Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
F		2418649

Summary

Illegal wildlife trade poses a serious threat to global biodiversity and ecological balance, involving transactions worth billions of dollars annually. This paper proposes corresponding project measures to tackle this issue.

Firstly, we aligned the characteristics of various organizations with the requirements of the project and identified the United Nations Environment Programme as our client.

Later, to further convince the client, we collected relevant data on wildlife trade and World Development Indicators (WDI) from public data websites. We analyzed this data and then used it to construct a Risk Terrain Model (RTM). This model correlates illegal wildlife trade crime records with WDI data based on spatial information. By combining crime records and WDI data with spatial information through GIS, we employed the random forest algorithm to predict incidents of illegal wildlife trade occurrence in various regions worldwide. We evaluated the model's performance using ROC analysis. Subsequently, we assessed the importance of each indicator for each type of crime using permutation importance. We selected two indicators, incidence of tuberculosis and education expenditure, which ranked high in importance and were helpful for project planning.

Subsequently, considering the impact of government transparency on wildlife trade, we chose to use a linear regression model to fit the corresponding variables. However, we found that the relationship between the two was not significant. We speculate that both transparency and the scale of wildlife trade are influenced by a combination of various factors. Additionally, governments may formulate relevant policies based on their own circumstances to promote local economic development. This makes the linear correlation between transparency and related wildlife trade not strong.

Furthermore, based on wildlife trade data from various countries, we constructed a Trading Demand Network Model using trade information to build a network model. We discovered significant trading nodes, corresponding to the regions of the Americas and Asia. Measures can be implemented targeting these important nodes to curb illegal wildlife trade (IWT).

To quantitatively measure the impact of project implementation, we adjusted the values of corresponding parameters based on the actual project situation to observe the project's trajectory. When we increased the value of education expenditure, the model predicted a decrease in the crime rate. This demonstrates the feasibility of our model and project.

Finally, a sensitivity analysis was conducted on the Risk Terrain Model (RTM), where corresponding parameters were perturbed. The model maintained good performance even when the perturbed data accounted for approximately 5% of the total data. However, when the perturbed indicator values exceeded 10%, the model's performance significantly deteriorated. Our model exhibits strong robustness.

Keywords: IWT; United Nations Environment Programme; PCI; Risk Terrain Model; Trading demand network model

Table 3: World Development Indicators(WDI) Dataset Description

Key	**Description**
Number of indicators	1486
Number of Topics	90
Number of countries and areas	266
Year range	1960-2022

4.2 Data Pre-processing

4.2.1 Data Filling

For the World Development Indicators database, indicators with more than one year of missing data in a decade are excluded before interpolation, and indicators with more than 5 percent of countries missing are excluded, and domestic averages are prioritized for interpolation, followed by world averages.

4.2.2 Feature Engineering

For the World Development Indicators database, feature extraction was performed on the rest 170 variables after initial removal and interpolation. According to the correlation heatmap, there are many inter-correlated variables.

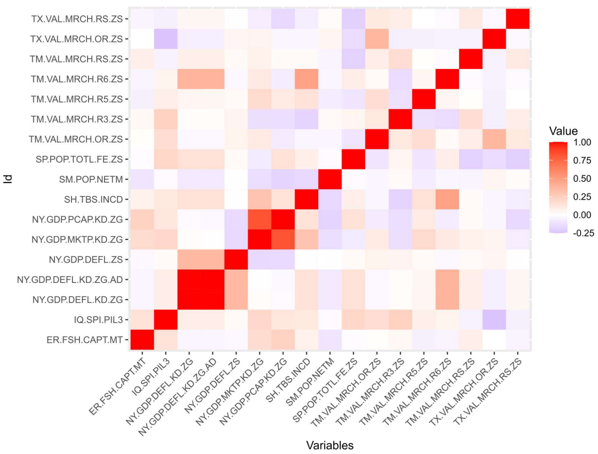


Fig. 2: Extracted WDI correlation heat map

Using the Random Forest algorithm, with the variables as independent variables and the total number of crimes in the country as the dependent variable, the top 10% of indicators based on Gini impurity were selected as the final indicators, which should contain relatively more information than others. The updated correlation heatmap shows that the correlation between variables has decreased, meaning there is less overlap in the information each variable contains.

1. WWF[3]
- 2023 World Wildlife Total Operating expenses $454,544,058
- 2023 World Wildlife Total program services $384,784,239
- 2022 World Wildlife Total Operating expenses $406,016,843
- 2022 World Wildlife Total program services $322,898,558

2. UNEP[4]:
- 2022-2023 Total budget $872,900,000
- 2022-2023 budget for nature actions $204,899,000
- 2021-2022 budget for nature actions $190,299,000

(2) The recent project undertaken (Example):
1. WWF
- Closing Asia's ivory markets
- Wildlife crime technology project
- Road map to zero-poaching in Selous

2. UNEP:
- Kunming-Montreal Global Biodiversity Framework
- Protect endangered waterbirds
- The cross-border effort to protect Caribbean wildlife

(3) Stakeholders/Related industries in Illegal Wildlife Trade:
Poachers, Traficckers, Customs, Financial Regulatory Institutions, the Environmental Protection Agency, the Department of Cultural Heritage/Artifact Conservation, and the National Parks/Nature Reserve Management Authorities, among others.
Within the United Nations system, an organization closely collaborating with UNEP in combating illegal wildlife trade is the United Nations Office on Drugs and Crime (UNODC). UNODC plays a role in combating transnational crimes, including the fight against illegal wildlife trade. Additionally, the United Nations Development Programme (UNDP) may collaborate with UNEP in jointly addressing incidents of illegal wildlife trade.

Conclusion:
Both WWF and UNEP possess ample funding and extensive project experience. However, as UNEP operates within the framework of the United Nations, it can collaborate more effectively, extensively, and in coordination with other UN agencies. This enables multi-agency assistance in collectively combating illegal wildlife trade. Therefore, we have chosen **UNEP** as our target client.

6 Model 1: Risk Terrain Mapping based on Random Forests

6.1 Algorithm Background

The illegal wildlife trade exhibits strong spatial dependence, being not only related to specific locations but also potentially associated with the economic and societal development levels of the respective countries or regions. The characteristics of exports and imports may also vary. Simply using machine learning methods to predict crime levels in different countries might overlook the importance of spatial location and geographical environment.

We chose Random Forests combined with GIS to perform crime analysis and prediction [5]. We adopt this strategy for two reasons: on the one hand, random forests excel in handling tabular data and possess the capability to capture inter-feature dependencies, making them more suitable for processing the dataset we use compared to other machine learning algorithms such as neural networks. On the other hand, considering that illegal wildlife trade (IWT) exhibits spatial dependencies, incorporating spatial information can make the model results more realistic and facilitate the formulation of more detailed plans.

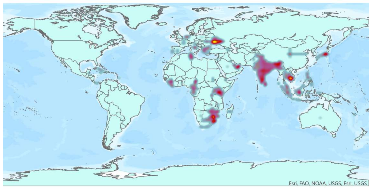
6.2 RTM based on Random Forests

The general workflow of the model is provided below [6]:

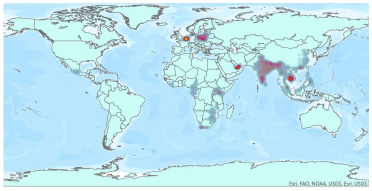
- Import the vector map data from Natural Earth and rasterize it, setting the grid size to 10000×10000 in meters. Pay special attention to the fact that the coordinate reference system (CRS) of this dataset is WGS84. Use latitude and longitude for coordinate mapping, corresponding to TRAFFIC.
- Map the coordinates of each indicator to the raster according to the corresponding country coordinates, thereby assigning a series of indicator values attributes to the grid.
- Count the coordinates of each record in TRAFFIC and tally them according to the grid, assigning a crime count attribute to each grid.
- Train a random forest model using the ”mlr” package in R, with the goal of predicting the crime count within each grid based on latitude and longitude coordinates, as well as indicator values.
- Evaluate the performance of the model using ROC (Receiver Operating Characteristic) analysis.
- Obtain the indicator with the highest permutation importance and use Shapley Regression to determine its importance across various grids.
- Results analysis.

6.3 Prediction Result

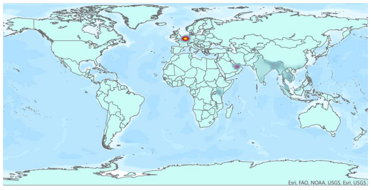
Using 70% of the records as the training set and the entire record dataset as the test set, we trained our model and predicted the four types of illegal wildlife trade crimes in different countries and regions worldwide. It’s important to note that the time range should also be from 2011 to 2020, and the predicted results can be considered to include potential crimes that were not included in the dataset. The heat map of predicted crime counts is depicted below using ArcGIS software. Notice that the four graphs are not directly comparable based on the color due to the difference in units.



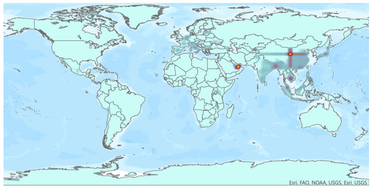
(a) Origin crime



(b) Discovery crime



(c) Transit crime



(d) Destination crime

Fig. 3: Predicted crime marked on global map

6.4 Evaluation Results

We adopted Receiver Operating Characteristic (ROC) as a measure of the model's merit. ROC is a metric based on a confusion matrix that describes the model's sensitivity and specificity, characterized by FPR and TPR, respectively.

		Predicted Value	
		Positive	Negative
Observed Value	Positive	True Positive Rate TPR	False Positive Rate FPR
	Negative	False Negative Rate FNR	True Negative Rate TNR

Fig. 4: Confusion matrix

Plotting as a graph with FPR on the horizontal axis and TPR on the vertical axis is the ROC curve, the more convex the curve is to the upper left indicates a better model.

The graph shows that our model has a relatively acceptable prediction performance, therefore we further analyze the dataset with the trained model.

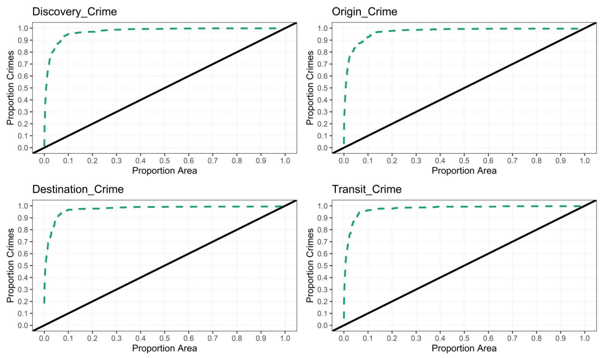


Fig. 5: ROC curve for Random Forests

6.5 Importance Analysis

Before continue, we need to clarify the difference between the analysis now and at feature engineering, that here we considered the spatial information while the other one did not. In addition, we also implemented a more detailed approach.

We first conducted a preliminary importance analysis to determine the most important indicators for each of the four types of crime records. Specifically, we utilized Permutation Importance to score different indicators. By examining the Mean Absolute Error (MAE) between the new predicted results and the original results after randomly shuffling an indicator, indicators with smaller errors are considered less important, while those with larger errors are considered more important. The results are shown in the following figure.

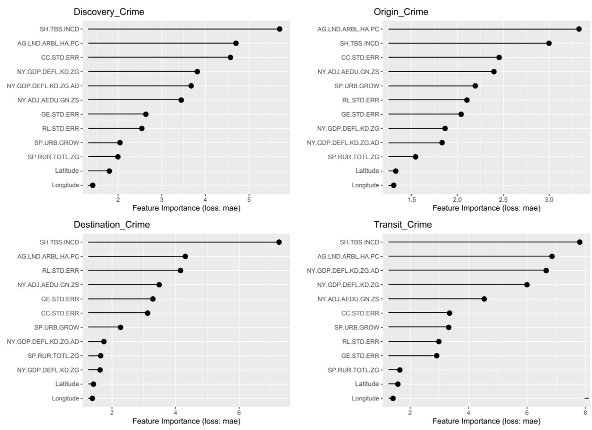


Fig. 6: Indicator permutation importance for different types of crime

The important indicator's corresponding description is provided below.

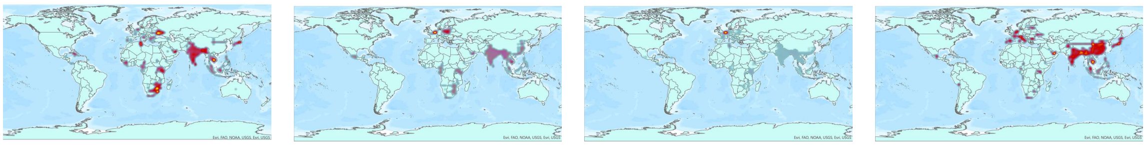
Table 5: Indicator Description

Key	Description
AG.LND.ARBL.HA.PC	Arable land (hectares per person)
CC.STD.ERR	Control of Corruption: Standard Error
GE.SED.ERR	Government Effectiveness: Standard Error
NY.ADJ.AEDU.GN.ZS	Adjusted savings: education expenditure (% of GNI)
NY.GDP.DEFL.KD.ZG	Inflation, GDP deflator (annual %)
NY.GDP.DEFL.KD.ZG.AD	Inflation, GDP deflator: linked series (annual %)
RL.STD.ERR	Rule of Law: Standard Error
SH.TBS.INCD	Incidence of tuberculosis (per 100,000 people)
SP.RUR.TOTL.ZG	Rural population growth (annual %)
SP.URB.GROW	Urban population growth (annual %)

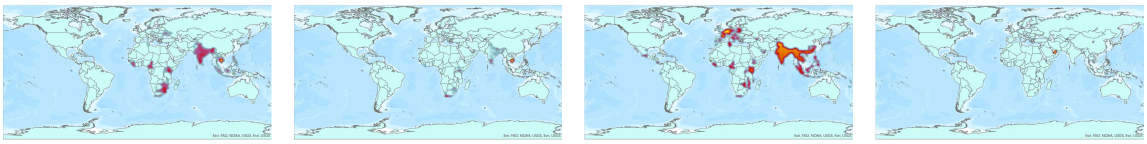
The figure illustrates that some indicators contribute significantly to the four types of crimes, such as education expenditure, incidence of tuberculosis, and inflation. Specifically, the "Rule of Law: Standard Error" indicator ranks particularly high in its contribution to IWT, especially in the import type, indicating a correlation between fluctuations in public confidence in the overall rule of law and illegal wildlife importation. Additionally, "Inflation, GDP deflator (annual %)" ranks particularly high in its contribution to transit trade, suggesting a strong association between inflation and the choice of transportation routes for illegal wildlife trade.

Based on project considerations, we have selected education expenditure and incidence of tuberculosis as the subjects of analysis.

Next, we examined the Shapley values of the important indicators of interest in various locations globally, thereby revealing the importance of these indicators in different countries and regions.



(a) Origin crime, education expenditure (b) Discovery crime, education expenditure (c) Transit crime, education expenditure (d) Destination crime, education expenditure



(e) Origin crime, incidence of tuberculosis (f) Discovery crime, incidence of tuberculosis (g) Transit crime, incidence of tuberculosis (h) Destination crime, incidence of tuberculosis

Fig. 7: Predicted crime marked on global map

By comparison, it is evident that education expenditure is strongly associated with the origin and import of illegal wildlife trade (IWT) across various regions. On the other hand, the incidence of tuberculosis is more closely related to the origin and transportation routes of IWT, with less significant contributions to imports.

6.6 Conclusion

Through the integration of crime records and World Development Indicators (WDI) data with spatial information using GIS, we employed the random forest algorithm to predict incidents of illegal wildlife trade (IWT) occurrence in various countries or regions worldwide. We evaluated the model using ROC analysis and found that the model demonstrated good predictive accuracy. Subsequently, we assessed the importance of each indicator for each type of crime using permutation importance and selected two indicators, incidence of tuberculosis and education expenditure, which ranked high in importance and were helpful for project planning, for further analysis. Lastly, using Shapley Regression, we calculated the contribution of these two indicators to different regions and found that education expenditure contributes significantly to both the origin and import of IWT across various regions, suggesting it could be a focus of our project efforts. Conversely, the incidence of tuberculosis contributes significantly to both the origin and transportation of IWT across different regions, indicating a likely correlation with increased opportunities for contact with wildlife due to illegal wildlife trade.

7 Model 2: Linear Regression-Relationship between CPI and the Rate of Wildlife Trade

1. Research Motivation:
Recognizing the substantial influence and extensive resource channels typically held by the public sector within a country/region, we hypothesize that the corruption level of public sector may impact the overall wildlife trade in that region. Thus, we aim to investigate the potential correlation between the CPI of public sector and the rate of wildlife trade.

2. Corruption Perceptions Index [7]:
(1) Definition:
CPI stands as the most widely utilized global corruption ranking, assessing how corrupt each country/territory's public sector is perceived to be.
The score reflects the perceived level of corruption within its public sector on a scale of 0-100, where 0 signifies high corruption, and 100 indicates a very clean public sector.

(2) Covered Manifestations:
The data sources used for compiling the CPI specifically encompass various manifestations of public sector corruption: bribery, diversion of public funds, nepotistic appointments in the civil service, access to information on public affairs and government activities, etc.

3. Introduction to the Dataset:
(1) Data Source:
The dataset has been compiled by team members through the collection of information from various sources such as CITES, the United Nations official website, and others[8] [9] [10]. The data has been meticulously curated and processed. Considering the potential impact of the Covid-19 pandemic on global trade, we have chosen to analyze the pre-pandemic period (2019).

10 Measurable Impact on Our Project

According to our project implementation plan mentioned above, we adjusted the education expenditure indicators of the hotspot countries in the hotspot map corresponding to the education expenditure indicators in the previous section upward to varying degrees, in the hope of predicting the changes in the corresponding global crime rate, as shown in the change map below.

The countries/areas are: Zimbabwe, South Africa, Mozambique, India, Netherlands, Ukraine, Cambodia, Thailand, Japan, Cuba, Barbados.

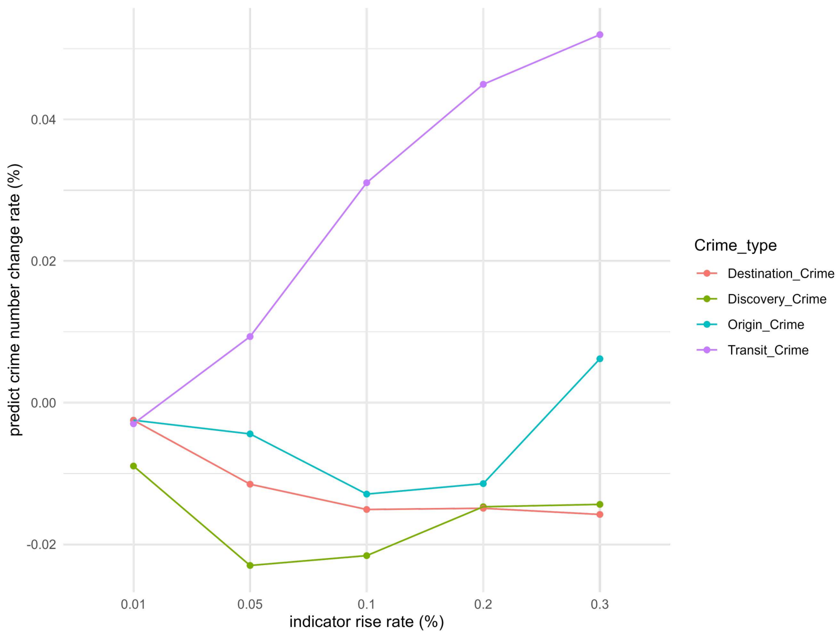


Fig. 15: Crime number change compare to origin (%)

Even by improving the education levels in these hotspot countries, we can promote a decrease in the global illegal wildlife trade (IWT) crime rates. This indicates that we can strategically invest in education funds to curb global IWT crimes. Particularly, when the indicators are increased to 10% above the original levels, the impact becomes insignificant, suggesting that education level is not the sole important influencing factor. On the other hand, even with a slight increase in indicators, we can still reduce a certain number of IWT crimes.

It's important to note that even a small percentage reduction in illegal wildlife trade (IWT), given a large baseline, can still result in a significant decrease in the actual number of wildlife deaths. We should relentlessly strive for this outcome.

11 Sensitivity Analysis

We want to assess the sensitivity of the random forest model because the data in the World Bank and TRAFFIC datasets may contain a significant amount of noise, leading to unreliable results. Therefore, we perturbed different proportions of the WDI data from TRAFFIC and observed changes in the ROC curve to determine whether the model's performance is affected.

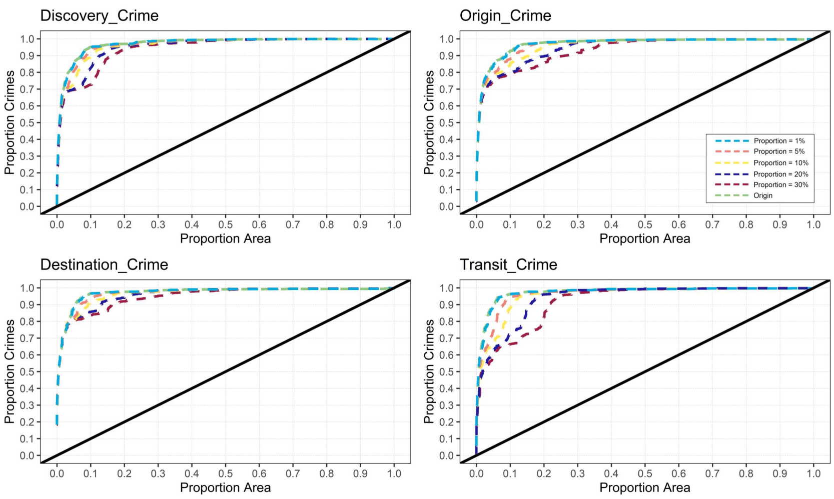


Fig. 16: Sensitive analysis of Random Forests based on ROC

We found that the model maintained good performance even when the perturbed data accounted for approximately 5% of the total data. However, when the perturbed indicator values exceeded 10%, the model's performance significantly deteriorated. This suggests that the model exhibits strong robustness, and its conclusions remain credible even when the original dataset may deviate from reality due to potential biases.

12 Model Evaluation and Further Discussion

12.1 Strengths

- Incorporating the World Development Indicators (WDI) into the model allows for a more comprehensive understanding of the relationship between various factors and illegal wildlife trade. This enhances the completeness and robustness of the model, enabling a more systematic analysis of the factors influencing illegal wildlife trade.

- Analyzing the relationship between various factors and illegal wildlife trade (IWT) spatially provides a more realistic perspective, leading to clearer and more convincing conclusions. This spatial analysis approach allows for a better understanding of how geographical factors influence IWT patterns, contributing to the overall effectiveness and persuasiveness of the model's findings.

12.2 Weaknesses

- Due to the challenges in obtaining illegal wildlife trade (IWT) data and the presence of numerous missing values in some datasets, the model we obtained may exhibit certain biases in its predictions compared to real-world situations.