# Relation Between Economic Freedom and GDP per Capita by Group Random

Orçun Çelik[1], Gökçenur Duran[2], Zeynep Gökçe Tarlakazan[3], Onur Tahmaz[4], and Chingis Safarov[5]

[1,2,3,4,5]Boğaziçi University

August 7, 2019

# Contents

**Abstract**

In our research, we tried to explore whether there is a relationship between the determinants of economic freedom and GDP per capita of countries. In order to do this we constructed a prediction model which allows us to see the impact of economic freedom variables on GDP per capita of countries. Linear Regression, Logistic Regression and K-Nearest Neighbors are three different methods of Machine Learning that we used in our project. We found that “property rights” is the best feature to predict GDP per capita of countries together with “financial freedom”, “trade freedom” and “unemployment”. For several reasons, we found that logistic regression is more suitable for our research.

# Chapter 1

# Introduction

Economic freedom is the freedom to prosper within a country without intervention from a government or economic authority. Individuals are free to secure and protect his or her human resources, labor and private property. The Heritage Foundation‘s Index of Economic Freedom is the most widely referenced index of economic freedom. The index refers to mostly institutional strength and functioning within a country. A country‘s score in the index is based on 12 qualitative and quantitative features grouped into four broad categories: Rule of Law (property rights, government integrity, judicial effectiveness); Government Size (government spending, tax burden, fiscal health); Regulatory Efficiency (business freedom, labor freedom, monetary freedom) and Open Markets (trade freedom, investment freedom, financial freedom). According to many studies and researches, that was made on the effect of Economic Freedom on a nation‘s prosperity, a country or state that upholds principles of economic freedom experiences more prosperity. We wanted to see the effect of these factors on development. Although GDP is not perfect to explain development levels of countries, it is broadly used as an important indicator of economic performance and a useful unit to make cross-country comparisons of average living standards and economic well-being. While GDP is the most widely used measure of a country‘s economic activity, per capita GDP is a better indicator of a nation‘s prosperity since it adjusts for population. In this project, by using different techniques of Machine Learning and the latest data available, we want to find out whether there is a connection between Economic Freedom and a nation‘s prosperity which can be expressed as per capita GDP.

In order to understand and analyze the data correctly, we calculated the world rank of the Index of Economic Freedom and GDP per capita according to the 5 regions: Asia-Pasific, Europe, Middle East and North Africa, Sub-Saharan Africa and Americas. There is a relation between two variables, but there are some outliers which we deleted as discussed in the next section.

# Chapter 2

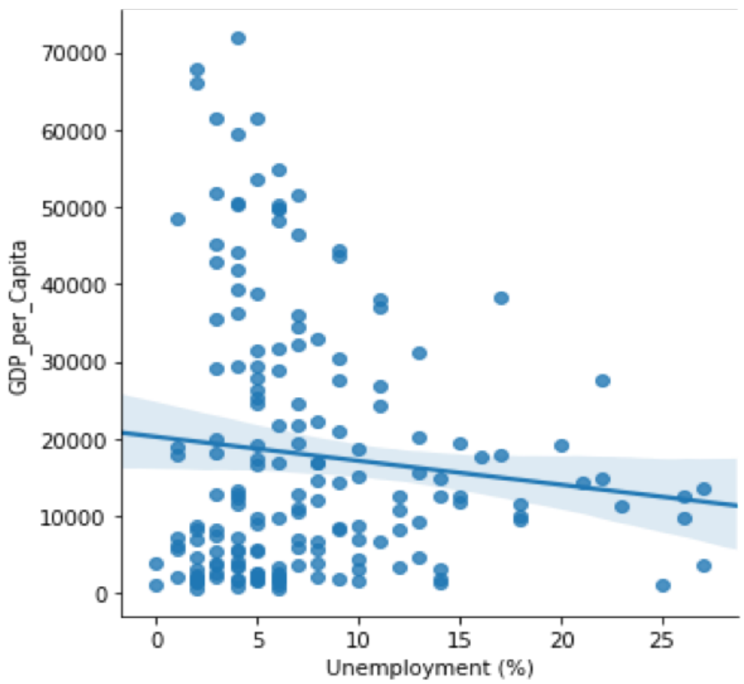
# Data Description

We used “Index of Economic Freedom” and “GDP per Capita” data-sets from ‘The Heritage Foundation 2019 Annual Report for the project. Index of Economic Freedom is evaluated concerned with a country’s interactions with the rest of the world, focusing on policies within a country, assessing the liberty of individuals to use their labor or finances without undue restraint and government interference. The 12 aspects of economic freedom of 186 countries measured in the Index are grouped into four broad categories:

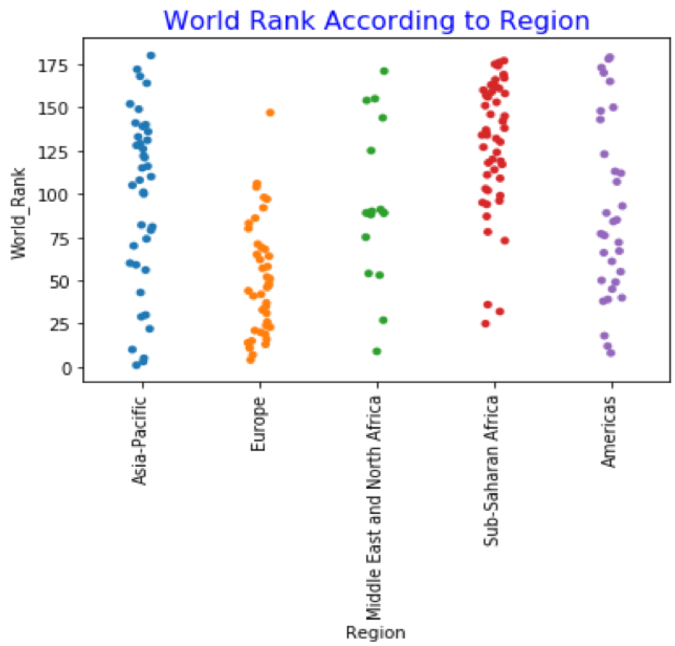
- Rule of law (property rights, judicial effectiveness, and government integrity)
- Government size (tax burden, government spending, and fiscal health)
- Regulatory efficiency (business freedom, labor freedom, and monetary freedom) and

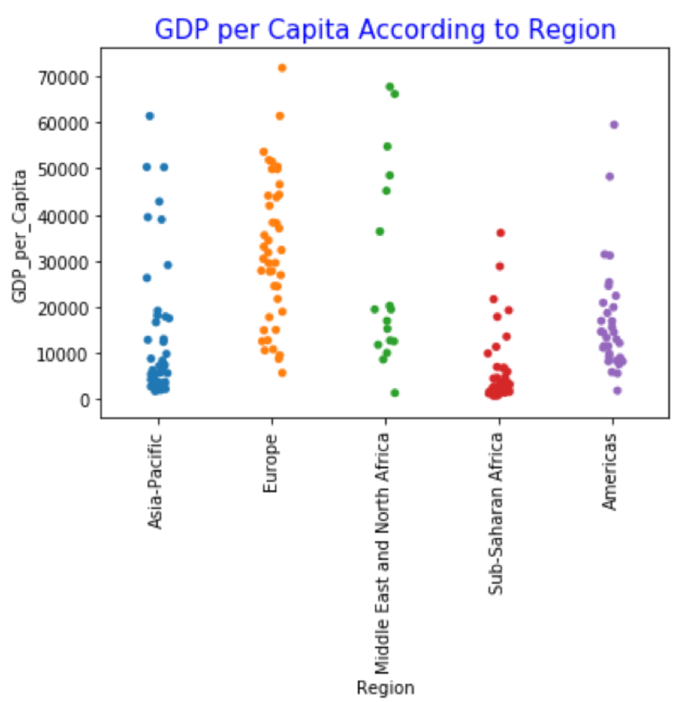
- Market openness (trade freedom, investment freedom, and financial freedom)

In addition, we added another variable to variables in Index of Economic Freedom: Unemployment. We showed its relationship with GDP per capita in the graph below.




In order to understand and analyze the data correctly, we calculated the world rank of the Index of Economic Freedom and GDP per Capita according to the 5 regions: Asia-Pasific, Europe, Middle East and North Africa, Sub-saharan Africa and Americas. As it is seen below, there is a relation between two variables, but there are some outliers. We deleted some outliers as discussed in the next section.


World Rank According to Region
World_Rank
0
25
50
75
100
125
150
175
Asia-Pacific
Europe
Middle East and North Africa
Sub-Saharan Africa
Americas
Region


GDP per Capita According to Region
GDP_per_Capita
0
10000
20000
30000
40000
50000
60000
70000
Asia-Pacific
Europe
Middle East and North Africa
Sub-Saharan Africa
Americas
Region

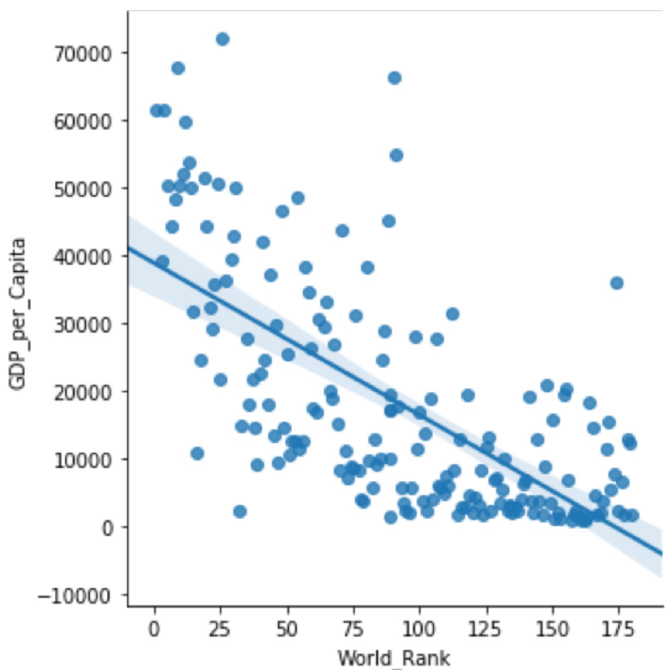
## 2.1 Data Cleaning and Exploratory Analysis

We started with importing the Python packages : Pandas, NumPy, Matplotlib and Seaborn. Then we cleaned our data by getting rid of the null values. Then we filled the null values with the countries' median specified by their region. We checked if we could clean the data from the null values. Some countries such as United Arab Emirates where the GDP per capita is high but the institutional and social relations are weak, created a problem in our analysis. Therefore, we detected and dropped these outlier values in GDP per capita. Then our data has 179 rows and 17 columns.

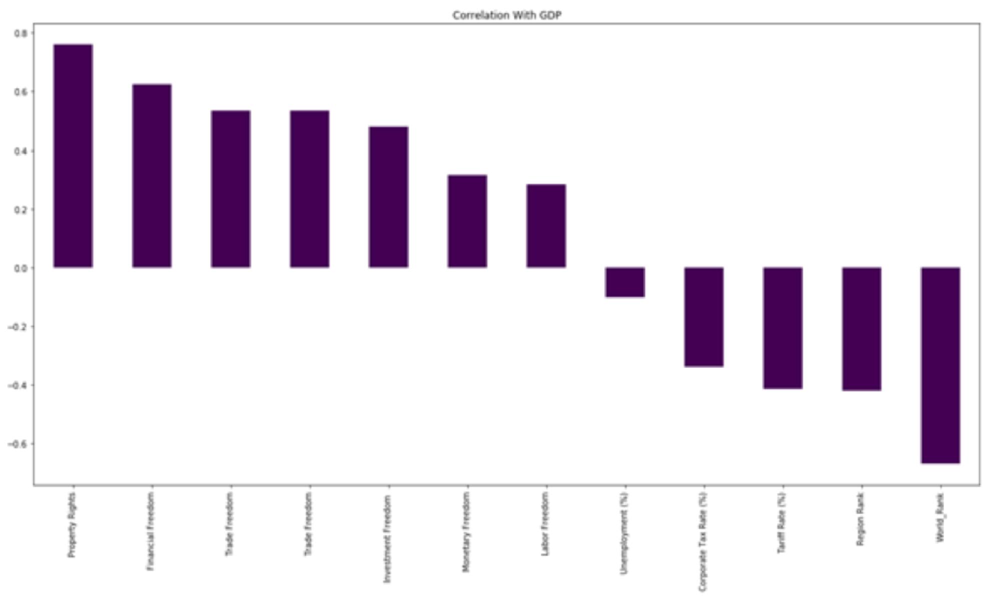
| | World_Rank | Region Rank | 2019 Score | Property Rights | Judical Effectiveness | Government Integrity | Tax Burden | Gov't Spending | Fiscal Health | Business Freedom | ... | Financial Freedom | Tariff Rat (% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 179.000000 | 179.000000 | 179.000000 | 179.000000 | 179.000000 | 179.000000 | 179.000000 | 179.000000 | 179.000000 | 179.000000 | ... | 179.000000 | 179.00000 |
| mean | 92.905028 | 20.737430 | 60.256983 | 51.608939 | 44.173184 | 40.597765 | 77.335196 | 64.083799 | 65.268156 | 63.089385 | ... | 48.156425 | 6.20111 |
| std | 50.759478 | 12.637482 | 10.869548 | 19.202151 | 17.753878 | 19.065976 | 13.216199 | 23.202303 | 31.780774 | 15.774850 | ... | 19.093446 | 5.50217 |
| min | 1.000000 | 1.000000 | 6.000000 | 8.000000 | 5.000000 | 8.000000 | 0.000000 | 0.000000 | 0.000000 | 5.000000 | ... | 0.000000 | 0.00000 |
| 25% | 50.500000 | 10.000000 | 54.000000 | 37.000000 | 31.000000 | 26.000000 | 71.000000 | 51.500000 | 38.500000 | 54.000000 | ... | 30.000000 | 2.00000 |
| 50% | 92.000000 | 20.000000 | 61.000000 | 50.000000 | 43.000000 | 36.000000 | 79.000000 | 69.000000 | 80.000000 | 63.000000 | ... | 50.000000 | 5.00000 |
| 75% | 135.500000 | 31.000000 | 67.000000 | 64.500000 | 53.500000 | 50.000000 | 86.000000 | 81.500000 | 90.000000 | 74.500000 | ... | 60.000000 | 9.00000 |
| max | 180.000000 | 47.000000 | 90.000000 | 95.000000 | 87.000000 | 97.000000 | 100.000000 | 97.000000 | 100.000000 | 96.000000 | ... | 90.000000 | 50.00000 |

8 rows × 24 columns

In addition, we used seaborn package to show that the Index of Economic Freedom is inversely proportional to GDP per Capita. The variable "world rank" below shows countries' rank on numerical value based on all 12 features of the Index of Economic Freedom.

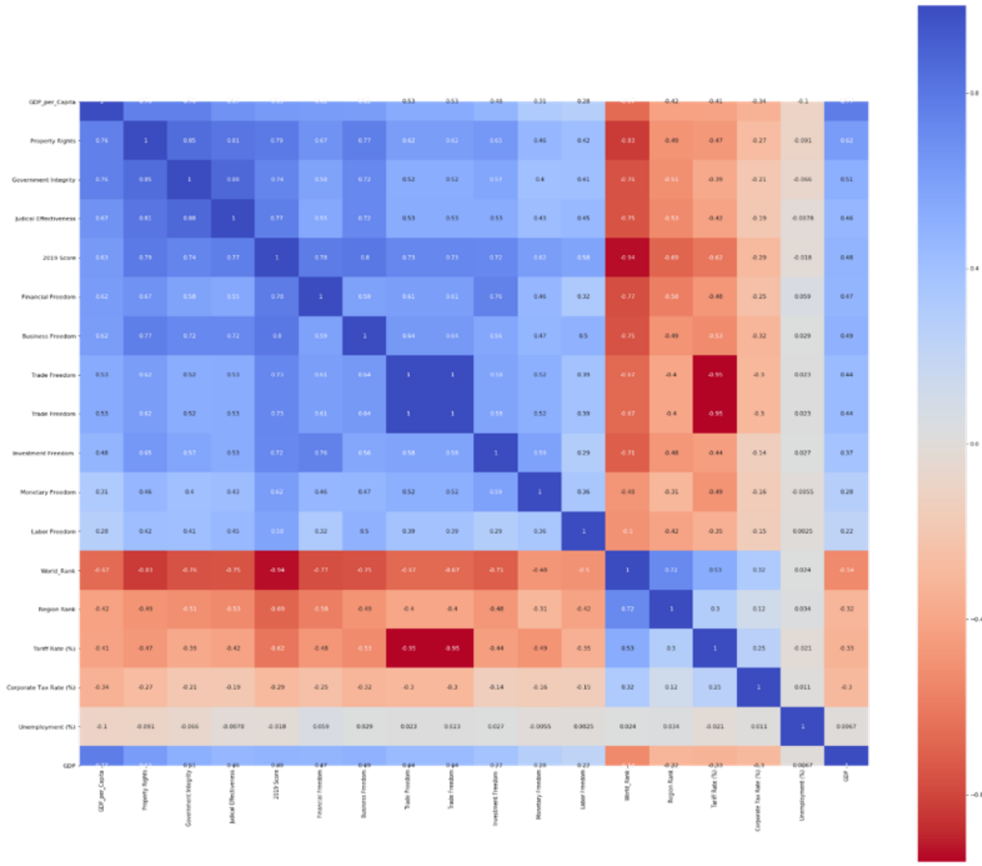



We checked the correlation between GDP per capita and our 12 variables in Economic Freedom Index as well as our other variable unemployment, one by one. As GDP per Capita is actually an aspect of the Index of Economic Freedom, we dropped it from the dataset and calculated high correlation coefficient between the GDP per Capita and the rest of the aspects of the Index of Economic Freedom. It leads to that the Property Rights, Financial Freedom, Trade Freedom, Investment Freedom, Monetary Freedom, Labor Freedom are positively and Unemployment Rate, Corporate Tax Rate, Tariff Rate, Region Rank and World Rank are negatively correlated with the GDP per Capita.




Some variables in our data may be linearly related to each other, which may cause multicollinearity problem for our regression. For example, according to the methodology used to measure these scores, the more effective legal protection of property increases property right scores, which also has an effect in judicial effectiveness and government integrity. Other variables such as fiscal health and government spending may also be highly related to each other. So we need to get rid of this potential multicollinearity problem among the predictors for our regression.

We created a heat-map of correlations of all variables with each other, as presented in Figure 1. We removed the variables that are highly correlated (with 0.7 and over) with property rights, as it seems like the most related to GDP in our calculations. We dropped the variables “Government Integrity”, ”Judicial Effectiveness”, ”2019 Score” and ”Business Freedom”.



## 2.2 Linear Regression Method

Our first method to find the relation between GDP per capita and our variables is regression method. All our variables of interest are continuous variables, so we can use linear regression. By linear regression, we try to fit a linear curve on our data set that will minimize the mean squared errors. We used the features from Economic Freedom Index, as well as other economic indicators such as unemployment as dependent variables. We used GDP per capita as our independent variable. We created a separate training data set and a test data set to calculate the regression to see how our model performs during training. To do this, we used "train_test_split" function from "scikit" library and we split our data into a training set and a test set. We used 20%of our data as test data. With our training and test data, we calculated the multi-variate linear regression in Python. The coefficient results of our regression are presented in figure 2.1.

We found "property rights" feature in Economic Freedom Index gives the highest coefficient when regressed on GDP per capita, while the other variables held constant. Financial freedom and trade freedom features turned also very high coefficients with respectively 242.439665 and 47.727688. On the other hand, unemployment turned an extremely big negative correlation with GDP per capita. Although unemployment and GDP per capita negative relation is pretty intuitive, investment freedom, monetary freedom and labor freedom features have negative high coefficients when regressed on GDP per capita. R2 is the variance of dependent variables that can be explained by the model. R2 of our model is about 73% which is enough to accept our model as significant, in our opinion.

| | coefficients |
|---|---|
| Property Rights | 321.132156 |
| Financial Freedom | 242.439665 |
| Trade Freedom | 95.455376 |
| Investment Freedom | -111.812723 |
| Monetary Freedom | -57.320367 |
| Labor Freedom | -31.992999 |
| World_Rank | 8.066796 |
| Region Rank | 0.382555 |
| Tariff Rate (%) | 7.195379 |
| Corporate Tax Rate (%) | -99.948264 |
| Unemployment (%) | -328.506359 |
| GDP | 15662.609573 |

Figure 2.1: Coefficients of Variables in Linear Regression

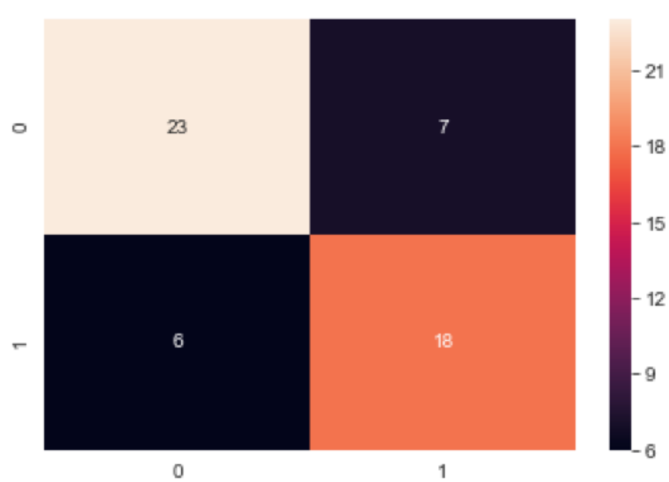
We also calculated an OLS regression using "statsmodel" Python package to compare to our machine learning model. The linear regression table is presented below. We found the coefficients for most parameters and their signs are similar. The R2 value of OLS regression is higher than our model, 0.873 and 0.73 respectively. However, the p values of OLS regression mostly turned extremely high. The only significant parameters were "property rights", "financial freedom" and "unemployment" with p values within 0.05 significance level.

```
                            OLS Regression Results
=====================================================================================
Dep. Variable:          GDP_per_Capita   R-squared (uncentered):                 0.883
Model:                             OLS   Adj. R-squared (uncentered):            0.873
Method:                  Least Squares   F-statistic:                            82.61
Date:                 Mon, 05 Aug 2019   Prob (F-statistic):                  5.39e-55
Time:                         22:12:07   Log-Likelihood:                       -1491.5
No. Observations:                  143   AIC:                                    3007.
Df Residuals:                      131   BIC:                                    3043.
Df Model:                           12
Covariance Type:             nonrobust
========================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Property Rights           322.6242     77.586      4.158      0.000     169.141     476.108
Financial Freedom         239.2182     67.513      3.543      0.001     105.662     372.775
Trade Freedom              11.7557    103.956      0.113      0.910    -193.894     217.405
Investment Freedom       -107.2885     56.104     -1.912      0.058    -218.276       3.699
Monetary Freedom          -64.9730     77.851     -0.835      0.405    -218.981      89.035
Labor Freedom             -36.3544     62.542     -0.581      0.562    -160.077      87.369
World_Rank                  3.5782     35.562      0.101      0.920     -66.771      73.928
Region Rank                -0.5102     88.227     -0.006      0.995    -175.044     174.024
Tariff Rate (%)          -150.3985    208.312     -0.722      0.472    -562.489     261.692
Corporate Tax Rate (%)   -104.3736     88.220     -1.183      0.239    -278.895      70.148
Unemployment (%)         -326.6509    129.885     -2.515      0.013    -583.595     -69.707
GDP                      1.571e+04   1896.749      8.284      0.000     1.2e+04    1.95e+04
==============================================================================
Omnibus:                        21.430   Durbin-Watson:                   2.275
Prob(Omnibus):                   0.000   Jarque-Bera (JB):               29.361
Skew:                            0.829   Prob(JB):                     4.21e-07
Kurtosis:                        4.476   Cond. No.                         479.
==============================================================================
```

## 2.3 Logistic Regression Method

So far, we calculated the correlations between our variables and GDP per capita, one by one, and the multi-variate linear regression of our variables on GDP per capita. But we intend to find not only the effect of single variables, but the groups of variables that explain GDP per capita the best. Therefore, we need to use a classification method. Our first classification method is logistic regression method. The biggest difference of logistic regression method from the linear regression is that it is used on binary variables, which is very useful for classification. We created a binary variable to split our variables into groups as high correlation and low correlation variables. We also created another column named "GDP" as a binary variable for high GDP countries and low GDP countries. We also used training data and test data sets with "scikit" library, taking 30% of data as test data. To find the regression results of the binary variables, we used logistic regression. As the optimal regularization hyperarameter "C" we tried different values ranging from 0.01 to 10. As we used it in a for loop, our parameter "C" is automatically selected giving the best accuracy score. We determined our K number as 5 since our confusion matrix gave the lowest false positive score in case K is 5.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.79 | 0.77 | 0.78 | 30 |
| 1 | 0.72 | 0.75 | 0.73 | 24 |
| accuracy | | | 0.76 | 54 |
| macro avg | 0.76 | 0.76 | 0.76 | 54 |
| weighted avg | 0.76 | 0.76 | 0.76 | 54 |



We found that the group of “property rights”, “financial freedom”, “trade freedom”, and “investment freedom” and a second group of all these variables together with “monetary freedom” give the best results with accuracy score of “0.796”.

```
---------------------------------------------------------------------------------
Optimal Regularization Parameter C:10
['Property Rights', 'Financial Freedom', 'Trade Freedom', 'Investment Freedom ', 'Monetary Freedom']
Average accuracy score on cv (KFold) set: 0.792
Accuracy score on test set is: 0.796
```

Validation and train scores for our model are presented in the graph below. As we can see in the graph, our logistic regression has a training score about 100% of accuracy score and validation score about 98%. Therefore, we can say our model may have an underfitting problem. Our model may give better results with more data. Since the number of countries are limited, more data can be obtained from different years, using time series models.

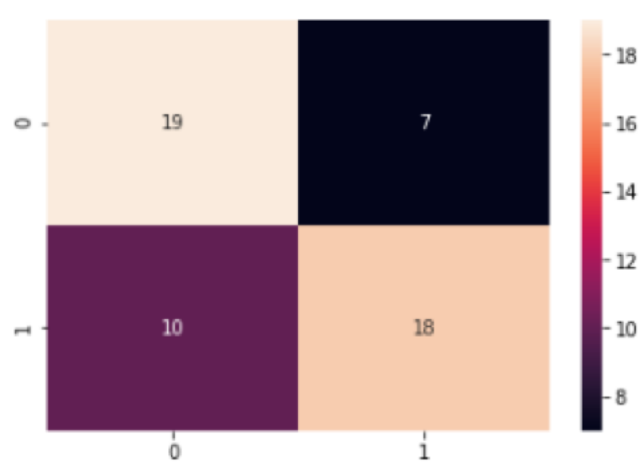

We also used a logit regression from “statsmodel” Python package to compare to our machine learning model. The regression table for logit is presented below. As we can see, the P values for all variables except "property rights" are very high, so we cannot accept them as significant.

```
Optimization terminated successfully.
         Current function value: 0.356960
         Iterations 7
                          Results: Logit
==================================================================
Model:               Logit              Pseudo R-squared:  0.480
Dependent Variable:  GDP                AIC:               58.5517
Date:                2019-08-05 19:09   BIC:               78.4415
No. Observations:    54                 Log-Likelihood:    -19.276
Df Model:            9                  LL-Null:           -37.096
Df Residuals:        44                 LLR p-value:       4.5919e-05
Converged:           1.0000             Scale:             1.0000
No. Iterations:      7.0000
------------------------------------------------------------------
                       Coef.  Std.Err.    z    P>|z|   [0.025 0.975]
------------------------------------------------------------------
Property Rights        0.2157   0.0659  3.2714 0.0011  0.0865 0.3450
Financial Freedom      0.0582   0.0478  1.2180 0.2232 -0.0354 0.1518
Trade Freedom         -0.0683   0.1055 -0.6480 0.5170 -0.2751 0.1384
Investment Freedom    -0.0703   0.0530 -1.3252 0.1851 -0.1742 0.0337
Monetary Freedom      -0.0636   0.0864 -0.7360 0.4618 -0.2328 0.1057
Labor Freedom          0.0059   0.0406  0.1452 0.8846 -0.0737 0.0855
World_Rank             0.0190   0.0243  0.7808 0.4349 -0.0286 0.0665
Region Rank           -0.0658   0.0540 -1.2175 0.2234 -0.1716 0.0401
Tariff Rate (%)       -0.1398   0.2457 -0.5688 0.5695 -0.6214 0.3419
Corporate Tax Rate (%) -0.0050  0.0809 -0.0619 0.9506 -0.1635 0.1535
==================================================================
```

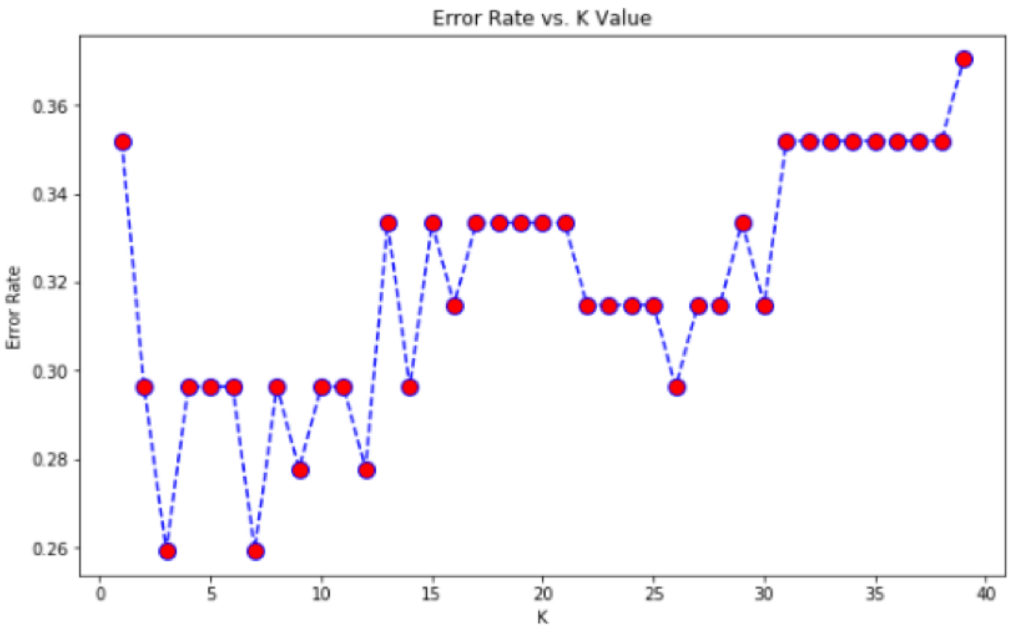
## 2.4 K-Nearest Neighbors Method

Another classification method we used is k-nearest neighbors or kNN method. K-nearest neighbors method uses the shortest distances to k number of values to split data into classes. This method uses entire data to train and is a widely used method for pattern recognition and classification. We used a similar algorithm for kNN method that the function automatically selects the best K number for optimal accuracy score in the range of 1 to 40. We found that "property rights" alone gives the highest accuracy score with 0.815 with k value 27. Accuracy is a good measure for the purposes of our model since we only try to understand the relations between variables and false positive and false negatives do not have significant costs. But we also calculated the precision, recall and f-1 scores for our model shown in the table below. As precision, recall and f-1 scores are all greater than 0.5, we can conclude the model as reliable enough to predict true positives.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.65 | 0.77 | 0.70 | 26 |
| 1 | 0.74 | 0.61 | 0.67 | 28 |
| micro avg | 0.69 | 0.69 | 0.69 | 54 |
| macro avg | 0.69 | 0.69 | 0.68 | 54 |
| weighted avg | 0.69 | 0.69 | 0.68 | 54 |



Our model function used k values between 1 and 40. The mean error rates for each k value is presented in Figure 8. The graph shows that we could not find nearly any value of k that will give us less than 25% of mean error rate. The range of mean error rates of our k values is between 26% and 37%.



# Chapter 3

# Conclusion

As described throughout the paper, there are numerous advancements in both technological and economics-wise affecting agricultural sector and its direct exposure to climate change. However, the key point is not to develop further novelties, but to be able to spread the use of these novelties such as climate-smart agriculture. There are traditional barriers to the use of these technological advancements as well as financial burden of the adaptation of these methods in where the public investment plays a crucial role.

The agricultural sector does not consist of only profit-maximizing and cost-minimizing businesses, but also provides food to the population worldwide. Therefore, it should not be perceived as the other businesses among private sector. The public sector's support for climate-smart agriculture would ultimately help the future of the country and even the world as a whole to become more secure with respect to nutrition and health.

# Bibliography

*https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8*

*Kane, Tim. "2019 Index of economic freedom." (2019).*

*https://stackabuse.com/k-nearest-neighbors-algorithm-in-python-and-scikit-learn*

*https://realpython.com/linear-regression-in-python/*