Datathon

- Team 2

Problem Statement

War and Refugee Crisis

Dataset: Global Slavery Index

- **Vulnerable population** are increasingly at risk of human trafficking during crises, but "***patterns and triggers***" are not well documented.

Acquisition of data

All Data used was provided as a resource in the hackathon guidelines and instructions

Links Used:

Global Slavery Index - https://www.walkfree.org/global-slavery-index/downloads/

Country ISO Codes- ISO 3166-1 - Wikipedia

Cleaning the data

- Reads raw data from 'data.xlsx' (GSI 2023 summary data)
- Performs data cleaning operations:
- Removes duplicate entries
- Drops empty columns
- Removes header rows and resets index
- Selects relevant columns including:
- Country demographics
- Modern slavery statistics
- Governance metrics
- Vulnerability scores
- Handles missing values ('NaN' and 'N/A')
- Exports cleaned data to 'cleaned_data.csv'

Interactive World Map

- Creates interactive global visualization
- Libraries such as pycountry and plotly were used to create a world map view
- Maps vulnerability scores geographically
- The cleaned data values from the csv file were mapped to the iso Alpha 3 codes with the help of pandas
- Rows without iso codes were dropped
- An Interactive choropleth map was created with locations pointing to ISO3 and projection as natural earth (color scale used: Blues)
- Color-codes countries by vulnerability
    - Data cleaning → Risk prediction → Geographic visualization
    - Enables understanding of global modern slavery risks
    - Provides both analytical and visual insights

Risk Prediction using Forest Model

- Installed required Python libraries: pandas, numpy, matplotlib, seaborn, scikit-learn
- Sets visualization styles using ggplot and viridis color palette
- Loads cleaned data and selected four key features for analysis:
  - Risk factors percentage
  - Effects of conflict
  - Estimated slavery numbers
  - Government response percentage
- Target variable was Total Vulnerability score
- Removes rows with missing values
- Created a correlation heatmap showing relationships between features
- Plotted the distribution of vulnerability score

Risk Prediction using Forest Model

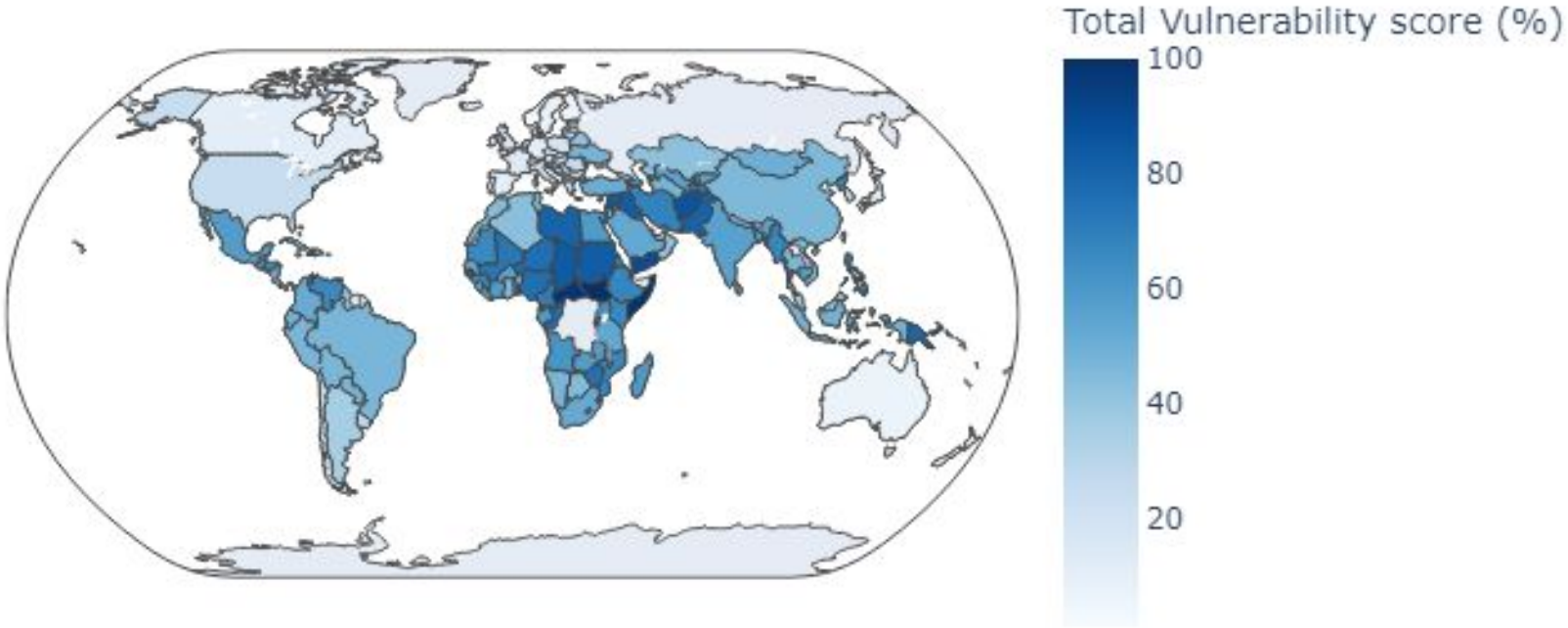
- Converts the continuous target variable into 3 categories using pd.qcut - Categories are labeled as 0, 1, 2 (representing low, medium, high vulnerability)

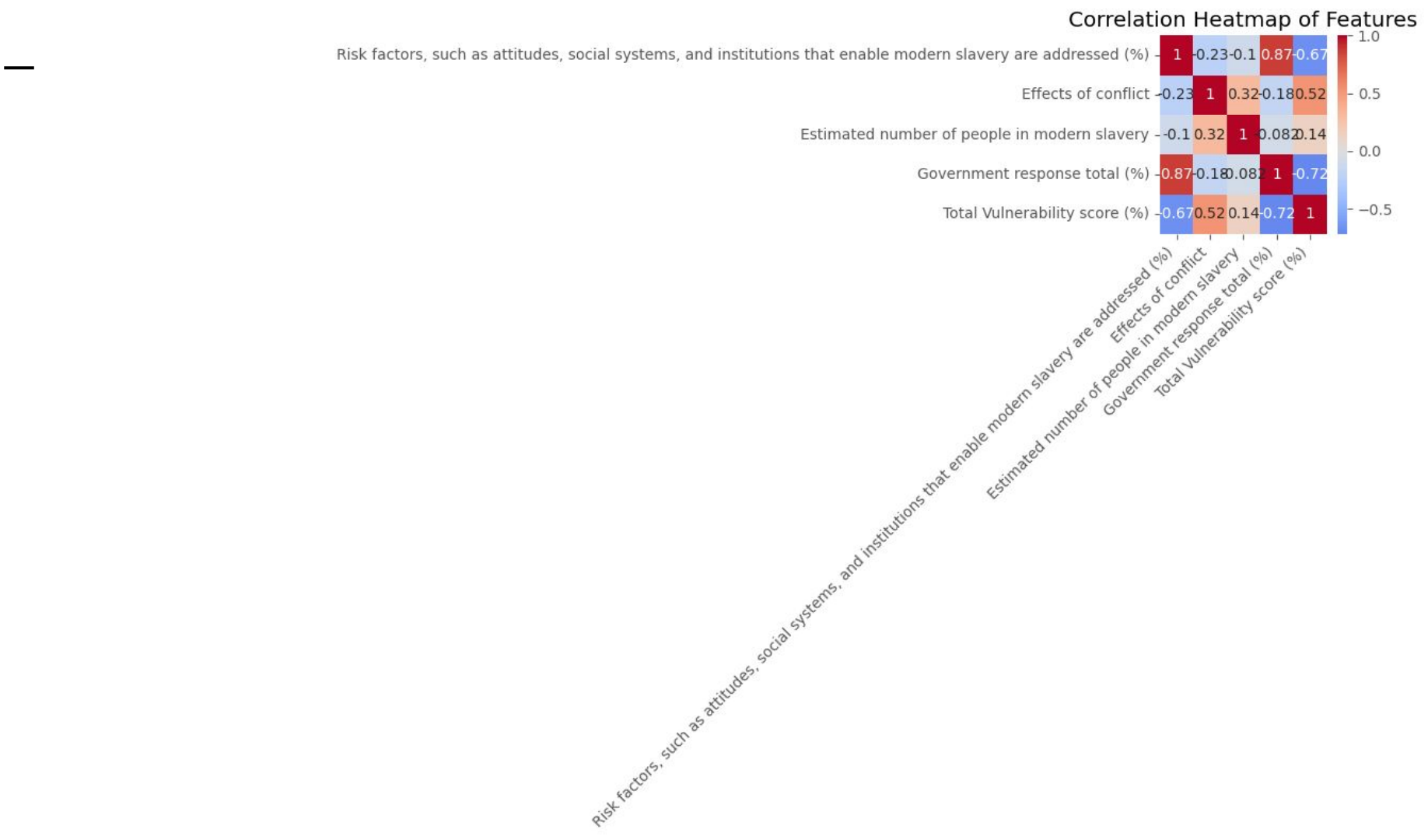
> Model Training & Evaluation

  - Splits data into training (80%) and testing (20%) sets
  - Creates a Random Forest Classifier with 100 trees
  - Trains the model on the training data
  - Makes predictions on test data
  - Calculates and prints model accuracy
  - Generates a detailed classification report
  - Creates a confusion matrix visualization
  - Feature Analysis
- Visualizes feature importance from the Random Forest model
- Creates a bar plot showing which features have the most impact on predictions
- Sorts features by importance for better visualization

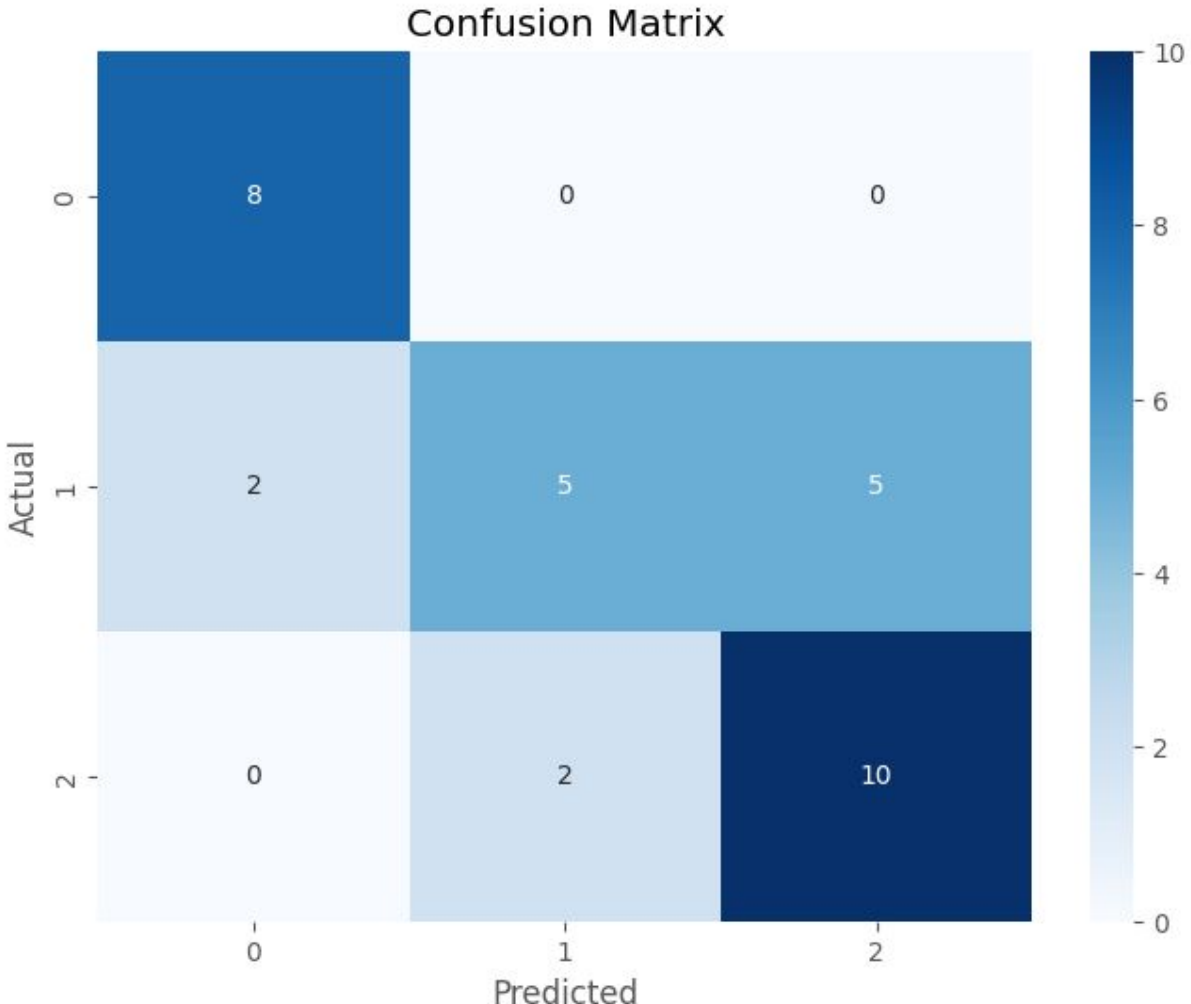
Outputs

Global Human Trafficking Vulnerability Map




Correlation Heatmap of Features
Risk factors, such as attitudes, social systems, and institutions that enable modern slavery are addressed (%)
Effects of conflict
Estimated number of people in modern slavery
Government response total (%)
Total Vulnerability score (%)
1 -0.23 -0.1 0.87 -0.67
-0.23 1 0.32 -0.18 0.52
-0.1 0.32 1 -0.082 0.14
0.87 -0.18 0.082 1 -0.72
-0.67 0.52 0.14 -0.72 1
1.0
0.5
0.0
−0.5

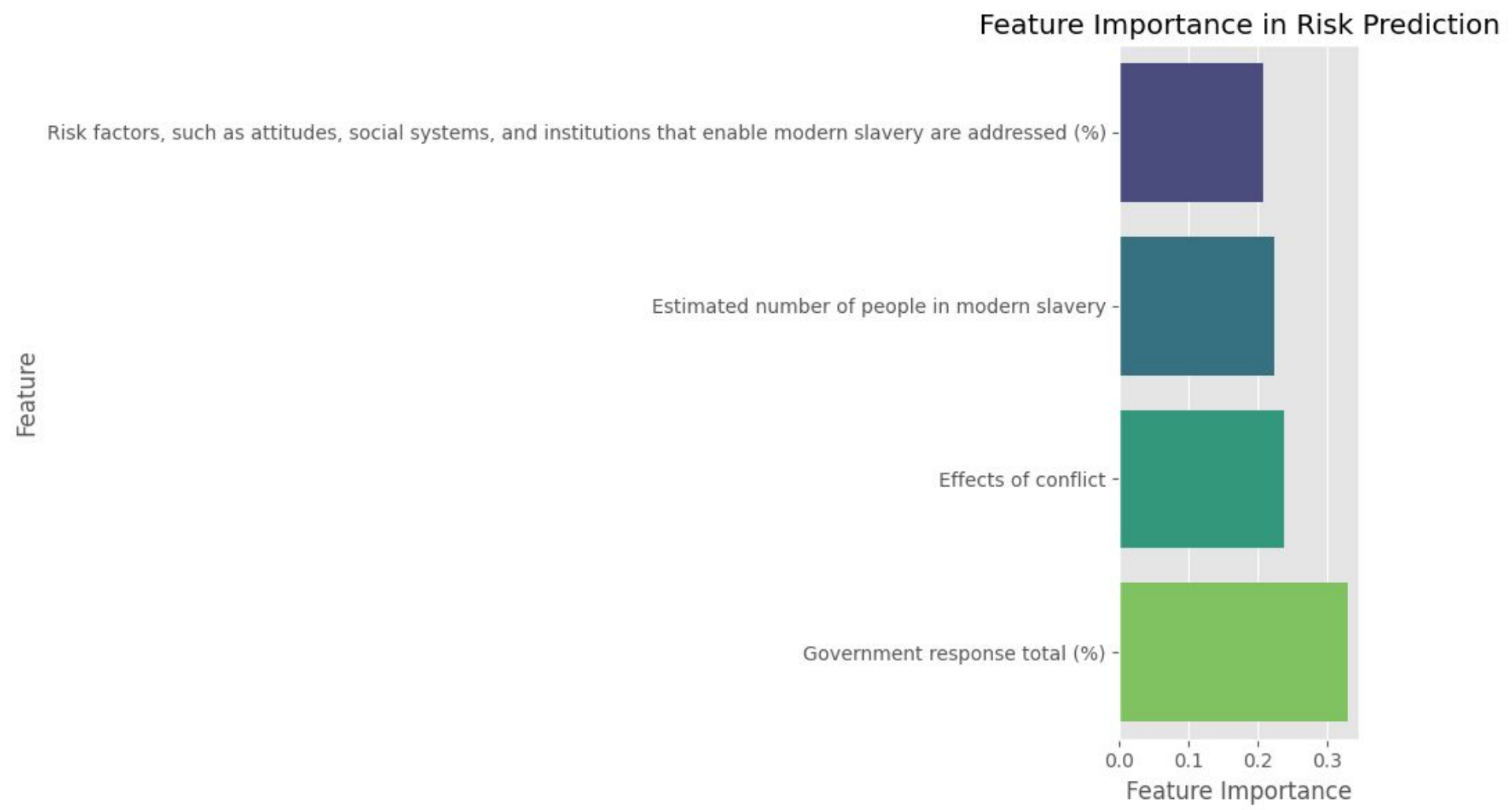

Distribution of Total Vulnerability Score
Count
Vulnerability Score
0
2
4
6
8
10
12
14
0
20
40
60
80
100



Model Accuracy: 0.72

Classification Report:

	precision	recall	f1-score	supp
0	0.80	1.00	0.89	8
1	0.71	0.42	0.53	12
2	0.67	0.83	0.74	12
accuracy			0.72	32
macro avg	0.73	0.75	0.72	3
weighted avg	0.72	0.72	0.70	


Feature Importance in Risk Prediction
Risk factors, such as attitudes, social systems, and institutions that enable modern slavery are addressed (%)
Estimated number of people in modern slavery
Effects of conflict
Government response total (%)
Feature
0.0
0.1
0.2
0.3
Feature Importance

Further Analysis

- Clustering
- Economic Indicators
- Network Analysis
- Time Series Analysis
- Policy Impact Analysis
- Risk Factor Decomposition
- Create composite risk scores