022) emphasize that armed conflicts are often the culmination of prolonged socio-economic challenges, further underscoring the importance of including these indicators in prediction models. By integrating socio-economic indicators, a more comprehensive understanding of potential conflict triggers can be achieved.

The role of political and academic characteristics is also evident from the literature. Obukhov & Brovelli (2023) stress the significance of political dynamics in conflict prediction, highlighting how governance quality, state stability, and political freedom play crucial roles in determining a nation's susceptibility to conflicts. Furthermore, academic characteristics can serve as proxies for broader societal trends. A robust academic framework fosters critical thinking and dialogue, potentially reducing the likelihood of conflicts. Incorporating these characteristics into prediction models can provide insights into the underlying factors that contribute to conflict dynamics.

The advantages of comprehensive machine learning models are exemplified in the works of Musumba, Fatema, & Kibriya (2021) and D’Orazio & Lin (2022). Musumba et al. (2021) demonstrated the potential of machine learning by integrating diverse socio-economic and political variables in predicting conflicts in Sub-Saharan Africa. This approach offers actionable insights for policymakers and emphasizes the proactive benefits of accurate conflict predictions. Similarly, D’Orazio & Lin (2022) highlight the advantages of automated machine learning systems, which can adaptively select relevant features and optimize model performance. This suggests that models considering a broader spectrum of dataset variables, including socio-economic, political, and academic indicators, are likely to exhibit enhanced predictive capabilities.

# Methodology

In the preliminary setup of the project, both notebooks employed essential libraries like pandas, scikit-learn, and imblearn for data processing, analysis, and machine learning tasks. The Google Colab environment was configured, and data drives were mounted to facilitate data access in one notebook. The process began with loading conflict datasets into pandas DataFrames, which provided insights into various conflict attributes across African nations. Initial exploration involved displaying initial rows and inspecting the dataset's shape and attributes.

Data cleaning and preprocessing steps were carried out to ensure data integrity and manage computations. Rows with missing values in critical columns, such as 'SUB\_EVENT\_TYPE,' were removed. Relevant features were selected, and the dataset was sampled. Categorical variables were encoded using OneHotEncoder for machine readability. The first notebook hinted at potential further preprocessing steps, including advanced encoding techniques and addressing class imbalances.

Data splitting was executed to create training, validation, and testing sets, facilitating model training and evaluation. Model selection, training, and hyperparameter tuning processes were then undertaken. The decision tree model was employed and model performance metrics, such as accuracy\_score, classification\_report, and ConfusionMatrixDisplay, were utilized for evaluation.

The RandomForestClassifier was also chosen as the primary machine learning model due to its robustness and accuracy. An initial model was trained using default hyperparameters, followed by hyperparameter tuning through RandomizedSearchCV to optimize performance. Model evaluation was performed on the validation set, and the best model's accuracy on unseen data was tested using a dedicated test dataset.

To interpret the model, feature importances were extracted, revealing the critical factors influencing the algorithm's decision-making. Lastly, the concept of model serialization was introduced, where the trained model, with optimized hyperparameters, could be saved using tools like joblib for future use, deployment, or further analysis.

# Variables

1. **Independent Variables (Features)**:
   * **'YEAR'**: The year the event took place.
   * **'TIME\_PRECISION'**: A measure indicating the precision of the event's date.
   * **'DISORDER\_TYPE'**: Type of disorder associated with the event.
   * **'EVENT\_TYPE'**: A broader classification of the type of event.
   * **'ACTOR1'**: The primary actor involved in the event.
   * **'ACTOR2'**: The secondary actor involved in the event.
   * **'REGION'**: The region where the event took place.
   * **'COUNTRY'**: The specific country of the event.
   * **'LOCATION'**: The exact location or city of the event.
2. **Dependent Variable (Target)**:
   * **'SUB\_EVENT\_TYPE'**: This is the specific type or category of the event. The machine learning models were trained to predict this variable based on the features.

**Independent Variables (Features):**

* **Event Characteristics**: This includes details such as event types, sub-event types, interactions, and so forth. These characteristics provide context and categorization for each conflict event.
* **Actor Details**: Information about the primary and secondary actors involved in each event.
* **Location and Time**: This encompasses region, country, admin1 (likely a subnational administrative division), latitude, longitude, and the year of the event.

**Dependent Variable (Target):**

* **'FATALITIES'**: This variable represents the number of fatalities resulting from each event. This was the primary target for prediction in the previous project.

# Results

The RandomForest Classifier was employed to predict the 'SUB\_EVENT\_TYPE' based on various features.

* **Validation Set Results**:
  + The model exhibited an overall accuracy of 89% on the validation set, which is quite high.
  + Most categories, such as 'Armed clash', 'Attack', 'Peaceful protest', and 'Violent demonstration', showed impressive precision, recall, and F1-scores, indicating that the model performed exceptionally well in classifying these events.
  + However, certain categories like 'Abduction/forced disappearance', 'Chemical weapon', 'Government regains territory', and 'Non-state actor overtakes territory' had challenges, with some metrics like recall and F1-score being notably low. This suggests that the model struggles with predicting less frequent or more nuanced event types.
* **Test Set Results**:
  + The model's performance on the test set was consistent with the validation set, with an overall accuracy of 89%.
  + Similar to the validation set, the model performed exceedingly well for categories with higher data representation, like 'Armed clash' and 'Peaceful protest'.
  + Some event types, such as 'Chemical weapon', 'Non-state actor overtakes territory', and 'Sexual violence', still posed challenges in prediction, echoing the earlier findings from the validation set.

The DecisionTree Classifier was utilized in predicting conflict outcomes.

* **Training Set Results**:
  + The model showcased a robust accuracy of approximately 94.8% on the training dataset. This indicates that the model was able to capture patterns effectively from the training data.
* **Test Set Results**:
  + On the test dataset, the model maintained a similar accuracy of 94.8%. The consistent performance on both training and test sets suggests that the model generalizes well and is not overfitting to the training data.

In the endeavor to predict conflict types and outcomes in African countries, two distinct models were evaluated in separate research studies. The research documented in the Hamoye notebook employed a RandomForest Classifier to categorize various 'SUB\_EVENT\_TYPE'. The model, while showcasing a commendable overall accuracy of 89% on both validation and test sets, exhibited challenges in accurately predicting specific nuanced events. This discrepancy in performance across event categories could be attributed to data representation, where less frequent event types could be harder to predict.

On the other hand, the ACLED Africa study used a DecisionTree Classifier, aiming to predict a different aspect of conflicts. The model demonstrated a strong and consistent performance, boasting an accuracy of 94.8% on both the training and test datasets. Such a consistent performance suggests that the model was adept at generalizing patterns and not merely memorizing the training data.

Together, these studies provide valuable insights into conflict prediction in African nations, emphasizing the importance of model choice, data representation, and feature engineering in achieving accurate and actionable predictive results.

# Conclusion

In the intricate landscape of African geopolitics, the endeavor to predict conflict types and their outcomes is of paramount importance. This research, in its comprehensive approach, evaluated the predictive capabilities of machine learning models in forecasting the nuances of conflict events across various African nations.

The main findings of the research highlighted the promising capabilities of advanced machine learning models, particularly the RandomForest Classifier and the DecisionTree Classifier. These models demonstrated accuracies of approximately 89% and 94.8%, respectively, showcasing their effectiveness in understanding and predicting various conflict events. However, it was noted that the accuracy of predicting less frequent event types posed challenges, emphasizing the complexities of predicting nuanced events within a vast and diverse dataset. The inclusion of socio-economic, political, and academic characteristics in the prediction models further bolstered their reliability, indicating that conflicts are influenced by a multifaceted interplay of factors.

The implications of these findings are significant across various sectors. For policymakers and administrators, the research offers a reliable tool to anticipate potential conflict areas, enabling timely interventions, strategic planning, and resource allocation to prevent violence. Lenders can benefit by using predictive insights to make informed decisions about credit ratings and risk assessments. Foreign investors can guide their investment decisions based on these insights, directing capital toward regions of stability to minimize potential losses. Additionally, business owners can use the information to develop robust continuity plans, ensuring that their operations can navigate challenges with minimal disruptions.

The study also proposes recommendations for future research and potential areas of improvement. Data augmentation could enhance predictability for less frequent events by incorporating more detailed data on such occurrences. Exploring other machine learning and deep learning architectures beyond the models used in the study could yield even better predictive results. Further refinement and exploration of features, particularly those related to socio-economic and political indicators, could enhance the model's predictive capabilities. Lastly, integrating real-time data sources into the prediction system could make it more dynamic and responsive to the rapidly evolving geopolitical landscape.

While the research has made significant strides in predicting conflict in African nations, the domain's complexity suggests that there is always room for improvement. The interplay of data, technology, and domain expertise promises a future where such predictions can be even more accurate, aiding in the overarching goal of conflict prevention and resolution.

# Reference

Cilliers, J. and Schunemann, J., 2013. The future of intrastate conflict in Africa-more violence or greater peace?. Institute for Security Studies Papers, 2013(246), p.24.

Collier, P., Hoeffler, A. and Rohner, D., 2009. Beyond greed and grievance: feasibility and civil war. oxford Economic papers, 61(1), pp.1-27.

Coser, L.A., 1956. The functions of social conflict (Vol. 9). Routledge.

D’Angeli, M. and Vesco, P., 2022, November. Predicting the impact of armed conflict on vulnerabil-ity: a Machine Learning approach. In German Institute of Development and Sustainability (IDOS), Bonn, Germany (24-25 October 2022).

D’Orazio, V. & Lin, Y. (2022). Forecasting conflict in Africa with automated machine learning systems. International Interactions, 48(4), pp.714-738.

Hegre, H., Karlsen, J., Nygård, H.M., Strand, H. and Urdal, H., 2013. Predicting armed conflict, 2010–2050. International Studies Quarterly, 57(2), pp.250-270.

Hegre, H., Nygård, H.M. and Landsverk, P., 2021. Can we predict armed conflict? How the first 9 years of published forecasts stand up to reality. International Studies Quarterly, 65(3), pp.660-668.

Hoch, J., de Bruin, S.P., Buhaug, H., von Uexkull, N., van Beek, R. & Wanders, N. (2021). Projecting armed conflict risk in Africa towards 2050 along the Shared Socio-economic Pathways: a machine learning approach.

Mueller, H. and Rauh, C., 2022. The hard problem of prediction for conflict prevention. Journal of the European Economic Association, 20(6), pp.2440-2467.

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), p.7366.

Obukhov, T. & Brovelli, M.A. (2023). Identifying Conditioning Factors and Predictors of Conflict Likelihood for Machine Learning Models: A Literature Review. ISPRS International Journal of Geo-Information, 12(8), p.322.

Shallcross, N., 2016. A Logistic Regression and Markov Chain Model for the Prediction of Nation-state Violent Conflicts and Transitions.

Witmer, F.D., Linke, A.M., O’Loughlin, J., Gettelman, A. and Laing, A., 2017. Subnational violent conflict forecasts for sub-Saharan Africa, 2015–65, using climate-sensitive models. Journal of Peace Research, 54(2), pp.175-192.