GLOBAL MODERN SLAVERY.

AN ANALYSIS
BY RIOPA MARGRET NAMUNYAK

OVERVIEW

- The purpose of this analysis is to provide actionable insights for combating modern slavery by examining key factors that contribute to its prevalence. By analyzing various features such as the lack of basic needs, the effectiveness of criminal justice mechanisms, and overall vulnerability scores, the study identifies critical areas that require targeted interventions. The analysis also evaluates the performance of predictive models to understand which factors are most influential in identifying at-risk populations and regions. The goal is to inform policy-making, enhance intervention strategies, and improve resource allocation to effectively reduce and ultimately eradicate modern slavery. These insights will help stakeholders, including governments, NGOs, and international organizations, to develop more efficient and focused approaches to addressing modern slavery across different contexts and regions.

BUSINESS UNDERSTANDING

- The primary objective of analyzing the Global Slavery Index (GSI) dataset is to develop a classification model that accurately categorizes countries based on their vulnerability to different types of modern slavery, specifically forced labor, human trafficking, and child exploitation. This classification will enable stakeholders to identify and prioritize regions where specific interventions are most needed, thereby contributing to more effective and targeted efforts in combating modern slavery globally.

Stakeholders:

- **Non-Governmental Organizations (NGOs**): NGOs that are focused on human rights and anti-slavery initiatives will benefit from the classification model by gaining insights into where their efforts can have the most impact, allowing them to tailor interventions to the specific types of slavery prevalent in different regions.
- **Governments and Policy Makers**: Governments can use the classification results to enhance their policy frameworks, strengthen law enforcement, and allocate resources more efficiently
- **International Bodies (e.g., United Nations, International Labour Organization):** These organizations can use the model's outputs to monitor global trends, coordinate international responses, and support countries in addressing their specific modern slavery challenges.
- **Academics and Researchers:** Scholars studying modern slavery can use the classification as a foundation for further research into the causes and solutions for different types of modern slavery.

DATA UNDERSTANDING

- The Global Slavery Index (GSI) dataset provides detailed data on modern slavery, including socio-economic, political, and demographic indicators across various countries. The dataset aims to measure the prevalence of modern slavery, encompassing different forms such as forced labor, human trafficking, and child exploitation. Additionally, it includes data on government responses, vulnerability factors, and regional differences, providing a comprehensive view of the global state of modern slavery.

Source of the data

The data for the Global Slavery Index is sourced from multiple authoritative bodies and research organizations, including but not limited to:

1. **Walk Free Foundation**: The primary organization behind the Global Slavery Index, which conducts extensive research and data collection on modern slavery
2. **International Labour Organization (ILO):** Provides estimates and data on forced labor and other forms of modern slavery.
3. **United Nations (UN):** Offers data on human trafficking and child exploitation through various UN agencies.
4. **World Bank:** Supplies demographic and socio-economic indicators such as population, inequality, and governance.
5. **National Surveys**: Data collected from national-level surveys conducted in various countries to assess vulnerability to modern slavery.

These sources are integrated into the GSI dataset, ensuring that the data reflects a wide range of reliable inputs, though it's important to note that data collection methods may vary across countries.

DATA ANALYSIS

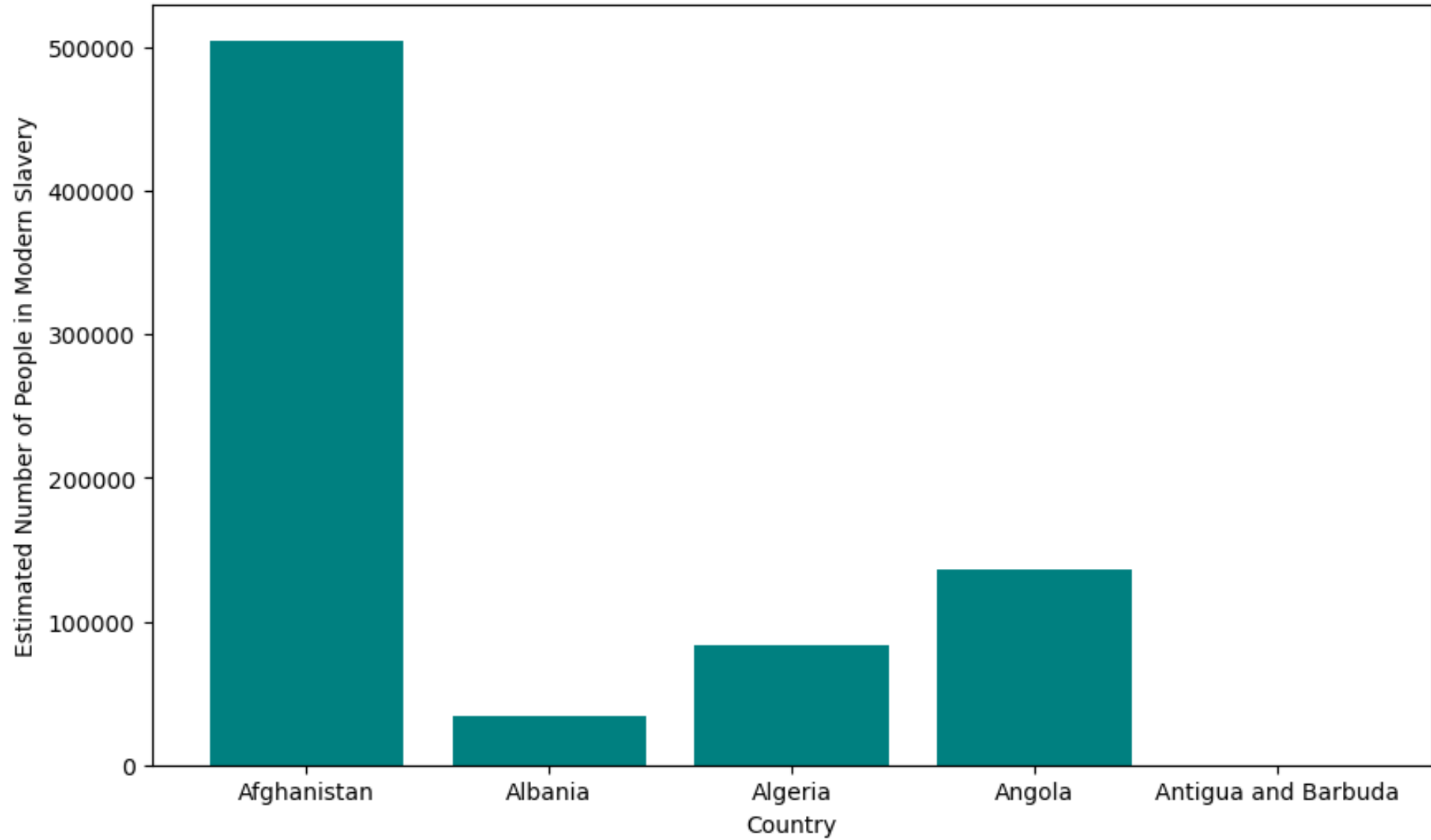Our analysis involved several key steps:

- **Importing Libraries**: I imported several libraries from Basic data manipulation, Preprocessing models, to Decision Tree model , Logistic Regression Model, Hyperparameter Tuning Model among others
- **Data Cleaning**: We handled missing values, checked for duplicates, log transformed some column and did a standardization of the whole dataset
- **Exploratory Analysis**: We examined various factors such as Estimated number of people in modern slavery, Estimated prevalence to modern slavery , did analysis to understand the estimated number of people in modern slavery among East Africa countries..
- **Graphs and Visualizations:** We created multiple visualizations to understand the data better. For instance:

- ✓ **Bar Charts**: Showing Estimated Number of people in Modern Slavery.
- ✓ **Line Chart**: Illustrated Estimation Prevalence for Modern Slavery for the First Five Countries
- ✓ **Confusion Matrix**: Confusion Matrix of Logistic Regression and Decision Trees.
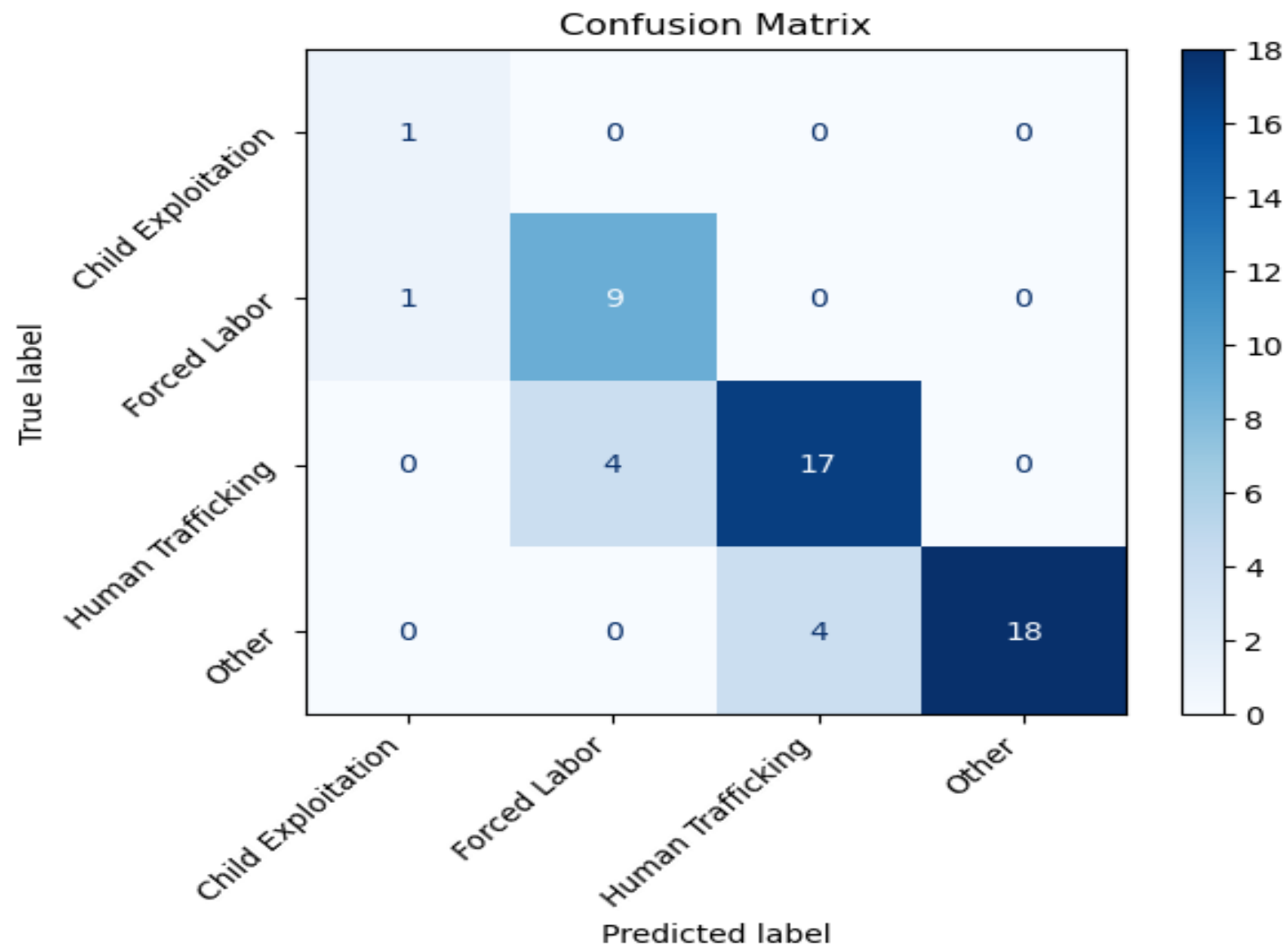- ✓ **Decision Tree Plot**: A plot showing the nature of a decision Tree.
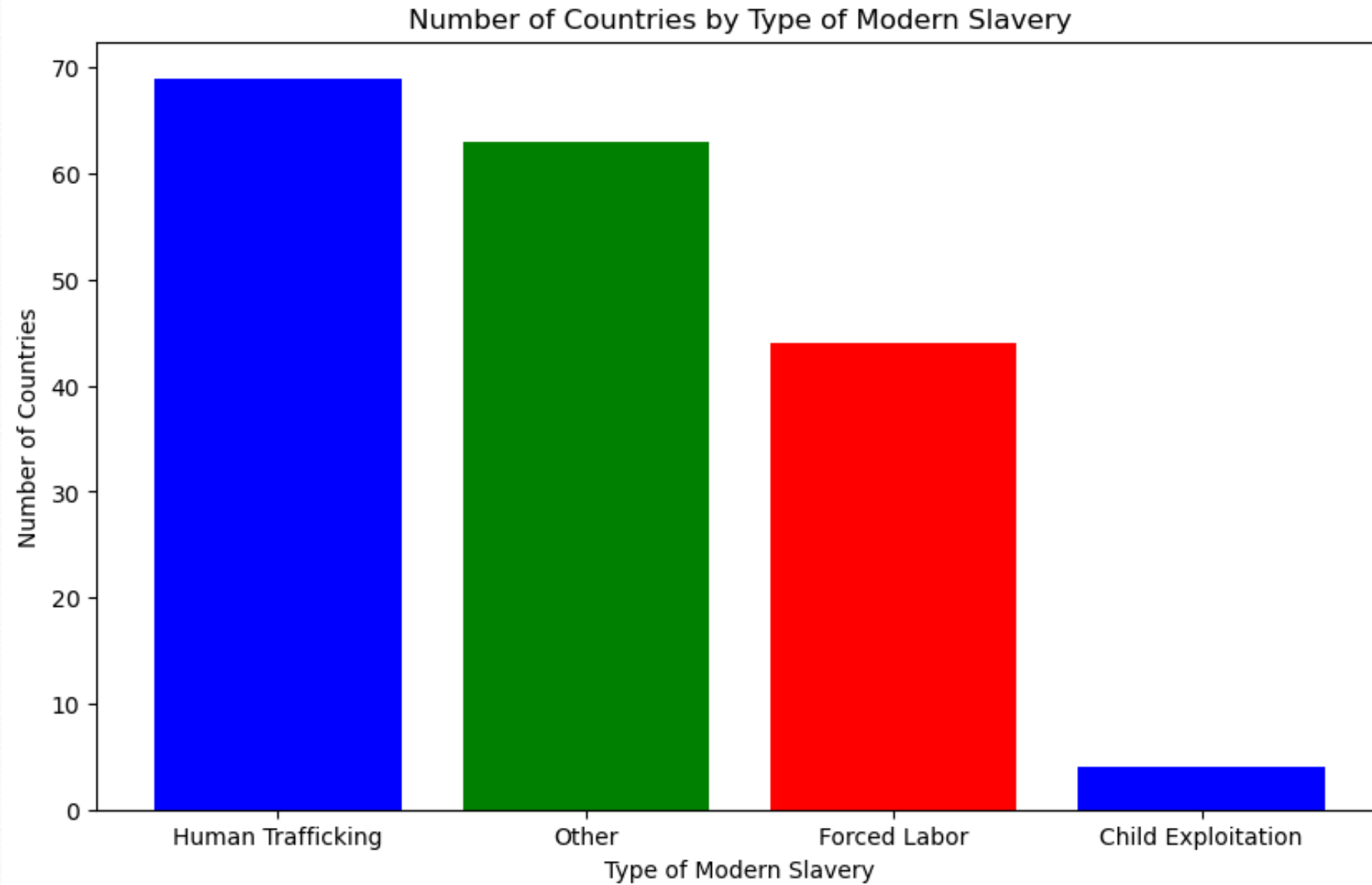
Number of Countries by Type of Modern Slavery
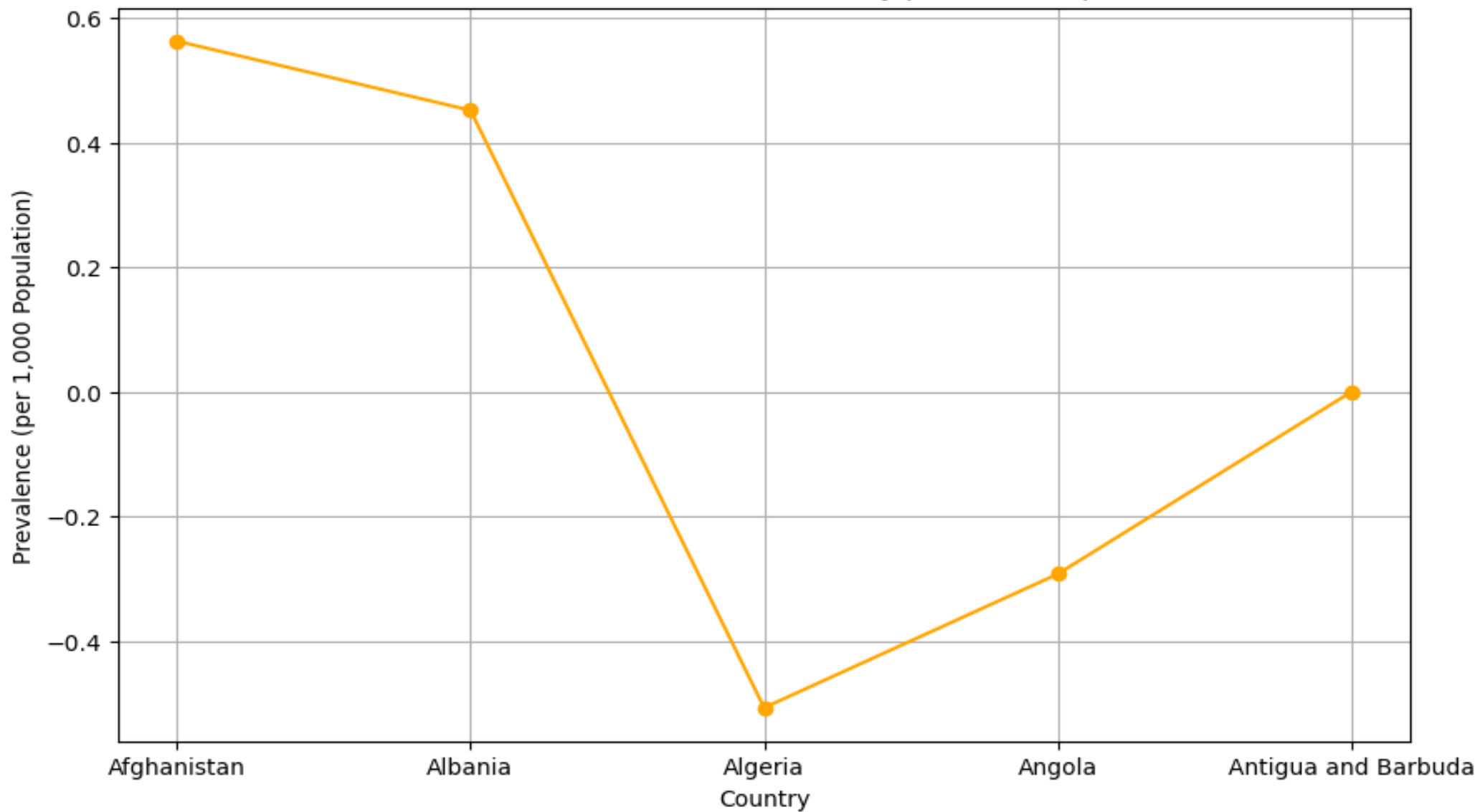
Estimated Number of People in Modern Slavery by Country

Confusion Matrix

Receiver Operating Characteristic Curve

Confusion Matrix for Decision Tree
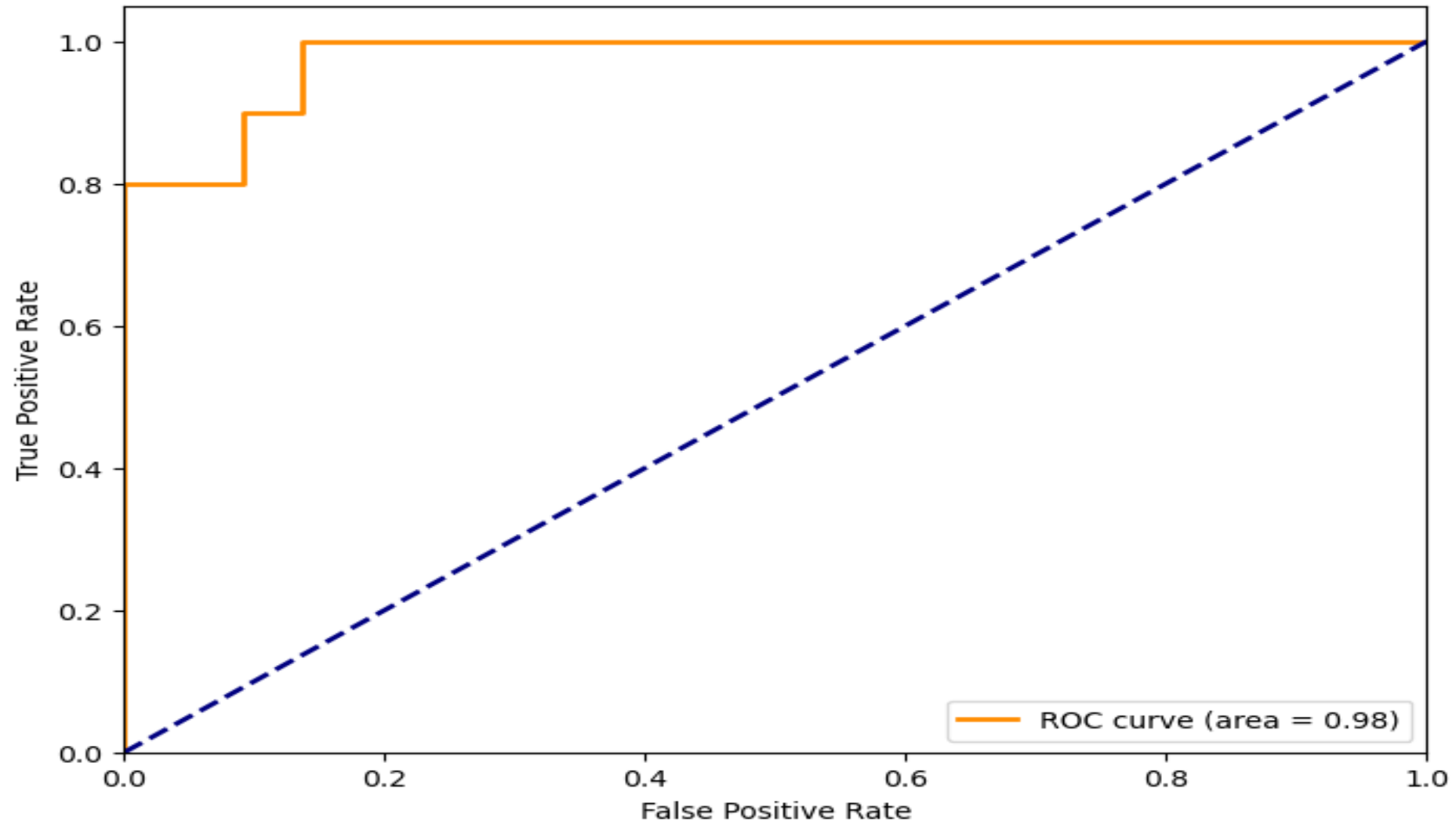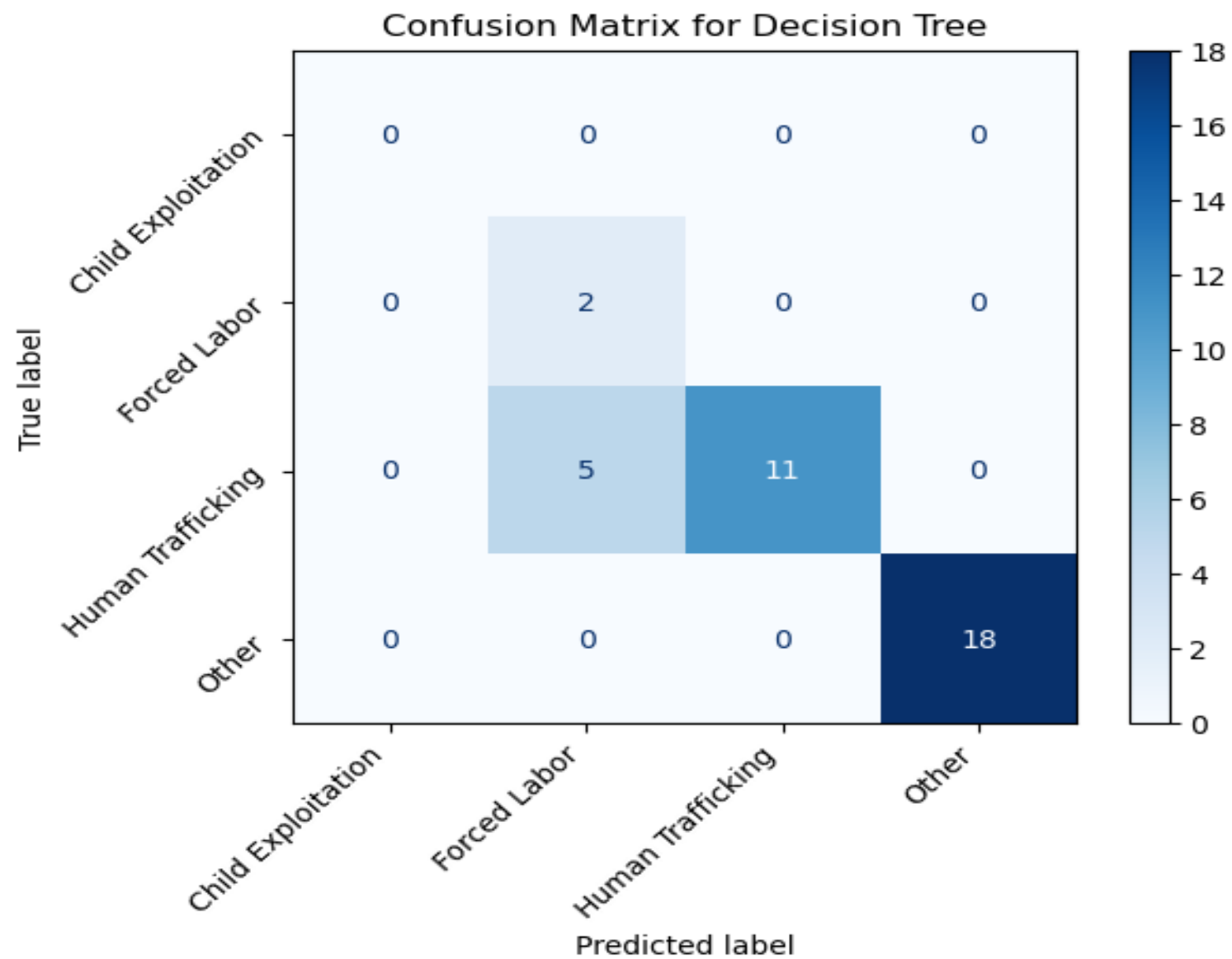
Decision Tree Visualization

Root node: remainder__Criminal justice mechanisms function effectively to prevent modern slavery (%) <= 0.206
gini = 0.681
samples = 144
value = [4, 42, 53, 45]
class = Human Trafficking

Left child: remainder__Lack of basic needs <= 0.62
gini = 0.484
samples = 90
value = [0, 37, 53, 0]
class = Human Trafficking

Right child: remainder__Lack of basic needs <= 0.93
gini = 0.291
samples = 54
value = [4, 5, 0, 45]
class = Other

remainder__Total Vulnerability score (%) <= 0.985
gini = 0.135
samples = 55
value = [0, 4, 51, 0]
class = Human Trafficking

cat__Country_Togo <= 0.5
gini = 0.108
samples = 35
value = [0, 33, 2, 0]
class = Forced Labor

remainder__Effects of conflict <= 0.767
gini = 0.15
samples = 49
value = [4, 0, 0, 45]
class = Other

gini = 0.0
samples = 5
value = [0, 5, 0, 0]
class = Forced Labor

gini = 0.0
samples = 51
value = [0, 0, 51, 0]
class = Human Trafficking

gini = 0.0
samples = 4
value = [0, 4, 0, 0]
class = Forced Labor

cat__Country_Mongolia <= 0.5
gini = 0.057
samples = 34
value = [0, 33, 1, 0]
class = Forced Labor

gini = 0.0
samples = 1
value = [0, 0, 1, 0]
class = Human Trafficking

gini = 0.0
samples = 45
value = [0, 0, 0, 45]
class = Other

gini = 0.0
samples = 4
value = [4, 0, 0, 0]
class = Child Exploitation

gini = 0.0
samples = 33
value = [0, 33, 0, 0]
class = Forced Labor

gini = 0.0
samples = 1
value = [0, 0, 1, 0]
class = Human Trafficking

Recommendations on the Models Performance

1. Address Class Imbalance:
Observation: The Decision Tree model showed high accuracy (86.1%) but struggled with minority classes, such as "Forced Labor," where it achieved low precision (0.29) and recall (1.00). This suggests the model may be overfitting to the majority classes.

Recommendation: Implement techniques like SMOTE (Synthetic Minority Over-sampling Technique) or class-weight adjustments to better balance the dataset. This can help models better identify and predict minority classes, improving overall model performance, especially for underrepresented categories.

2.Model Selection Based on Context:
Observation: The Decision Tree model demonstrated strong overall accuracy and excelled in certain classes, like "Other," where it achieved perfect precision and recall. However, it showed weaknesses in handling minority classes.

Recommendation: For tasks requiring high accuracy and the ability to capture complex patterns, ensemble methods such as Random Forests or Gradient Boosting are recommended. These methods could improve performance by reducing the variance seen in single decision trees. For tasks where interpretability is critical, a simpler model like Logistic Regression might be preferable.

3.Hyperparameter Tuning:

Observation: The performance differences between the tuned Decision Tree and Logistic Regression models were marginal, indicating that both models might already be close to their optimal performance for this dataset.

Recommendation: While further tuning might yield only slight improvements, it is still worthwhile to explore hyperparameter adjustments, particularly for the Decision Tree model, to prevent overfitting and improve its handling of minority classes.

4.Consideration for Ensemble Methods:

Observation: The Decision Tree model's high accuracy suggests that tree-based methods are effective for this dataset, but its performance on minority classes was less satisfactory.

Recommendation: Experiment with ensemble methods like Random Forests or Gradient Boosting. These approaches can combine multiple decision trees to reduce overfitting and enhance the model's ability to generalize, particularly in handling imbalanced datasets.

.

5.Improvement in Data Representation:

Observation: The performance of both models highlights the importance of data representation, especially in how features like "Lack of Basic Needs" and "Criminal Justice Mechanisms" are utilized.

Recommendation: Consider feature engineering to create new variables or improve existing ones. This might involve creating interaction terms, normalizing variables, or using domain-specific knowledge to refine the input data. Improved feature representation can significantly enhance model accuracy and reliability.

CONCLUSIONS

- Both Logistic Regression and Decision Tree models demonstrated strong overall accuracy, with Decision Trees slightly outperforming Logistic Regression in accuracy. However, the performance varied across different classes:

- Logistic Regression: This model provided a balanced performance across classes and handled the dataset's imbalance better than the Decision Tree, making it suitable for tasks requiring straightforward decision boundaries and high interpretability.

- Decision Tree: This model achieved higher overall accuracy but struggled with minority classes, such as "Forced Labor." This suggests that while Decision Trees are powerful for capturing complex relationships, they are sensitive to class imbalance and may require techniques like boosting or balancing to optimize their performance.

- Both models are effective, but the choice between them should be guided by the specific requirements of the task—whether it's interpretability and generalization (favoring Logistic Regression) or maximizing accuracy through capturing complex patterns (favoring Decision Trees or ensemble methods). Balancing the dataset and tuning hyperparameters are essential steps to further enhance model performance.

THANK YOU

Thank you for your attention. We look forward to implementing these findings that will help stakeholders to develop more efficient and focused approaches to addressing modern slavery across different contexts and regions.

.

This project is available on my github (https://github.com/Namnyak/P3-PROJECT)

My contact info: kinyeinamnyak@gmail.com