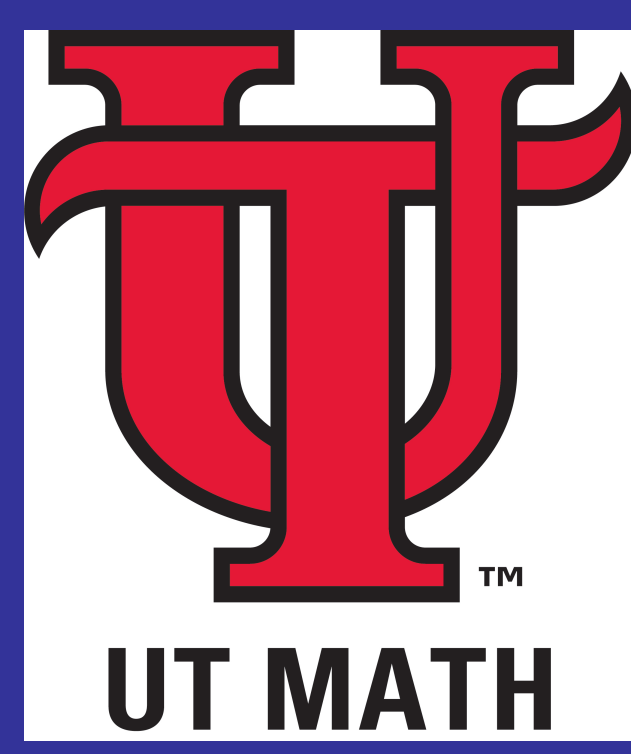




Predicting Ecological Debt using Machine Learning Models

Darri Stuber, Riley Rutigliano, Mentor : Dr. Binod Rimal

Department of Mathematics, University of Tampa

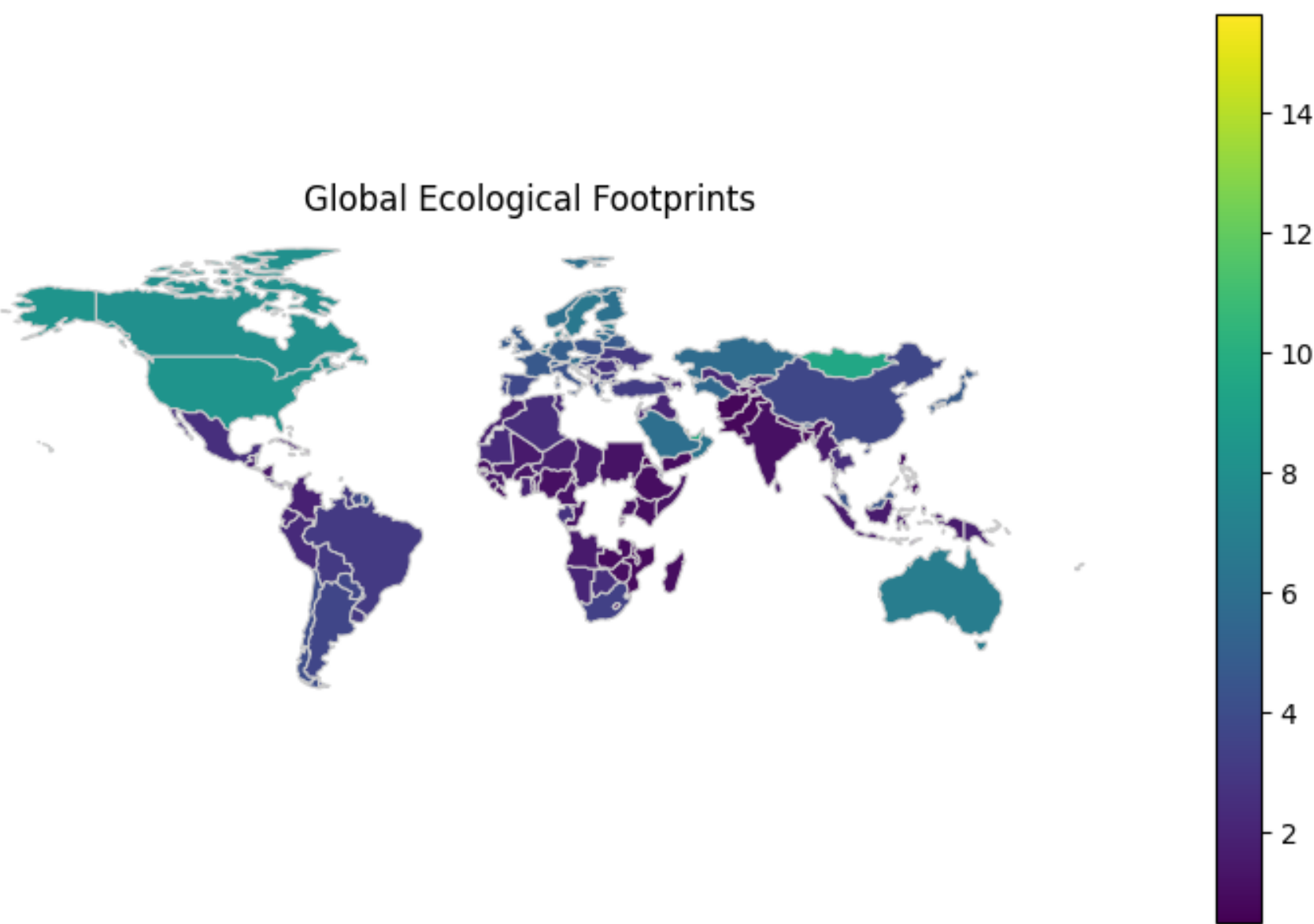


Abstract

In the modern age of consumption and intensive resource use, it is important to look introspectively at human behaviors over time. In this project, we implemented a data analytical process to indicate the degree to which natural resources are overused. Based on their ecological footprint (EF) and biocapacity with an extensive set of related defining characteristics, countries that overuse their resources are deemed ecological debtors, while those that underuse them are ecological creditors. We predicted creditor/debtor class using three classification models: logistic regression, decision trees, and random forest. Through our comparison, we deemed the random forest model to produce higher accuracy and f1-score values than the other two models. The resultant model is advantageous to use when evaluating the future environmental footprint of all countries throughout the world.

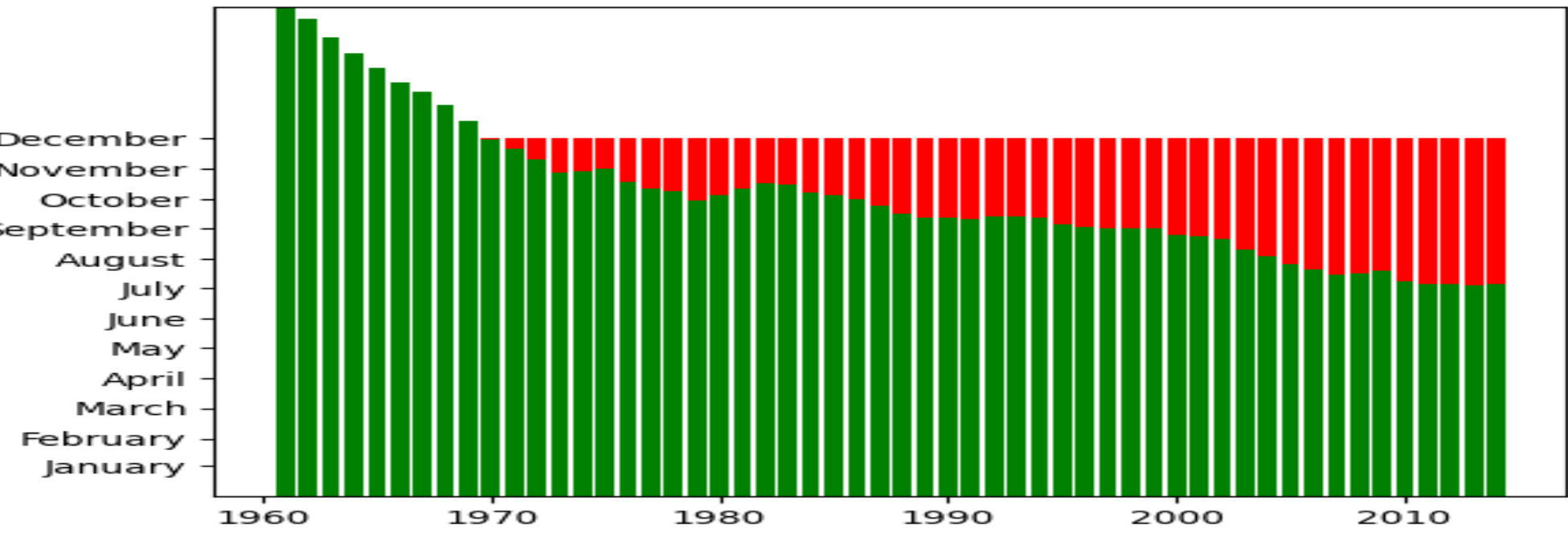
Ecological Footprint

EF is a measure of how much land and water area is required to produce the resources and services consumed by an individual, population, or activity. It is expressed in terms of global hectares (gha), which is a standardized measure that accounts for the differing productivity of land and water resources in different regions of the world. Biocapacity is the land and water area that is naturally available for use. These measurements include two components: the demand for natural resources and the capacity of ecosystems to regenerate those resources. The demand for natural resources includes the consumption of food, fiber, timber, and energy, as well as the use of land for buildings and infrastructure. The capacity of ecosystems to regenerate those resources includes the ability of land and water resources to produce food, fiber, timber, and other resources, as well as the capacity of forests, oceans, and other ecosystems to absorb carbon dioxide and other pollutants from consumption waste[Global Footprint Network]. If the EF exceeds the biocapacity (available land and water resources), it indicates that the consumption of natural resources is unsustainable - countries with this condition are deemed ecological debtors. A country with an EF less than their biocapacity is described as an ecological creditor. Analyzing the relationship between each country's ecological footprint and biocapacity allows for the identification of patterns to be used in effective policy strategies [Niccolucci et al., 2012].



Earth Overshoot Day

Earth Overshoot Day refers to the date for each year in which humans' consumption of resources and services surpasses those that can be sustainably offered on an annual basis. Each year, after the overshoot day has occurred, any resources used cannot be sustainably replenished or processed [Network]. This is how humans are cementing our ecological debt. In the subsequent graph, the annual debt is shown in red, corresponding to the date (shown in months) in which human have began to overuse that years' allotted resources. As shown, the overshoot day for 2014 was August 29th. Other sources alarmingly note that since then, the annual overshoot day has moved up to July 28th 2022 [Network].



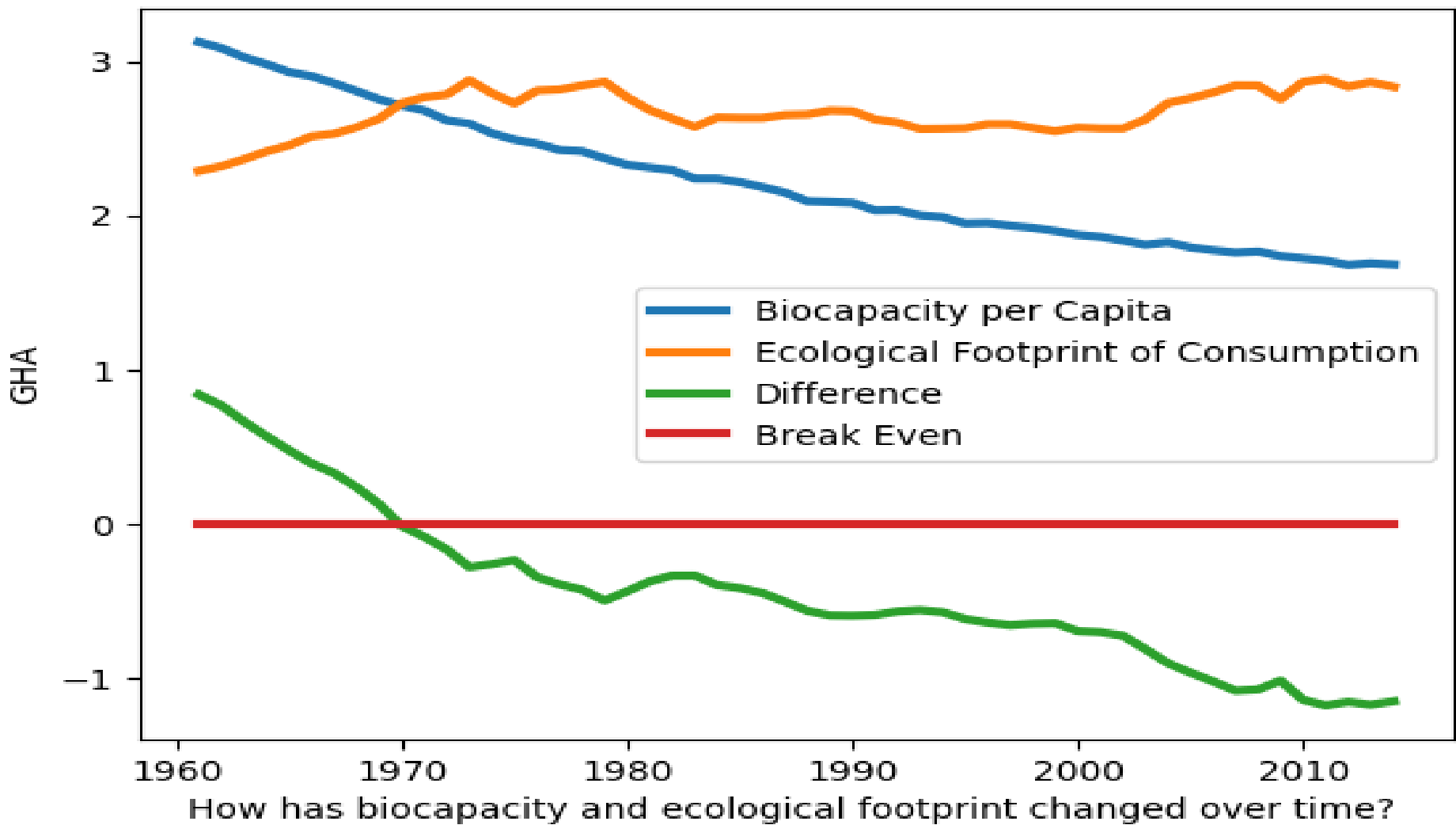
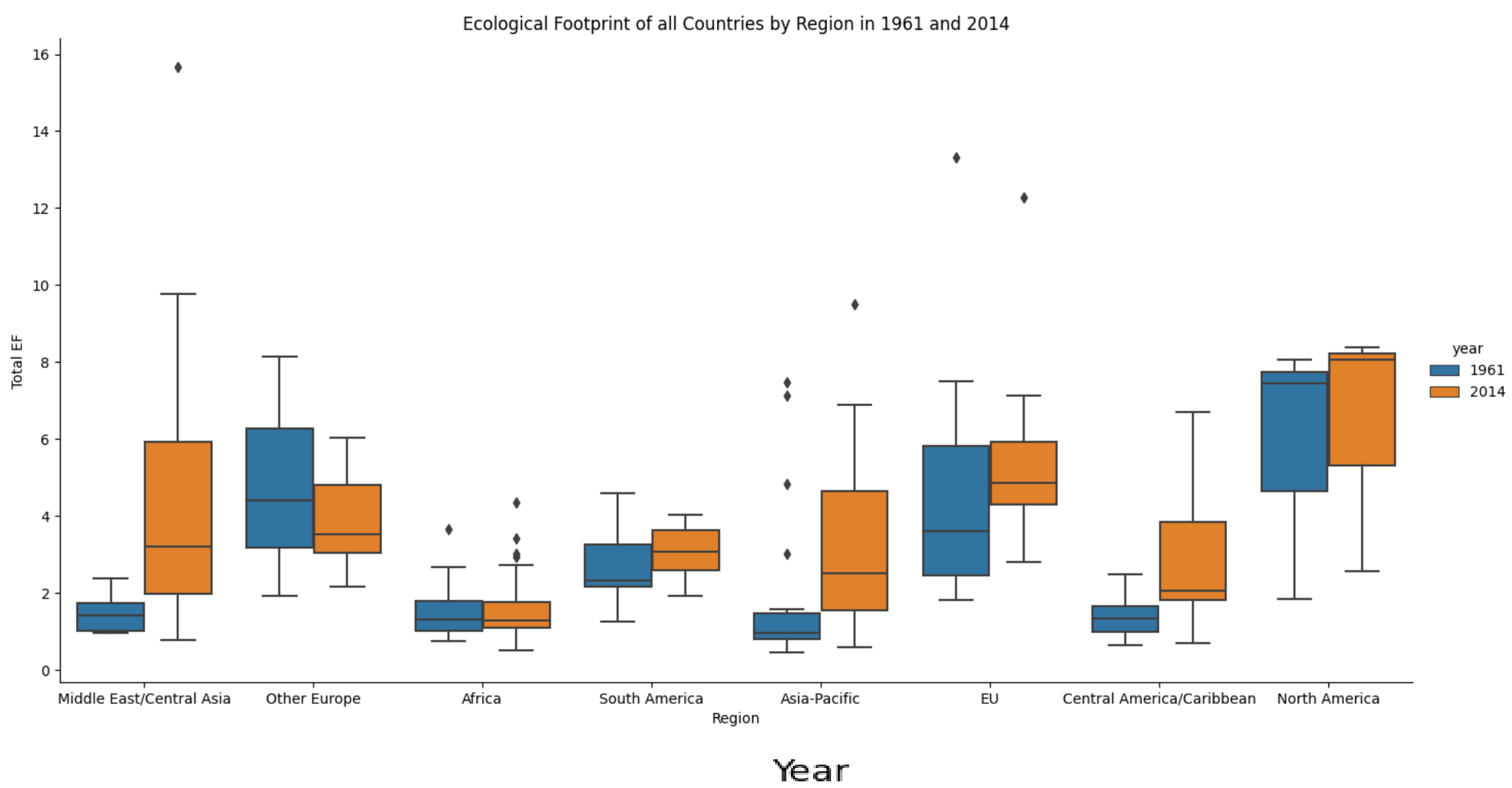
Environmental Footprint and Biocapacity Dataset

The data set chosen in the Machine Learning Repository is "Ecological Footprint and National Biocapacity". It contains 16 different attributes; 14 are numerical and 2 are categorical. We also have 61,568 instances in the data set. The characteristics of the data set is multivariate, meaning that there are many different data types.

Variables	Description
Country	Name of Country (string)
Year	Year of the data (integer)
BC_crop_land	Bio-capacity per capita for cropland (float)
BC_grazing_land	Bio-capacity per capita for grazing land (float)
BC_forest_land	Bio-capacity per capita for forest land (float)
BC_fishing_ground	Bio-capacity per capita for fishing ground (float)
BC_built_up_land	Bio-capacity per capita for built-up land (float)
BC_total	Total Bio-capacity throughout all land types (float)
EFC_crop_land	EF consumption per capita for cropland (float)
EFC_grazing_land	EF consumption per capita for grazing land (float)
EFC_forest_land	EF consumption per capita for forest land (float)
EFC_fishing_ground	EF consumption per capita for fishing ground (float)
EFC_built_up_land	EF consumption per capita for built-up land (float)
EFC_total	Total EF consumption per capita throughout all land types (float)
Differences	Difference of Bio-capacity and EF Consumption per capita (float)
Status	Determination if the difference is positive or negative (Creditor/Debtor)

Table 1. List of variables and their description used in the dataset [Network, 2021].

Data Visualization



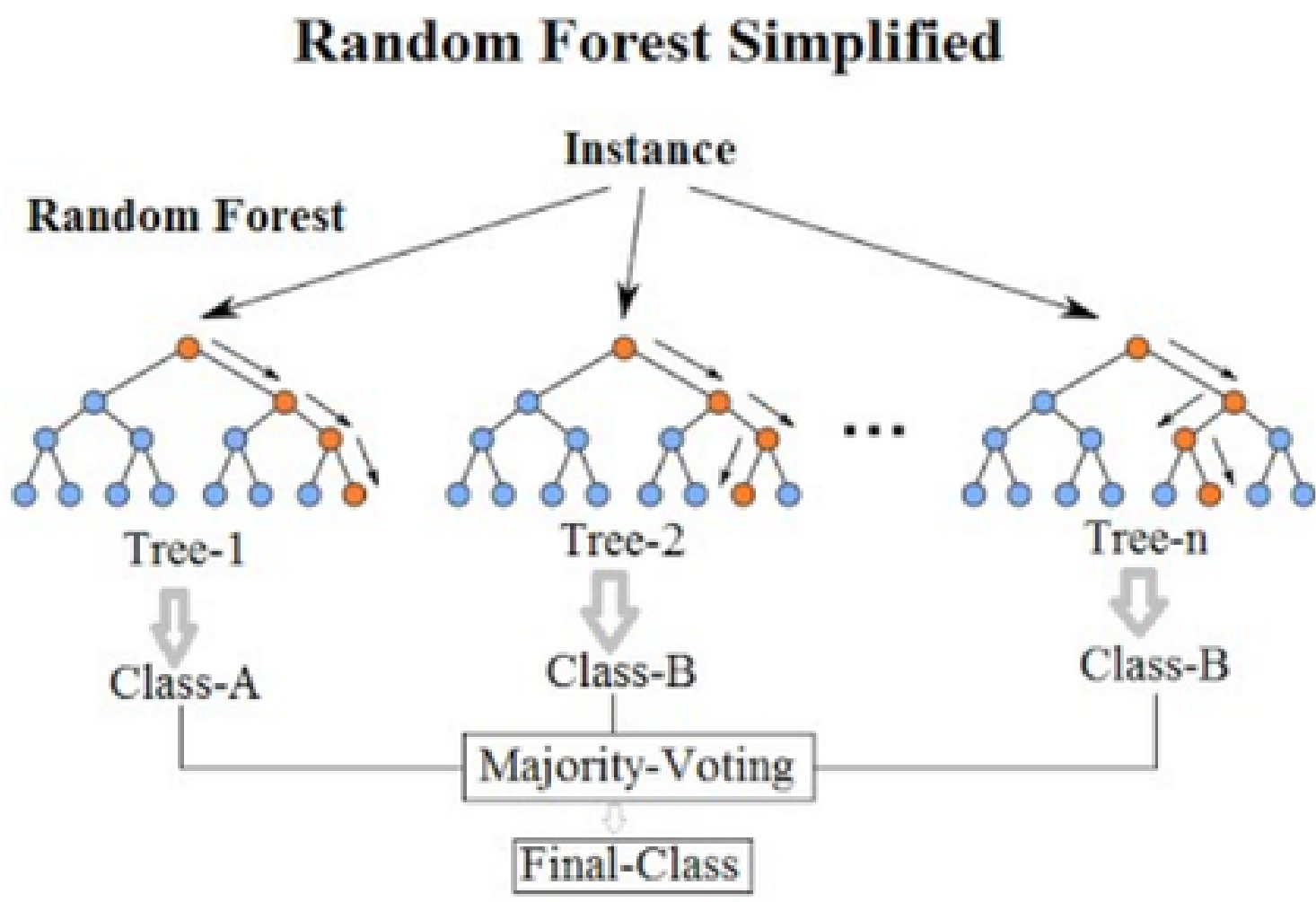
Machine Learning Models

▪ **Logistic Regression:** The logistic regression model is a probabilistic model to predict the default class ($y=1$) for the binary classification problem. It is defined mathematically as

$$P(X) = Prob(y = 1|X) = \frac{1}{1 + exp(-\beta X)}$$

where y is the predicting variable, X is/are predictors, and β is a regression coefficient. In our case, it is $Prob(\text{Status} = \text{True} | X)$ i.e. the probability of being ecological debtor is True given other input information as X .

▪ **Decision Tree and Random Forest** [Wikipedia contributors, 2023].



Performance of Three Machine Learning Models

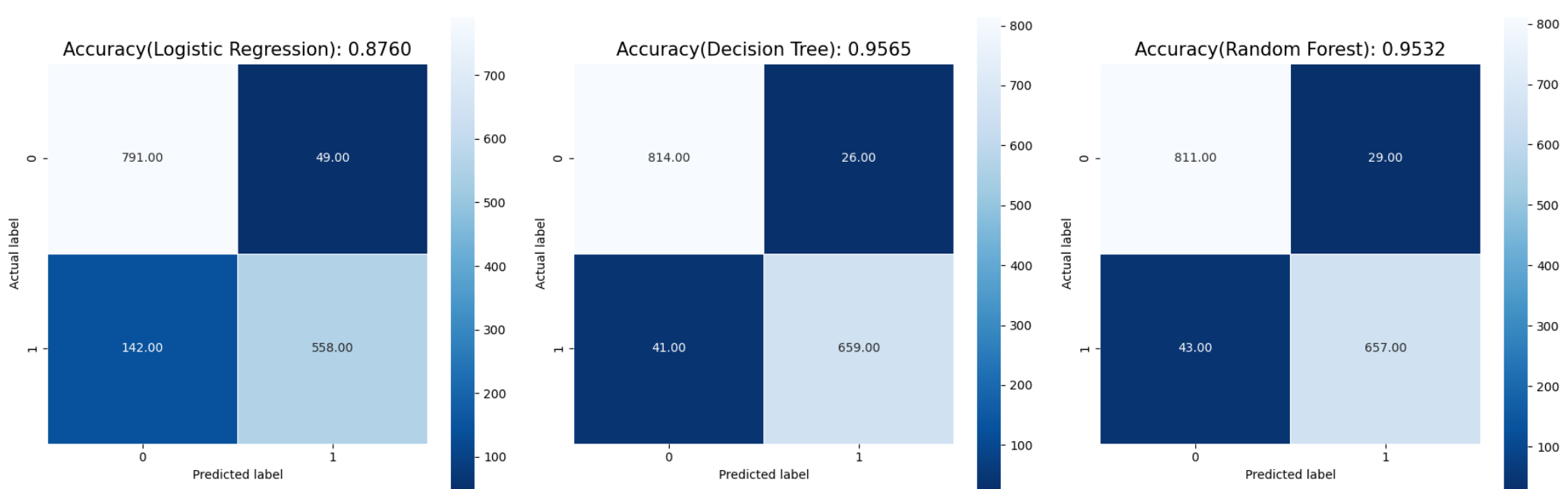


Figure 1. Confusion matrix for the classification of Ecological Class (Creditor/Debtor) from Logistic Regression, Decision Tree, and Random Forest respectively.

Model	Class	Precision	Recall	F1-score	Wt. F1-score	Accuracy
Logistic Regression	0	0.85	0.95	0.89	0.88	0.88
	1	0.93	0.79	0.85		
Decision Tree	0	0.91	0.94	0.92	0.94	0.94
	1	0.92	0.89	0.91		
Random Forest	0	0.925	0.97	0.96	0.95	0.95
	1	0.96	0.94	0.95		

Table 2. Performance scores of three models on predicting ecological debt/credit status

References

Global Footprint Network. Ecological footprint. <https://www.footprintnetwork.org/our-work/ecological-footprint/>. Online.
Earth Overshoot Network. Last year, earth overshoot day fell on July 28. <https://www.overshootday.org/?>. Online.
Global Footprint Network. "ecological footprint and national biocapacity". <https://www.kaggle.com/datasets/thedevastator/uncovering-links-between-ecological-footprint-an?select=NFA2021>. [Online].
V Niccolucci, E Tiezzi, FM Pulselli, and C Capineri. Biocapacity vs ecological footprint of world regions: A geopolitical interpretation. *Ecological Indicators*, 16:23–30, 2012.
Wikipedia contributors. Random forest — Wikipedia, the free encyclopedia, 2023. URL https://en.wikipedia.org/w/index.php?title=Random_forest&oldid=1150821923. [Online; accessed 20-April-2023].