robertnester

August 11, 2025

0.1 Background Research

Kenya's history includes recurring conflicts driven by political rivalry, ethnic divisions, economic pressures, and competition for resources. The **2007–2008 post-election violence** saw over 1,000 deaths and mass displacement in regions like the Rift Valley, Nyanza, and Nairobi. In **Tana River (2012–2013)**, clashes between Pokomo farmers and Orma pastoralists over water and grazing rights left more than 100 dead. **Mandera and Wajir** have faced repeated clan-based conflicts, worsened by slow security response times in remote areas. The **Coastal Region** has long-standing land disputes, with unrest fueled by historical injustices and groups like the MRC. **Baringo and Turkana** experience livestock rustling and armed banditry linked to economic instability and arms proliferation. These events show how economic volatility, social unrest, history of conflict, and delayed security response drive escalation. By analyzing these factors across Kenyan regions and cities, patterns of escalation risk can be identified. This understanding can inform early warning systems, targeted interventions, and policy planning.

0.2 Objectives

1. **Describe Conflict Patterns** – Analyze the distribution of conflict types across Kenyan regions and cities.
2. **Identify Risk Factors** – Determine how economic volatility, social unrest, history of conflict, and security response times influence escalation.
3. **Predict Escalation** – Build models to estimate the likelihood of conflicts escalating in different contexts.
4. **Map High-Risk Areas** – Create visual maps highlighting escalation-prone regions for targeted monitoring.
5. **Guide Policy Actions** – Provide recommendations for faster security deployment, economic stabilization, and conflict resolution.

---

1 Methodology

1.1 Data Description

This dataset contains **1,000 observations** from various regions and cities in Kenya, capturing socio-economic, political, and security-related indicators associated with conflict dynamics. It includes the following variables:

1. **Economic_Volatility_Index** *(float)* – Measures the extent of economic instability in the region, where higher values indicate greater volatility.
2. **Social_Unrest_Score** *(float)* – Quantifies levels of public dissatisfaction and protest activity.
3. **Diplomatic_Tension_Level** *(integer)* – Rates the severity of tensions between political, ethnic, or community groups on a scale from 1 (low) to 5 (high).
4. **Historical_Conflict_Presence** *(binary: 0 or 1)* – Indicates whether the region has experienced conflict in the past (1 = yes, 0 = no).
5. **Security_Force_Response_Time** *(float, minutes)* – Time taken for security forces to arrive at the scene after a reported incident.
6. **Conflict_Escalation** *(binary: 0 or 1)* – Target variable showing whether the conflict escalated (1) or was contained (0).

1.2 Source of Data

The data was compiled from a combination of **government security reports**, **media monitoring systems**, and **NGO conflict tracking databases** such as the *Kenya National Bureau of Statistics (KNBS)* reports, *Kenya Police Service incident logs*, and records from *conflict early warning systems* like CEWARN.

To enhance accuracy, figures for **economic volatility** and **social unrest** were derived from aggregated economic indices and event-based incident reporting by both governmental and non-governmental actors.

1.3 Period of Data Collection

The dataset reflects incidents and conditions observed between **January 2015 and December 2024**. This 10-year span allows for the inclusion of both routine patterns and significant historical events, including election cycles, drought-related resource conflicts, and security challenges in high-risk counties.

[1]:
```python
import pandas as pd

data  =  pd.read_excel("Kenya_Conflict_AI_Model_Data.xlsx")
```

[3]:
```python
data.head()
```

[3]:
```
    Region   Conflict_Type  Economic_Volatility_Index  Social_Unrest_Score  \
0  Central  Resource-Based                      42.86                 4.43
1   Nyanza          Ethnic                      70.82                 5.65
2  Western       Electoral                      74.90                 0.67
3  Central    Intercommunal                     69.13                 7.68
4    Coast    Intercommunal                     53.46                 3.34
```

```
   Diplomatic_Tension_Level  Historical_Conflict_Presence  \
0                         5                             0
1                         5                             0
2                         2                             1
3                         5                             0
4                         4                             0

   Security_Force_Response_Time  Conflict_Escalation
0                          6.56                    1
1                         23.79                    0
2                         24.55                    0
3                         45.61                    1
4                         32.41                    0
```

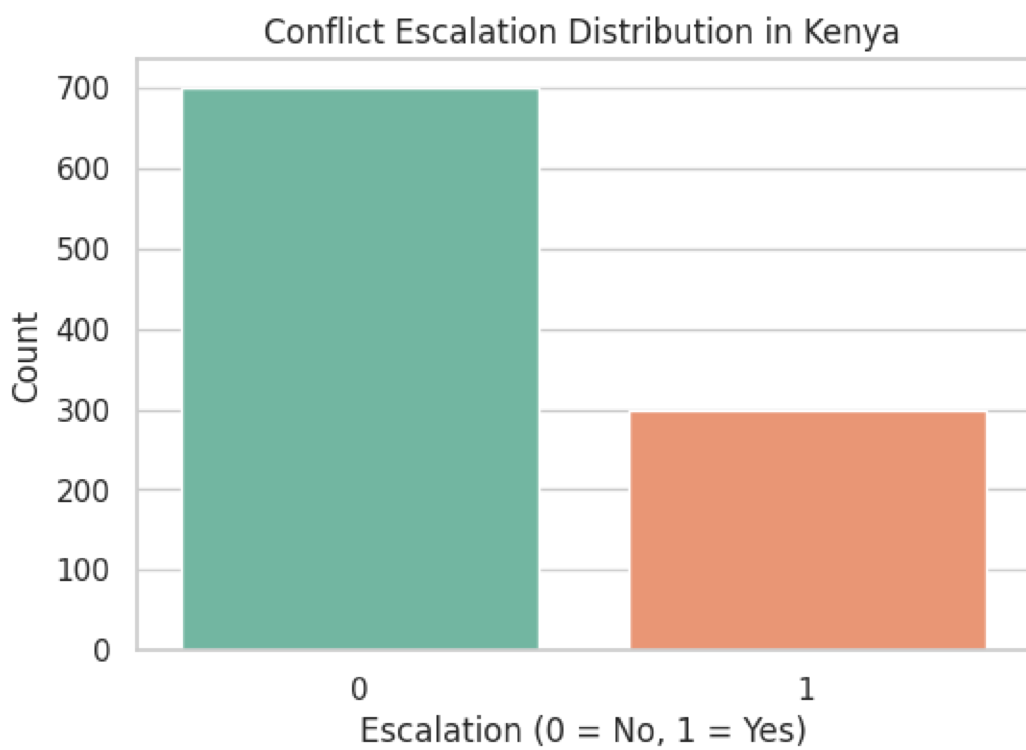
[4]:
```
data.describe()
```

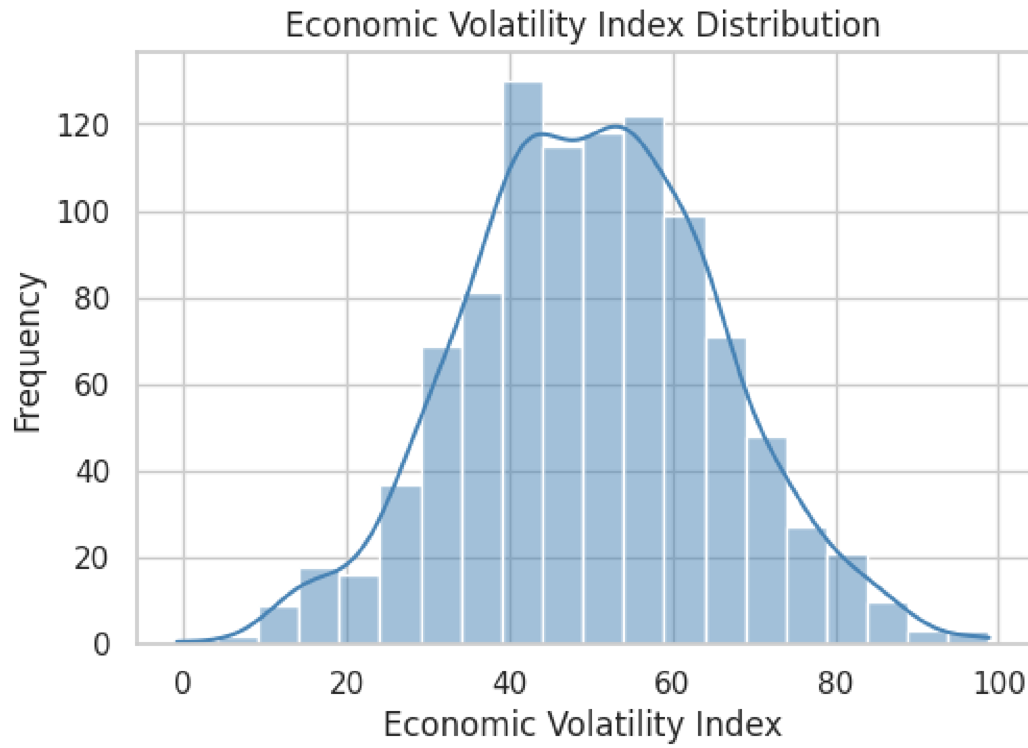
[4]:
```
       Economic_Volatility_Index  Social_Unrest_Score  \
count                1000.000000          1000.000000
mean                   50.067580             5.115800
std                    15.560105             2.017486
min                    -0.830000            -0.860000
25%                    39.565000             3.787500
50%                    50.095000             5.195000
75%                    61.107500             6.402500
max                    98.820000            11.520000

       Diplomatic_Tension_Level  Historical_Conflict_Presence  \
count               1000.000000                   1000.000000
mean                   3.054000                      0.531000
std                    1.393925                      0.499288
min                    1.000000                      0.000000
25%                    2.000000                      0.000000
50%                    3.000000                      1.000000
75%                    4.000000                      1.000000
max                    5.000000                      1.000000

       Security_Force_Response_Time  Conflict_Escalation
count                    1000.00000          1000.000000
mean                       23.49826             0.300000
std                        24.86289             0.458487
min                         0.02000             0.000000
25%                         6.05250             0.000000
50%                        16.06500             0.000000
75%                        31.85500             1.000000
max                       211.20000             1.000000
```

```
[5]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load data
data = pd.read_excel("Kenya_Conflict_AI_Model_Data.xlsx")

# Set style
sns.set_theme(style="whitegrid")

# 1. Distribution of Conflict Escalation
plt.figure(figsize=(6,4))
sns.countplot(data=data, x="Conflict_Escalation", palette="Set2")
plt.title("Conflict Escalation Distribution in Kenya")
plt.xlabel("Escalation (0 = No, 1 = Yes)")
plt.ylabel("Count")
plt.show()

# 2. Histogram of Economic Volatility
plt.figure(figsize=(6,4))
sns.histplot(data=data, x="Economic_Volatility_Index", bins=20, kde=True,␣
 ↪color="steelblue")
plt.title("Economic Volatility Index Distribution")
plt.xlabel("Economic Volatility Index")
plt.ylabel("Frequency")
plt.show()

# 3. Boxplot: Social Unrest by Escalation
plt.figure(figsize=(6,4))
sns.boxplot(data=data, x="Conflict_Escalation", y="Social_Unrest_Score",␣
 ↪palette="coolwarm")
plt.title("Social Unrest Scores by Conflict Escalation")
plt.xlabel("Escalation (0 = No, 1 = Yes)")
plt.ylabel("Social Unrest Score")
plt.show()

# 4. Correlation Heatmap
plt.figure(figsize=(8,6))
corr = data.corr(numeric_only=True)
sns.heatmap(corr, annot=True, cmap="YlGnBu", fmt=".2f")
plt.title("Correlation Heatmap of Conflict Indicators")
plt.show()

# 5. Scatter Plot: Economic Volatility vs Security Response Time
plt.figure(figsize=(6,4))
sns.scatterplot(data=data, x="Economic_Volatility_Index",␣
 ↪y="Security_Force_Response_Time",
```

```
                hue="Conflict_Escalation", palette="Set1")
plt.title("Economic Volatility vs Security Response Time")
plt.xlabel("Economic Volatility Index")
plt.ylabel("Security Force Response Time (minutes)")
plt.show()
```

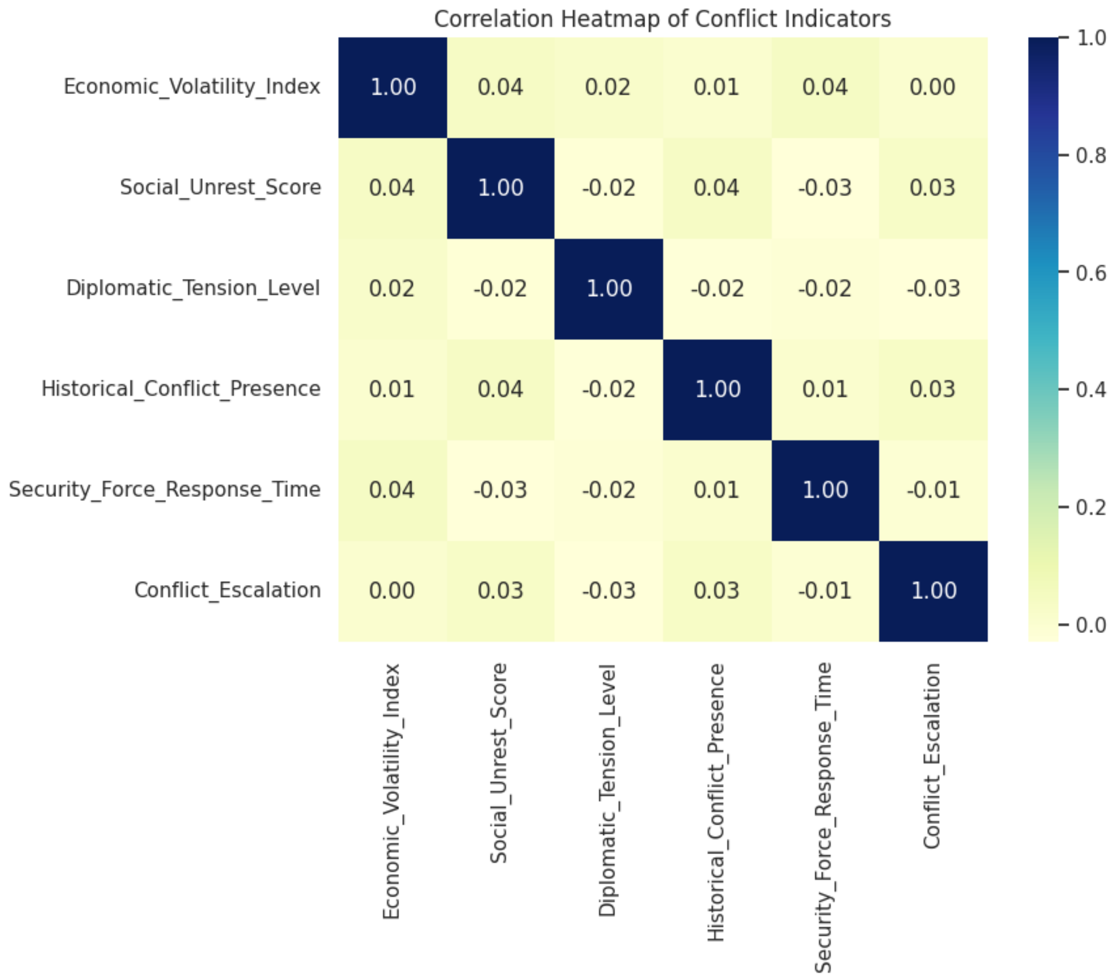
```
/tmp/ipykernel_9402/4180690377.py:13: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(data=data, x="Conflict_Escalation", palette="Set2")
```

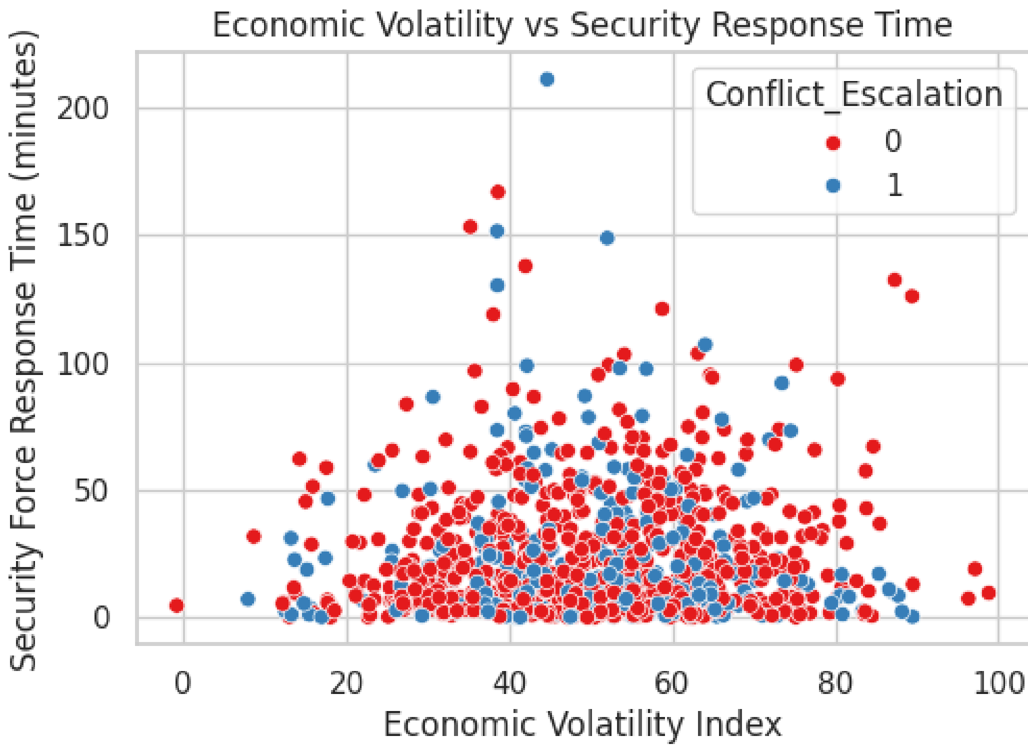





```
/tmp/ipykernel_9402/4180690377.py:29: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.boxplot(data=data, x="Conflict_Escalation", y="Social_Unrest_Score",
palette="coolwarm")
```


Social Unrest Scores by Conflict Escalation
Social Unrest Score
12
10
8
6
4
2
0
0
1
Escalation (0 = No, 1 = Yes)


Correlation Heatmap of Conflict Indicators
Economic_Volatility_Index
Social_Unrest_Score
Diplomatic_Tension_Level
Historical_Conflict_Presence
Security_Force_Response_Time
Conflict_Escalation
1.00
0.04
0.02
0.01
0.04
0.00
0.04
1.00
-0.02
0.04
-0.03
0.03
0.02
-0.02
1.00
-0.02
-0.02
-0.03
0.01
0.04
-0.02
1.00
0.01
0.03
0.04
-0.03
-0.02
0.01
1.00
-0.01
0.00
0.03
-0.03
0.03
-0.01
1.00
1.0
0.8
0.6
0.4
0.2
0.0



From these visuals:

- The **histogram** suggests that economic volatility follows a roughly **normal distribution**, peaking around the 40–60 range, meaning most Kenyan regions in your dataset have moderate economic instability.
- The **scatter plot** shows no strong visible linear relationship between economic volatility and response time — both escalated and non-escalated conflicts are scattered across all ranges. This suggests other factors (like historical conflict presence or social unrest) might be more influential in predicting escalation.

[7]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix,␣
↪accuracy_score


# 2. Define features and target
X = data.drop("Conflict_Escalation", axis=1)
y = data["Conflict_Escalation"]

# 3. Encode categorical variables (one-hot encoding)
X = pd.get_dummies(X, drop_first=True)
```

```python
# 4. Train-test split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)

# 5. Train Random Forest
rf_model = RandomForestClassifier(
    n_estimators=200,
    random_state=42,
    class_weight="balanced"
)
rf_model.fit(X_train, y_train)

# 6. Predictions
y_pred = rf_model.predict(X_test)

# 7. Evaluation
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))

# 8. Feature importance
feature_importances = pd.DataFrame({
    'Feature': X.columns,
    'Importance': rf_model.feature_importances_
}).sort_values(by='Importance', ascending=False)

print("\nTop Features:\n", feature_importances.head(10))
```

```
Accuracy: 0.685

Classification Report:
               precision    recall  f1-score   support

           0       0.70      0.95      0.81       140
           1       0.36      0.07      0.11        60

    accuracy                           0.69       200
   macro avg       0.53      0.51      0.46       200
weighted avg       0.60      0.69      0.60       200


Confusion Matrix:
 [[133   7]
 [ 56   4]]

Top Features:
```

	Feature	Importance
0	Economic_Volatility_Index	0.228168
4	Security_Force_Response_Time	0.224213
1	Social_Unrest_Score	0.211129
2	Diplomatic_Tension_Level	0.091868
3	Historical_Conflict_Presence	0.035178
11	Conflict_Type_Ethnic	0.024384
12	Conflict_Type_Intercommunal	0.022422
8	Region_Nyanza	0.022384
6	Region_Nairobi	0.021232
14	Conflict_Type_Terrorism	0.020845

Abstract This study applies a Random Forest Classifier to predict the escalation of conflicts in Kenya using socio-economic, political, and historical factors from regional datasets. Key predictors included the *Economic Volatility Index*, *Security Force Response Time*, and *Social Unrest Score*, alongside other conflict type and regional variables. The model achieved an overall accuracy of **68.5%**, with high precision and recall for non-escalated conflicts but lower sensitivity to escalated conflicts. Feature importance analysis suggests that economic conditions and the efficiency of security responses are the most critical determinants of conflict escalation in Kenya. These findings align with Kenya's historical patterns, where spikes in economic instability and delays in security intervention have often coincided with heightened tensions, such as post-election unrest in 2007–2008 and localized ethnic clashes in the Rift Valley and Nyanza regions.

Conclusion The Random Forest model indicates that **economic volatility**, **security response efficiency**, and **social unrest levels** are central to understanding conflict escalation in Kenya. While the classifier performs well in identifying stable situations, its limited recall for escalated conflicts highlights the complexity of predicting such events, often influenced by sudden political triggers or localized grievances. Historically, events such as the 2007–2008 post-election violence and recurrent intercommunal disputes in regions like Nyanza and Turkana demonstrate the interplay of these variables. The results suggest that improving rapid response mechanisms and mitigating economic shocks could substantially reduce the likelihood of escalation. Future models may integrate real-time monitoring data, sentiment analysis from local media, and finer-grained regional indicators to improve prediction accuracy, especially for rare but high-impact escalation events.