**Implementing Feature Selection and Machine Learning to Predict Economic Recessions in African Countries**

Introduction

Economic recessions have devastating impacts: companies make fewer sales, people lose work, and the overall standard of living collapses. These deleterious implications are amplified in developing and underdeveloped countries, which are typically unable to provide the social safety nets and economic support necessary to circumvent the detrimental welfare implications of recessions. This is especially true in African countries, which are often dependent on volatile commodities and vulnerable to capital-flight. The ability to forecast recessions in advance would enable the governments of African countries to take the necessary precautionary measures (e.g., monetary policy responses), which may enable them to prevent the recession or mitigate its impacts. Most machine learning-based economic forecasting models have been trained to predict recessions in advanced, western economies and thus do not apply to African countries. Therefore, in this project, we will use feature selection and machine learning to predict economic recessions in African countries.

Data & Data Processing

A dataset composed of 49 different explanatory variables (including macroeconomic indicators and commodity prices) was used. It was curated by blending the Penn World Table Productivity dataset, the Bank of Canada's Commodity Indices, and the World Bank's GDP dataset.First, highly correlated features were removed (one feature from a pair of features with a correlation coefficient greater than 0.75 was deleted) as they increase the model's variance and worsen its performance.Then, the dataset was split based on years. Data from 2000 to 2011 was used as training data and data from 2011 to 2017 was used as testing data. This is preferable to a random split as in the real-world, predicting future recessions using current and past data is what matters.

Feature Selection

To distinguish important features from the noise and irrelevant features, we will employ 3 feature selection algorithms on the training data: recursive feature elimination, LASSO- regression-based selection, and genetic algorithms. This not only improves the model’s performance but also shows policymakers which variables must be closely monitored. The results of the feature selection process are as follows:

Recursive feature elimination (it was made to select for 7 features and was configured to use the random forest algorithm): cwtfp, rkna, hc, rnna, emp\_to\_pop\_ratio, pl\_n, and csh\_c

LASSO: xr, hc, pl\_n, rnna, labsh, forestry\_change, agriculture\_change

Genetic algorithms (it was configured to use the random forest algorithm): cwtfp, rkna, rwtfpna, rnna, emp\_to\_pop\_ratio, irr, labsh, hc

(Note: the names of the features in the dataset and what they mean are mentioned in the appendix)

Creating a Dataset Comprised of the Most Important Variables

I selected variables that had been recommended by at least two of the feature selection algorithms. This triangulation would help me understand the most important determinants of recessions. Using this approach, the following variables were identified as the most important.

1. Emp\_to\_pop\_ratio,

 2. hc

3. cwtfp

4. rnna

5. rkna

6. labsh

7. pl\_n

Training the Models & Evaluation

The models were trained using data from 2000 to 2011 on the most important features. Below, you will find a summary of the accuracies of each model

1. Support Vector Machine: 84.67%
2. Naive Bayes: 65.17%
3. Gradient Boosted Decision Tree: 100%
4. Multiple Logistic Regression: 70.4%
5. Genetic algorithm based random forest: 100%

Summary of Findings

Two of the models, the gradient boosted decision tree and the genetic algorithm based random forest, were able to classify each observation. In other words, they used data from 2000 to 2011 to correctly predict whether or not a recession would occur in any given year from 2011 to 2017 (using the macroeconomic data and commodity prices from the start of that year). However, this does not mean that the model would be able to forecast any recession. For example, it still can’t predict events such as political instability, which are especially important in developing countries.

**Appendix**

**R Code**: <https://github.com/PrathamDave/RecessionPredictionInAfricanCountries>

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** |  |
| pop | Population (in millions) |  |
| emp | Number of persons engaged (in millions) |  |
| emp\_to\_pop\_ratio | Ratio of Employed Persons to Total Population |  |
| hc | Human capital index, based on years of schooling and returns to education; see Human capital in PWT9. |  |
| ccon | Real consumption of households and government, at current PPPs (in mil. 2011US$) |  |
| cda | Real domestic absorption, (real consumption plus investment), at current PPPs (in mil. 2011US$) |  |
| cn | Capital stock at current PPPs (in mil. 2011US$) |  |
| ck | Capital services levels at current PPPs (USA=1) |  |
| ctfp | TFP level at current PPPs (USA=1) |  |
| cwtfp | Welfare-relevant TFP levels at current PPPs (USA=1) |  |
| rconna | Real consumption at constant 2011 national prices (in mil. 2011US$) |  |
| rdana | Real domestic absorption at constant 2011 national prices (in mil. 2011US$) |  |
| rnna | Capital stock at constant 2011 national prices (in mil. 2011US$) |  |
| rkna | Capital services at constant 2011 national prices (2011=1) |  |
| rtfpna | TFP at constant national prices (2011=1) |  |
| rwtfpna | Welfare-relevant TFP at constant national prices (2011=1) |  |
| labsh | Share of labor compensation in GDP at current national prices |  |
| irr | Real internal rate of return |  |
| delta | Average depreciation rate of the capital stock |  |
| xr | Exchange rate, national currency/USD (market+estimated) |  |
| pl\_con | Price level of CCON (PPP/XR) (2011 as base year) |  |
| pl\_da | Price level of CDA (PPP/XR) (2011 as base year) |  |
| pl\_gdpo | Price level of CGDPo (PPP/XR) (2011 as base year) |  |
| csh\_c | Share of household consumption at current PPPs |  |
| csh\_i | Share of gross capital formation at current PPPs |  |
| csh\_g | Share of government consumption at current PPPs |  |
| csh\_x | Share of merchandise exports at current PPPs |  |
| csh\_m | Share of merchandise imports at current PPPs |  |
| csh\_r | Share of residual trade and GDP statistical discrepancy at current PPPs |  |
| pl\_c | Price level of household consumption (2011 as base year) |  |
| pl\_i | Price level of capital formation (2011 as base year) |  |
| pl\_g | Price level of government consumption (2011 as base year) |  |
| pl\_x | Price level of exports (2011 as base year) |  |
| pl\_m | Price level of imports (2011 as base year) |  |
| pl\_n | Price level of the capital stock, price level of USA in 2011=1 |  |
| total | Annual Bank of Canada commodity price index - Total |  |
| excl\_energy | Annual Bank of Canada commodity price index - Excluding Energy |  |
| energy | Annual Bank of Canada commodity price index - Energy |  |
| metals\_minerals | Annual Bank of Canada commodity price index - Metals and Minerals |  |
| forestry | Annual Bank of Canada commodity price index - Forestry |  |
| agriculture | Annual Bank of Canada commodity price index - Agriculture |  |
| fish | Annual Bank of Canada commodity price index - Fish |  |
| total\_change | Year-on-Year Percentage Change Annual Bank of Canada commodity price index - Total |  |
| excl\_energy\_change | Year-on-Year Percentage Change Annual Bank of Canada commodity price index - Excluding Energy |  |
| energy\_change | Year-on-Year Percentage Change Annual Bank of Canada commodity price index - Energy |  |
| metals\_minerals\_change | Year-on-Year Percentage Change Annual Bank of Canada commodity price index - Metals and Minerals |  |
| forestry\_change | Year-on-Year Percentage Change Annual Bank of Canada commodity price index - Forestry |  |
| agriculture\_change | Year-on-Year Percentage Change  Annual Bank of Canada commodity price index - Agriculture |  |
| fish\_change | Year-on-Year Percentage Change Annual Bank of Canada commodity price index - Fish |  |
| growthbucket | "1" = Recession; "0" = No\_Recession |  |