A $75 Billion Dollar Question: Do African Countries Suffer a Systematic Sovereign Credit Rating Bias?

Elikplimi Komla Agbloyor[a,b], Lei Pan[c,d *], Richard Adjei Dwumfour[c], Dennis Nsafoah[e]

[a] *Department of Finance, University of Ghana Business School, Ghana*

[b] *University of Stellenbosch Business School, South Africa*

[C] *School of Accounting, Economics and Finance, Curtin University, Australia*

[d] *Centre for Development Economics and Sustainability, Monash University, Australia*

[e] *Department of Economics and Finance, Niagara University, USA*

Abstract

Using data for 132 countries from 2000 to 2023, we investigate whether there is a systematic credit ratings bias against African countries based on the ratings provided by the major credit rating agencies using machine learning techniques. This is not a trivial question as estimates of the costs of this potential subjectivity are in excess of $75 billion. Following the COVID-19 pandemic and the Russian-Ukraine war, African countries argued that the major credit ratings were very quick to downgrade them. Our empirical analysis offers some credence to the arguments of African countries. We find that during this period, African countries received more adverse ratings compared to other countries. Further, the ratings of African countries were less stable as more non-African countries did not experience changes in their ratings. Indeed, our findings show that the difference or gap in the credit ratings of African countries compared to non-African countries widened after 2015. The results from our machine learning predictions in the full sample, we do not find evidence of a credit ratings bias against African countries as the African dummy does not rank as a top predictor of credit ratings. However, after controlling for sample selection bias, we find strong evidence of a credit ratings bias against African countries. The African dummy increased in importance and was now the number 3 predictor of credit ratings. The African dummy ranked ahead of many important economic, social, political and institutional variables as a predictor of credit ratings.

Keywords: Credit ratings; Africa; Machine learning

JEL Classifications: F34, G12, G24, H63, H81

* Corresponding author.

E-mail address: ekagbloyor@ug.edu.gh (E.K Agbloyor), lei.pan@curtin.edu.au (L. Pan), r.dwumfour@postgrad.curtin.edu.au (R.A Dwumfour), dnsafoah@niagara.edu (D. Nsafoah)

1. Introduction

A recent study by the United Nations Development Programme (UNDP) estimates that African countries could save up to $74.5 billion if credit rating assessments are less subjective in nature. The potential of credit misjudgement, whether intentional or not, has huge consequences for African countries where some of the poorest countries in the world are located. For countries, sovereign ratings (SR), sovereign debt ratings (SDR) or sovereign risk ratings (SRR) - often used interchangeably - are an assessment of their ability to pay their debts in full and on time (Broto and Molina, 2016; Fuchs and Gehring, 2017). Credit ratings have undeniably become a crucial part in determining the conditions within which countries access the global capital markets (Yalta and Yalta, 2018).

In developing countries where there is usually limited information and data, investors have increasingly grown to rely on the opinions of credit rating agencies (CRAs) for sovereign rating/sovereign risk rating (Ferri, 2004). Especially for institutional investors, most are keen on holding securities above a certain rating threshold (investment grade securities) mostly to meet regulatory requirements (Cantor and Packer 1996). This has implications for African countries that have consistently argued that their credit ratings do not reflect their risk exposure. A credit downgrade or a lower credit rating than deserved will increase a country's perceived risk and drive up its cost of borrowing especially during the time of crises (Alsakka and Gwilym, 2013). In some instances, it can entirely close the door to international capital markets as the country can be viewed as too risky leading to a skyrocketing in the cost of borrowing which will effectively preclude a country from being able to borrow.

Following the COVID-19 pandemic, African countries argue that the major credit ratings are very quick to downgrade them. In February 2022, the African Peer Review Mechanism (APRM), which the African Union sets up to provide support to African countries in areas such as credit ratings, denounced the ratings issued by Moody's downgrading Ghana's long-term foreign currency sovereign from B3 to Caa1 (African Union, 2022). This reinforced a long-term held believe that the main foreign ratings agencies, that is Standard's and Poors (S&P), Moody's and Fitch, are systematically biased against African countries. The Government of Ghana complained, and the APRM agreed, that the ratings assigned to Ghana were mainly based on a desk review (the analyst had never visited Ghana), that some key data such as budgetary control measures and upfront fiscal adjustments were not considered, that inaccurate balance of payments data were used in the ratings assessment, that only one primary analyst was assigned to rate a large economy such as Ghana amongst others (see African Union, 2022). In his 2023 State of the Nation Address, the President of Ghana, Nana Addo Danquah Akufo-Addo accused the major ratings agencies of a systematic bias against African countries. Other countries such as Angola, Egypt, Kenya, Nigeria and South Africa have also expressed public discontent with their assigned ratings. As a way of addressing this perceived bias, the APRM announced in March 2022 that the APRM is in the process of exploring the setting up of its own credit rating agency to provide alternative and complementary ratings to those provided by the major credit rating agencies (African Union, 2022).

The credit rating agencies however explain their low ratings of African countries based on factors such as high levels of political risks, low economic development, high levels of borrowing, relatively undiversified economies and limited financial flexibility (Bretton Woods, 2022). Indeed, Ghana subsequently defaulted on both its domestic and foreign currency debt after its credit downgrades. It could be that the ratings downgrade accelerated Ghana's decline into financial bankruptcy. However, credit rating agencies argue that their ratings are through the cycle and reflect the entire economic cycle (both good and bad economic environments) and not just current economic conditions. Consequently, such ratings are supposed to be more accurate and stable thus reflecting the fundamental credit quality of the borrower or obligor.

Following from the hot debate, one may ask: Are credit rating agencies biased in their credit ratings? This question has received some attention in the literature albeit the evidence has remained mixed. For instance, Fuchs and Gehring (2017) use ordinary least squares (OLS) and ordered probit model to examine whether there is bias in the ratings by CRAs for their home countries and find that not only do these major CRAs have bias for their home countries by assigning higher ratings, but they also assign higher ratings to countries that align economically, geopolitically, and culturally with their home countries. Erdem and Varli (2014) estimate the drivers of sovereign credit ratings of S&P for emerging economies and find, among other things, that S&P's rating of Turkey is biased against the country, especially when it comes to giving higher rating levels. In contrast, Alexe *et al.* (2003) use a nonrecursive regression model in examining the determinants of sovereign ratings and found no bias in the ratings by S&P's and Moody's as their prediction of ratings to unrated countries aligned with the actual ratings assigned to the countries by these CRAs. Yalta and Yalta (2018), in a later study, use the seemingly unrelated regressions (SUR) estimation and find evidence of a strong home country bias of these CRAs for the US but no bias against other individual countries. Ozerturk (2014) reveal that CRAs endogenous rating fee is shown to be reducing in the accuracy of the rating. Through comparing efficiency of different equilibria reached in a credit rating game that allows communication between issuer and two credit rating agencies before announcing ratings, Farkas (2021) document that CRAs can learn about each other's signals by exchanging messages with the issuer, conflicts of interest therefore causes CRAs to provide biased ratings. Moreover, CRAs find it optimal to selectively offer biased ratings based on issuer messages when these messages are informative about signals.

It is clear that previous studies have largely been inconclusive as to whether CRAs are biased or not. Most of the studies have tended to use linear regression models such as OLS or ordered probit models - especially due to the discrete nature of the dependent variable owing to the probabilistic nature of the credit ratings. The drawback of using these two methods is that OLS is limited in predicting discrete dependent variables, while the ordered probit models, which are mostly used in the literature assume a specific functional form and distributional assumptions, such as the normal distribution of error terms (Greene, 2017; Wooldridge, 2010). Violations of these assumptions can lead to biased or inefficient estimates. These earlier models are typically limited to modelling linear relationships between predictors and the outcome. They may not capture complex nonlinear patterns present in the data. In ordered probit models, researchers need to manually select and engineer predictor variables (features) that are deemed relevant to the outcome. This process requires some subject selection of variables and may overlook important relationships or interactions that are not explicitly considered.

Our paper differs from the previous studies in that we do not weigh into the debate on the consequences or otherwise of these credit ratings; our focus in this paper is to investigate whether African countries suffer a systematic rating bias or bear an African risk premium or cost. We contribute to the literature by adopting new techniques that are better at predicting outcome variables especially those that are discrete in nature. Specifically, we use machine learning (ML) models such as tree-based, boosting and neural network techniques. The advantages of these models compared to the traditional linear estimation techniques are that: i) whether a country may default on their loans or not is probabilistic in nature and is a prediction problem which machine learning models are better at compared to traditional econometric methods used in the previous studies (see Amini *et al*., 2021); ii) ML and AI models can capture complex nonlinear relationships between predictors and the outcome in the dataset (see Amini et al., 2021; Gu *et al*., 2020). Indeed, many of the macroeconomic variables such as a country's debt-to-GDP ratio, foreign currency reserves, balance of payments (BoP), inflation, interest rates and oil prices may change in a non-linear fashion in relation to the probability of default of their sovereign debt and may also interact in unpredictable ways when a country is in financial distress; iii) ML models can automatically learn feature representations and model interactions, allowing for more accurate predictions (Bishop and Nasrabadi 2006; Hastie *et al*., 2009). Therefore, our approach allows us to identify and untangle the complex, high-dimensional, and interactive effects that exist between the features that predict a country's credit risk rating. In our paper, we include as one of the key explanatory variables a dummy reflecting whether a country is based in Africa or not. To the best of our knowledge, no studies so far have used ML techniques to assess whether CRAs are biased or not against developing countries, for that matter, Africa. This paper aims to fill this gap in the literature.

The rest of the paper is organized as follows. Section 2 describes the dataset and presents the methodology used in this study. Section 3 discusses the empirical findings. Section 4 concludes and also provides policy recommendations.

2. Data and Methodology

2.1 Data

Our data is obtained from numerous sources. The data on credit ratings was sourced from Thomson Reuters Datastream. We sourced data on all three major ratings agencies, namely, S&P, Moodys and Fitch. We retrieved credit ratings for all 132 major economies across the world.[1] We constructed an African dummy which took on a value of one for African countries

[1] List of countries is provided in Appendix A.

and zero otherwise. In general, the data sources include data on debt levels, reserves, financial development, economic growth, institutional quality, infrastructural quality, banking stability, commodity price volatility, inflation and so on. We obtained data on other features or predictors of credit ratings from numerous sources. The sources of the data are detailed in Appendix B.

2.1.1 Train/Validation/Test data split

We split the data into train, validation and test sets to enable us judge the performance of our machine learning models on unseen or test data. We employ a Train/Validation/Test ratio split of 60/20/20. In addition, we perform Cross Validations by adopting the K-Folds method and a K of 10 on the training data. Consequently, the training data was split into 10. This implies that the machine learning algorithms trained or learnt on 9 (K-1) folds of the data and evaluated performance on 1. Consequently, instead of obtaining one measure of performance, we had K or 10 metrics of performance. High cross validation scores signal that the algorithms are stable, consistent, have low variance and are likely to replicate similar performance on unseen or test data.

2.2 Machine learning algorithms

The present study adopts various machine learning algorithms to predict the credit ratings of countries. The algorithms employed include Logistic Regressions, Bagging, Random Forest, AdaBoost, and Gradient Boost. The model selection is based on the models with the highest Recall score, the consistent models based on the other evaluation metrics (mainly Precision and the F1 scores), and finally, the models with the least over-fitting. Machine learning models are known to overfit the data quickly. The consequence of overfitting is that the machine learning models perform extremely well on the training data but have lower performance on unseen data or test data.

2.2.1 Synthetic minority over-sampling techniques (SMOTE)

We use the Synthetic Minority Over-Sampling Technique (SMOTE) to increase the number of observations for classes that have low numbers of observations. This is important as classes with low observations are unlikely to be predicted accurately as the machine learning models will be biased towards classes with higher numbers or observations.

2.2.2 Hyper-parameter tuning

Finally, we perform hyper-parameter tuning to search over the sample space and to optimize our models by selecting the best parameters. Hyper-parameter tuning in essence is a form of optimization and can help to improve or extract extra levels of performance from the machine learning models. The hyper-parameters are specified by the modeller as they cannot be learnt from the data. The parameters to be tuned include the learning rate, the number of estimators, the tree depth, minimum sample leaves, max features, and maximum sample size.

2.3 Model performance evaluation

We use the confusion matrix to help evaluate the performance of the algorithms. The confusion matrix gives the number of True Positives[2], True Negatives[3], False Positives[4] and False Negatives[5]. Based on the metrics from the confusion matrix, we obtained the Accuracy ratio, the Recall Score, the Precision Score and the F1 Score. The Recall score tells us the percentage of actual credit ratings that the models can predict. Recall is defined as:

$$Recall = \frac{True\ Positives}{\text{True Positives} + \text{False Negatives}} \tag{1}$$

The Precision Score on the other hand tells what percentage of our predictions of countries in a particular rating category are actually correct. Precision is defined as:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{2}$$

Finally, we obtained the F1 Score which is a harmonic mean of Precision and Recall. The F1 score is defined as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

The F1 score ranges between zero and one. A value of one represents perfect recall and prediction.

2.4 Feature importances

Finally, to understand the main drivers or predictors of credit ratings, we obtain the feature importance. These extracted features are the features or variables that the machine learning models suggest are important in predicting credit ratings. The feature importances, in essence, assign rankings and weights to each predictor/feature/independent variable. Our hypothesis is that the African dummy is an important predictor of credit ratings. This suggests the presence of a bias against African countries as the continent should ideally not be a key predictor of credit ratings of countries.

3. Discussion of Results

In this section, we present the results of our empirical analysis. We first of all begin by presenting some stylized facts. Then we proceed to present results from the algorithms estimated. Finally, we present the feature importances which show the top predictors of credit ratings.

3.1 Distribution of credit ratings

[2] This represents the prediction of those who fall in a rating category and actually did.
[3] This represents our prediction of countries who would not fall in a rating category and indeed they did not fall in that rating category.
[4] This represents the prediction of countries who fall in a rating category, but they did not.
[5] This represents the prediction of those who would not fall in a rating category but actually fell in that category.

Our preferred measure of credit ratings is the Fitch ratings because it has the largest observations based on the sample data. We use forward filling and also backfilling to fill in missing values. Though employing backfilling has some disadvantages in the sense that a country is likely to improve its economic and fiscal management when it is in the process of seeking a credit rating and consequently the first observed credit rating may not necessarily reflect previous ratings were they to be available, we believe that the benefit of having a balanced panel which is required for the algorithms we employed outweigh this disadvantage. In addition, this treatment is adopted for all countries thus it does not confer an advantage or disadvantage to any particular country.

Due to the fact that some original categories did not have sufficient observations, we developed a new rating scale, as shown in Table 1. We have more data per rating category when we employ this approach, making a prediction of a category more reliable. However, the cost is that it is difficult to detect small changes in credit ratings. For example, a country rated BB+ will have to drop by 3 notches (BB, BB- and B+) to B+ before we detect a rating downgrade in our data set. Consequently, we have less variation in the ratings dataset.

[Insert Table 1 Here]

Figure 1 shows the distribution of the credit ratings. The lowest category in terms of the number of credit ratings are 0, 1 and 7. Thus, as expected, countries with substantial risk of default or where default is imminent (0 and 1) have some of the lowest count of observations. Categories 5, 6 and 7, representing Upper Medium Grade, High Grade and Prime countries, have the next highest counts. This is not surprising as well as the number of countries with a high credit rating should not also be very high. The bulk of countries with average ratings (4, 3 and 2 representing Lower Medium Grade, Non-Investment Grade and Highly Speculative) have the highest counts. Even though the shape is not a nicely shaped bell distribution, it gives the sense of a normal distribution since most of the mass is in the middle and less of the mass is in the distribution's tails. From Figure 1, we see that more countries have a higher credit rating compared to those who have a low credit rating pulling the mean to be above the median. The average country (based on the mean) in our sample is rated non-investment grade based on our categorization. The box plot does not show the presence of outliers in our ratings data.

[Insert Figure 1 Here]

Figure 2 compares credit ratings between African countries and non-African countries. Panel A shows a visual representation of credit ratings by continent. The map clearly shows that Africa and South America tend to have low ratings whilst, for example, North America and Oceania have high credit ratings. In Panel B, the figure clearly shows, as expected, that non-African countries have an average higher credit rating (4 = lower medium grade) compared to African countries (between non-investment grade and highly speculative). It is interesting to note that the gap in the difference in the average credit rating is not constant over time. In particular, the gap or difference in credit ratings seems to have widened after the year 2015 due mainly to a lower average rating for African countries. The difference in average ratings is statistically significant (*p*-value of 0.000) based on the Mann-Whitney U test.

[Insert Figure 2 Here]

To better understand the differences in ratings between African and non-African countries, we compute the proportion of upgrades, no changes in ratings and downgrades for these countries. The results are illustrated in Figure 3. In terms of downgrades, a slightly higher proportion of African countries (0.03/3%) compared to non-African countries (0.026/2.6%) are downgraded over the sample period. The two sample proportion Z test shows that there is no significant difference between downgraded African and non-African countries (*p*-value = 0.525). In terms of upgrades, a higher proportion of non-African countries (0.03/3%) compared to African countries (0.01/1%) are upgraded over the sample period. Indeed, the two sample proportion Z test shows that there is a significant difference between upgraded African and non-African countries (*p*-value = 0.008). That is, over the sample period, more non-African countries experienced upgrades compared to African countries. However, in terms of no rating actions, African countries (0.917/91.7%) experienced more non-changes in their ratings compared to non-African countries (0.903/90.3%). This difference, though, is not statistically significant (*p*-value = 0.256).

[Insert Figure 3 Here]

3.2 COVID-19/Russia-Ukraine war and rating actions across African and non-African countries

We are interested in investigating whether there are systematic differences in rating actions between African and non-African countries during the COVID-19 and the Russia-Ukraine war. In part, this is motivated by the observed widening of spreads since 2015. In addition, the voices of ratings bias against African countries are loudest during the COVID-19 crisis and the Russia-Ukraine war. The COVID-19 pandemic and the Russia-Ukraine war are seismic shocks to the global financial system. We contribute to the debate by showing whether African countries received different ratings compared to non-African countries during the period of COVID-19 and the Russia-Ukraine war. Our analysis does not necessarily imply or answer questions about causality. To examine this question, we interact three key variables. These are COVID-19 and Russia-Ukraine war dummy (measured as 1 for the year 2020, 2021, 2022 and 2023 and 0 otherwise), an African dummy (measured as 1 for African countries and 0 otherwise) and a ratings change variable (upgrades measured as 1 for upgrades and 0 otherwise, downgrades measured as 1 for downgrades and 0 otherwise and no change measured as 1 for no change in ratings and 0 otherwise). For instance, an African country (the African dummy takes on the value of 1) downgraded (the downgrade variable takes on the value of 1) during the COVID-19 and Russia-Ukraine war (the COVID-19/Russia-Ukraine war variable takes on the value of 1) will assume a value of 1. Consequently, this variable will take on a value of 1 because it represents an African country downgraded during the COVID-19/Russia-Ukraine war period.

3.3 Are there systematic differences between African and non-African countries downgraded during COVID-19/Russia-Ukraine war

The Chi2 test (*p*-value = 0.000) shows a statistically significant difference between African and non-African countries that were downgraded during the COVID-19 pandemic and the Russia-Ukraine war. To investigate this further, we extract the count and proportions of African and non-African countries that were downgraded during the period of COVID-19 and the Russia-Ukraine war. The results are presented in Table 2. We can see that 8 African countries had ratings downgrades whilst 13 non-African countries experienced a downgrade. In terms of proportions, 7.4% of African countries experienced downgrades whilst 3.1% of non-African countries experienced downgrades. This is statistically significant with a *p*-value of 0.04. Consequently, we can conclude that significantly more African countries were downgraded during the period of COVID-19 and the Russia-Ukraine war.

[Insert Table 2 Here]

3.4 Are there systematic differences between African and non-African countries upgraded during COVID-19/Russia-Ukraine war

For upgrades (Chi2 test *p*-value = 0.006), we find a statistically significant difference between African and non-African countries during the period of COVID-19 and the Russia-Ukraine war. The results are presented in Table 3. It can be seen that 3 African countries had ratings upgrades whilst 8 non-African countries experienced an upgrade. In terms of proportions, 2.8% of African countries experienced upgrades whilst 1.9% of non-African countries experienced upgrades. This is not statistically significant (*p*-value of 0.57). Consequently, we can conclude that there is no difference between African and non-African countries which are upgraded during the period of the COVID-19/Russia-Ukraine war regarding the proportions of countries.

[Insert Table 3 Here]

3.5 Are there systematic differences between African and non-African countries with no ratings actions during COVID-19/Russia-Ukraine war

For no ratings changes (Chi2 test *p*-value = 0.000), our findings suggest that there is a statistically significant difference between African and non-African countries during the period of COVID-19/Russia-Ukraine war. The results are reported in Table 4. Our results show that 97 African countries have no ratings changes whilst 399 non-African countries have no ratings changes. In terms of proportions, 90% of African countries experienced no changes whilst 95% of non-African countries experienced no changes. This is statistically significant (*p*-value of 0.044). Consequently, we can conclude that a higher proportion of non-African countries experienced no ratings changes during the COVID-19/Russia-Ukrain war period.

[Insert Table 4 Here]

Taken together, the results suggest that African countries are received more adverse ratings compared to other countries during the COVID-19 pandemic and the Russia-Ukraine war. African countries, in terms of proportions, received more downgrades and fewer changes in their ratings.

3.6 Empirical results

Table 5 below shows the data we have in the training, validation and training data set. The class distribution on the training, validation and test data are similar and like the overall distribution in the overall data.

[Insert Table 5 Here]

3.6.1 Model building: original and over sampled

We present the results from the estimations using the original data. This data does not include a treatment for class imbalance. Class imbalance can affect the predictions, especially for under-represented classes. Figure 4 presents results from using the original data. We see that the Ada Boost estimator produces the lowest cross validation (CV) score. The Logistic estimator also underperforms the other estimators. The Random Forest estimator, on the other hand, produces the highest CV score. The average CV score for the bagging, random forest and gradient boost estimators is about 90%.

[Insert Figure 4 Here]

3.6.2 Model building: synthetic minority over sampling technique (SMOTE)

Given that we have a class imbalance problem, with some classes having low observations, we employ the synthetic minority over-sampling technique (SMOTE) to balance the data. Table 6 shows the distribution of credit ratings for the various classes. Class 0, Class 1 and Class 7, for example, have only 37, 109 and 89 observations, respectively, which could make predicting these classes unreliable. After applying SMOTE, each class now has 691 observations which is the observation of the largest class (Class 2 – Highly Speculative).

[Insert Table 6 Here]

After employing SMOTE, as shown in Figure 5 the performance of the three top estimators (bagging, random forest and gradient boost) improves slightly. The CV score has now moved from an average of about 90% to an average of about 95%. The performance of the Logistic regression estimator also improved. However, the performance of the Ada Boost estimator deteriorated after applying SMOTE. The best estimator is still the random forest estimator.

[Insert Figure 5 Here]

To extract an extra ounce of performance, we perform hyper-parameter tuning. We focus on tuning only the Random Forest, Bagging and Gradient Booster estimators. Table 7 presents the best parameters for our machine learning algorithms. These parameters are the ones we used in estimating our final algorithms. Figure 6 shows the evaluation metrics after running these models. We settled on the Random Forest because it has some of the highest evaluation metrics and is stable and consistent across both the validation and test data. The feature importances are, therefore, based on the Random Forest.

[Insert Table 7 and Figure 6 Here]

3.6.3 Feature Importances

In Figure 7, we present the feature importances from the full sample. The feature importances show the various predictors of the credit rating in terms of the importance of each variable in explaining the variance of the credit ratings variable. The feature importances are based on the Random Forest estimator. The quality of institutions is the biggest predictor of a country's credit rating. We proxy institutions by using the average of the six indicators (namely rule of law, regulatory quality, control of corruption, government effectiveness, voice and accountability, and political stability and absence of violence) sourced from the World Governance Indicators. The level of development of the financial sector (measured by financial market and financial institution development variables from the International Monetary Fund) is also another key predictor of credit ratings. The level of education, the labour force participation rate, central bank independence, the level of foreign exchange reserves, infrastructural development (measured by telephone lines), financial stability (non-performing loans), official development assistance, trade and the level of natural resource endowment and the extent of business disclosures turn out to be important predictors of a country's credit rating. Other predictors are the level of unemployment, debt service, savings, total debt, gross fixed capital formation, the current account balance, inflation, the real exchange rate, FDI, the level of external debt, gross domestic product (GDP) growth, US monetary policy, and financial market volatility. Interestingly, though Africa is not a top predictor (it is ranked number 32), it ranks above COVID-19, Highly Indebted Poor Countries (HIPC), the Russia-Ukraine war, and fragile and conflict countries as predictors of credit ratings.

[Insert Figure 7 Here]

Given that we observed a widening of spreads between African and non-African countries after 2015, we restricted our sample from 2015 to 2023 and retrained all our algorithms. The Feature Importances of this exercise is presented in Figure 8. We note that, the importance of the African dummy increases. The African dummy (now ranked 27) makes it into the Top 30 predictors as opposed to the Top 35 predictors previously.

[Insert Figure 8 Here]

Next, we retrain our algorithms using credit ratings from Moodys. The results are presented in Figure 9. We did not use ratings from S&P as an alternative source of credit ratings because it has the least observations amongst the ratings obtained from the three major ratings agencies. The results obtained using the Moodys credit ratings are qualitatively similar to that from Fitch. The top driver of credit ratings from Moodys is also the quality of institutions. The level of reserves increases in importance when we employ Moody's ratings. In the case of Moodys, the African dummy ranked as the the number 31 predictor of credit ratings.

[Insert Figure 9 Here]

Finally, in our last experiment, we address potential sample selection biases by randomly selecting a number of non-African countries that is equal to the number of African countries. The full sample had 132 countries made up of 106 non-African countries and 26 African countries. This introduces potential sample selection biases as the weight of non-African countries in the sample far outweighs the weight of African countries in the sample.

Consequently, to balance the sample, we included all 26 African countries and randomly selected 26 non-African countries. This approach consequently reduces the risk of overweighting non-African countries or underweighting African countries in the sample.

The results from this experiment are startling. This experiment provides the strongest evidence of an African bias in credit ratings. Africa now ranks as the number 3 predictor of credit ratings. The African dummy ranks ahead of many important economic, social, political and institutional variables such as financial stability, financial development, financial market volatility, natural resource endowment, debt levels, trade, inflation and economic growth.

[Insert Figure 10 Here]

4. Conclusion

In this paper, we investigate whether there exists a systematic ratings bias against African countries. This question is important as a potential bias can lead to very high borrowing costs for African countries. A downgrade (unfair) can lead to a death spiral and precipitate a default. For instance, once an African country is downgraded, investors are likely to sell their holdings of Eurobonds of the country because a number of investors, such as institutional investors, can only hold investment-grade bonds. An unfair downgrade can also effectively shut a country out of the international capital markets, compounding the country's woes. In addition, a downgrade (unfair) can lead to capital flow reversals leading to an exchange rate crisis and a banking crisis due to a rise in nonperforming loans (NPLs) because a depreciating currency can lead to imported inflation, higher cost of borrowing, lower consumer spending and reduce the ability of borrowers to service their loans. These signs were witnessed by countries such as Egypt, Ghana and Zambia during COVID-19. To examine the question at hand, we assemble credit ratings data for 132 economies from 2000 to 2023. We adopt machine learning techniques such as logistic regressions, ada boost, bagging, gradient boost and random forests. We use SMOTE to improve the precision of our predictions. We also employ hyper-parameter tuning to improve the performance of our models.

As expected, we find that the average rating of African countries is lower than non-African countries. The average rating of non-African countries has a score of 4, representing a lower medium grade in our data. On the other hand, the average rating of African countries has a rating of between 2 and 3 representing a non-investment grade/highly speculative investment based on our scaling. We also find that the gap or difference in credit ratings between African and non-African countries widened after 2015. Further, we investigated whether there is a significant difference in the ratings between African countries and non-African countries during the COVID-19 crisis and the Russian-Ukraine war. This is motivated by the loud 'cries' of a ratings bias by African countries during this period. Our findings show that a significantly larger proportion of African countries compared to non-African countries are downgraded during this period. In addition, we find that more non-African countries do not experience a change in their ratings compared to African countries suggesting that African countries have less stable ratings during this period.

Our study contributes to the debate on whether African countries suffer a systematic ratings bias. Based on the feature importances from the full sample, we find that the most important predictors of the credit ratings include the quality of institutions, financial development, energy use, cost of business start-up procedures, education, labour force participation, central bank independence, reserves, infrastructure and financial stability. Given that we observed a widening in the ratings between African and non-African countries from 2015, we retrain the algorithms on data for all countries by sub-setting the period from 2015 to 2023. The findings reveal that the feature importances do not change much. However, the African dummy's rank improved from Number 32 (Top 35 predictors) to Number 27 (within the Top 30 Predictors). Furthermore, we use data from an alternative source to ensure that our results are not driven by the choice of data from a particular ratings agency. In particular, we retrain the models using data from Moodys instead of Fitch. The results are qualitatively similar suggesting that our results are robust to the choice of rating agency. Finally, we controlled for potential sample selection biases by matching the number of non-African countries with the number of African countries. We did this by randomly selecting 26 non-African countries to match with our list of 26 African countries. Once we did this, we found very strong evidence of a credit ratings bias against African countries. The African dummy increased in importance and was now the number 3 predictor of credit ratings. The African dummy ranked ahead of many important economic, social, political and institutional variables such as financial stability, financial development, financial market volatility, natural resource endowment, debt levels, trade, inflation and economic growth.

Future studies should endeavour to entangle the question of causality and the direction of causality. Our findings should be interpreted with caution as machine learning models do not necessarily imply causality. Moreover, we use publicly available quantitative data to predict credit ratings. Rating agencies have access to non-public data and use some qualitative and subject assessments in their ratings. Consequently, a bias may still exist but it would be difficult to detect from publicly available data. From our findings, both African countries and credit rating agencies may have something to hold on to, though none may be happier!

On the policy front, we believe that regional credit rating agencies or credit rating agencies formed by economic blocks, such as the proposed credit rating agencies by the African Union and the proposed Credit Rating Agencies by the BRICS group, may be a good idea. These rating agencies are more likely to understand the regional or economic context of the countries that they cover. These rating agencies may also lead to the development of regional or economic financial centres. However, it is likely that global investors or investors from the Global South may view these agencies as offering highly favourable or undeserved ratings to the regions, countries, or companies that they cover. Overall, though, we believe that the advantages of regional credit rating agencies will outweigh the cost.

References

African Union. (2022). https://au.int/ar/node/41477. Retrieved on 20th June 2023.

Alexe, S., Hammer, P. L., Kogan, A. and Lejeune, M. A. (2003). A non-recursive regression model for country risk rating. *RUTCOR-Rutgers University Research Report RRR*, *9*, 1-40.

Alsakka, R. and Gwilym, O. (2013). Rating agencies' signals during the European sovereign debt crisis: market impact and spillovers. *Journal of Economic Behaviour & Organization*, 85, 144-162.

Amini, S., Elmore, R., Öztekin, Ö., & Strauss, J. (2021). Can machines learn capital structure dynamics?. *Journal of Corporate Finance*, *70*, 102073.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning 1st Edition*. Springer.

Boroto, C. and Molina, L. (2016). Sovereign ratings and their asymmetric response to fundamentals. *Journal of Economic Behaviour & Organization*, 130, 206-224.

Bretton Woods. (2022). Perceived versus real risk in African credit ratings. Retrieved from https://www.brettonwoods.org/article/perceived-versus-real-risk-in-african-credit-ratings

Cantor, R. and Packer, F. (1996). Determinants and Impact of Sovereign Credit Ratings. *FRBNY Economic Policy Review*, 2(2), 37–49.

Erdem, O. and Varli, Y. (2014). Understanding the sovereign credit ratings of emerging markets. *Emerging Markets Review*, *20*, 42-57.

Farkas, M. (2021). Competition, communication and rating bias. *Journal of Economic Behaviour & Organization*, 189, 637-656.

Ferri, G. (2004). More analysts, better ratings: do rating agencies invest enough in less developed countries?. *Journal of Applied Economics*, *7*(1), 77-98.

Fuchs, A. and Gehring, K. (2017). The home bias in sovereign ratings. *Journal of the European Economic Association*, *15*(6), 1386-1423.

Greene, W. H. (2012). *Econometric analysis 7th Edition*. Pearson Education.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223-2273.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. (2nd ed.). Springer.

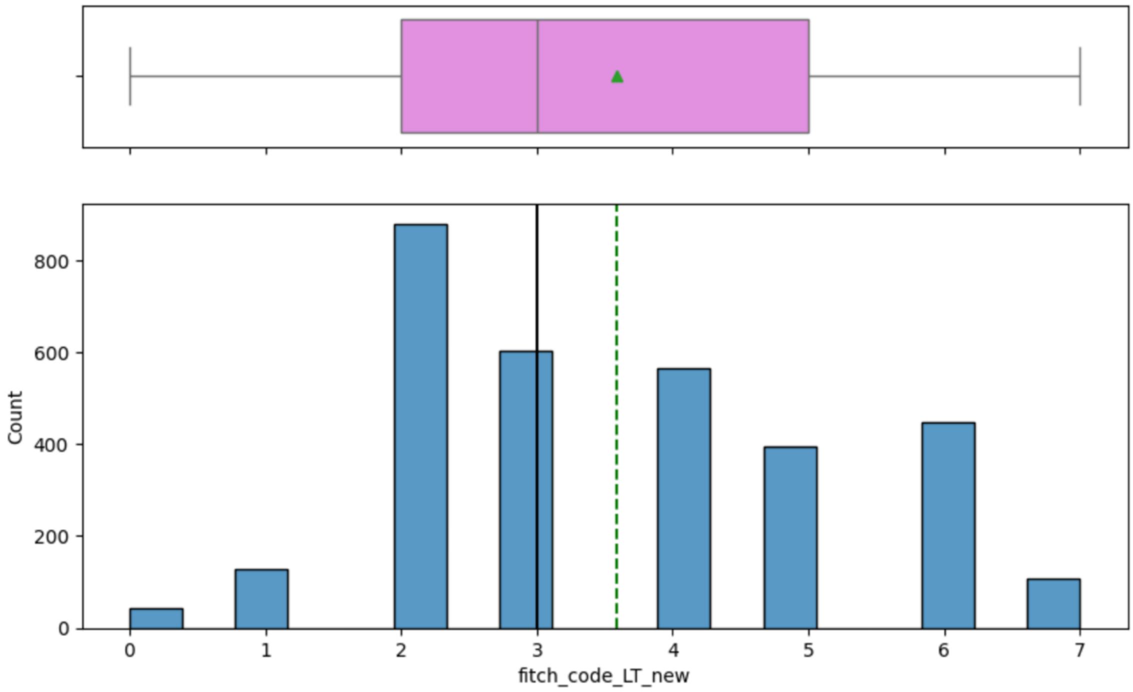
Ozerturk, S. (2014). Ratings as regulatory stamps. *Journal of Economic Behaviour & Organization*, 105, 17-29.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT Press.

Yalta, A. T. and Yalta, A. Y. (2018). Are credit rating agencies regionally biased? *Economic Systems*, *42*(4), 682-694.

Figures

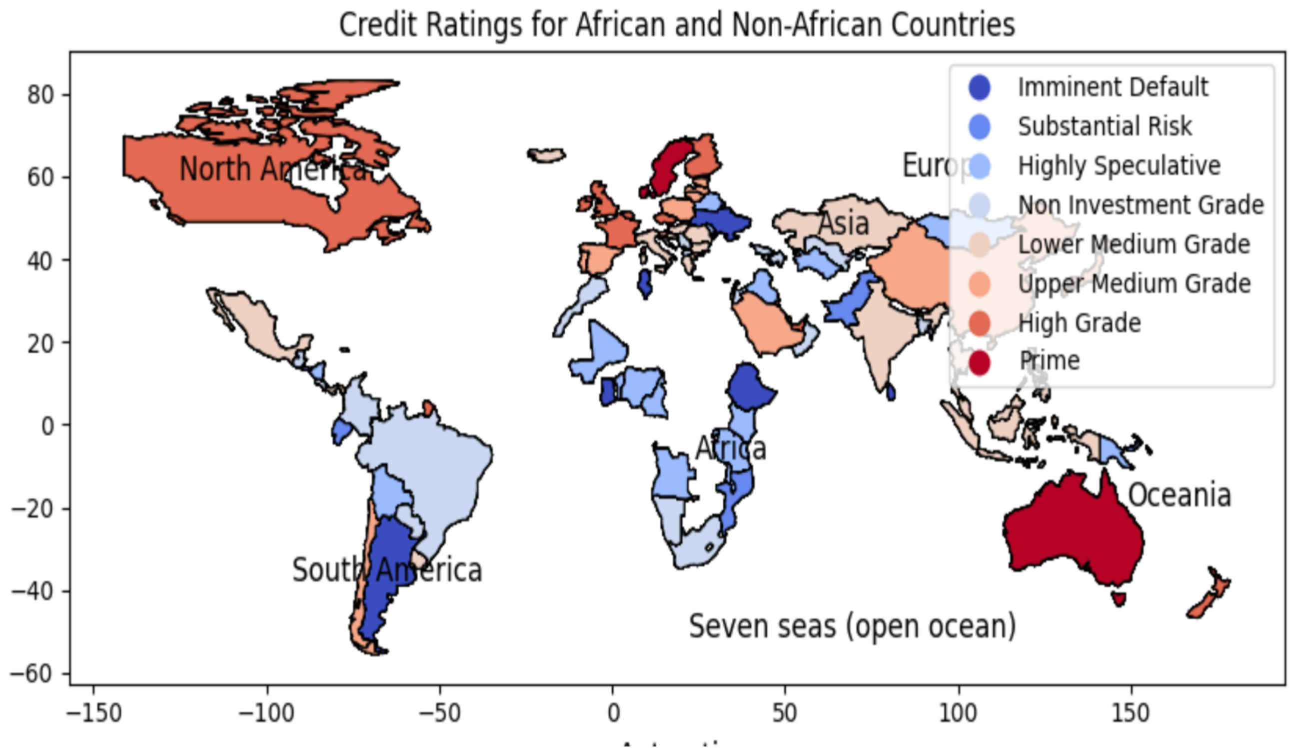
Figure 1: Credit ratings across the full sample



Note: fitch_code_LT_new stands for Fitch Long-term Issuer Rating.The green line represents the mean whilst the black line represents the median.

Figure 2: Comparison of credit ratings between African and non-African Countries

Panel A: Geographic Visually of Credit Ratings



Panel B: Credit Ratings Over Time and Mean Credit Ratings

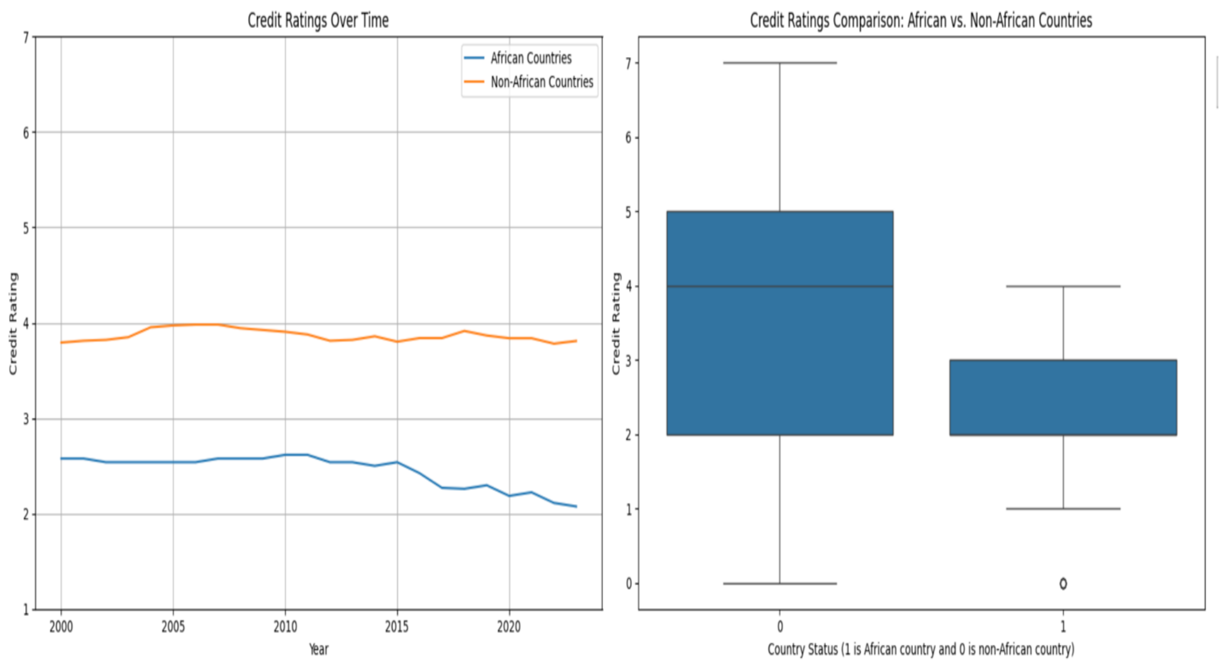


Figure 3: Rating changes: Africa vs. non-African countries

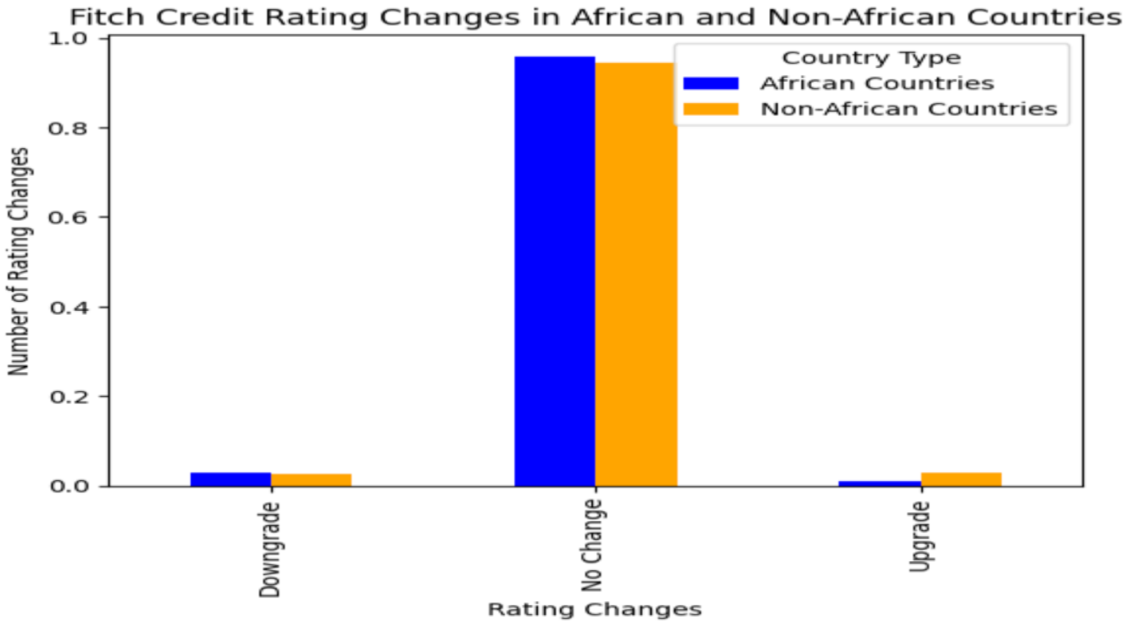


Figure 4: Model predictions using the original data

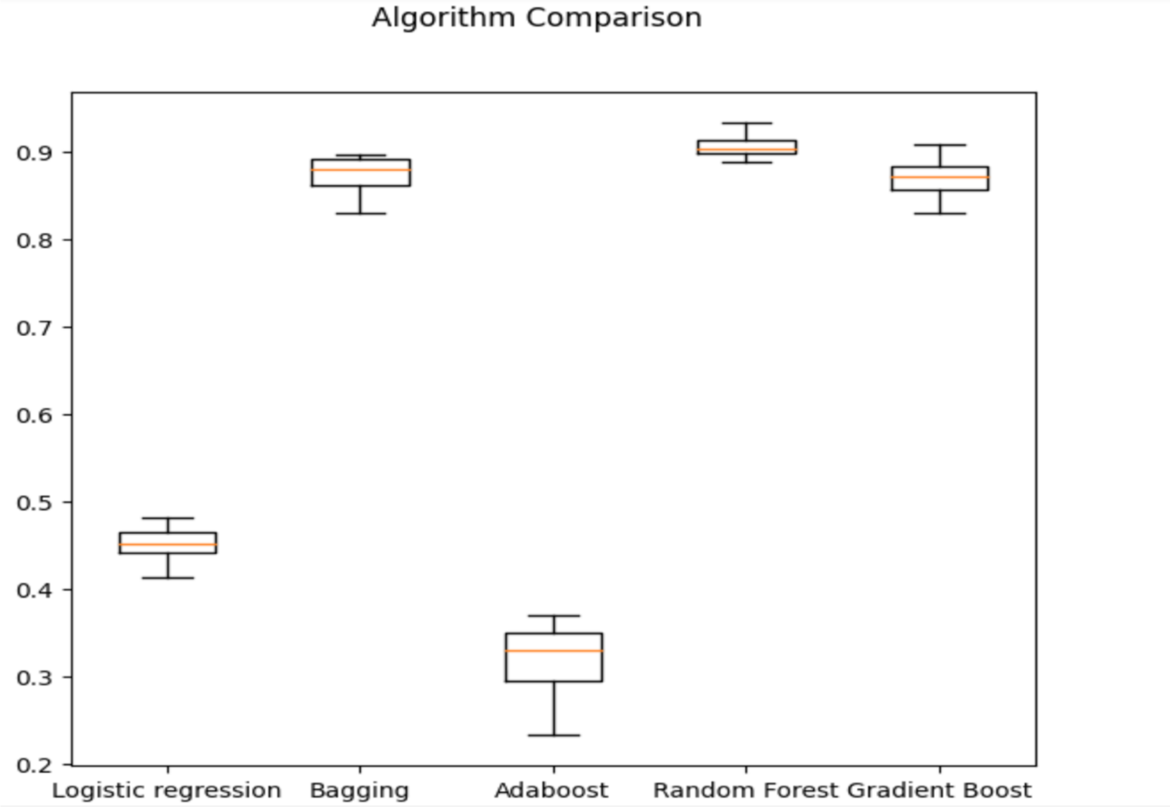


Figure 5: Model predictions using the over sampled data

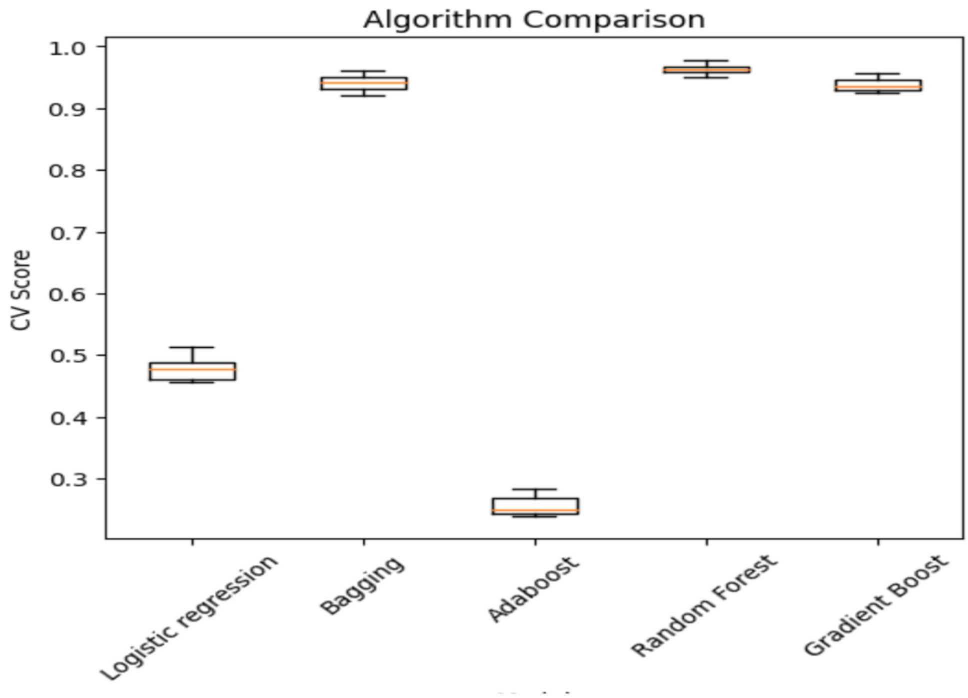
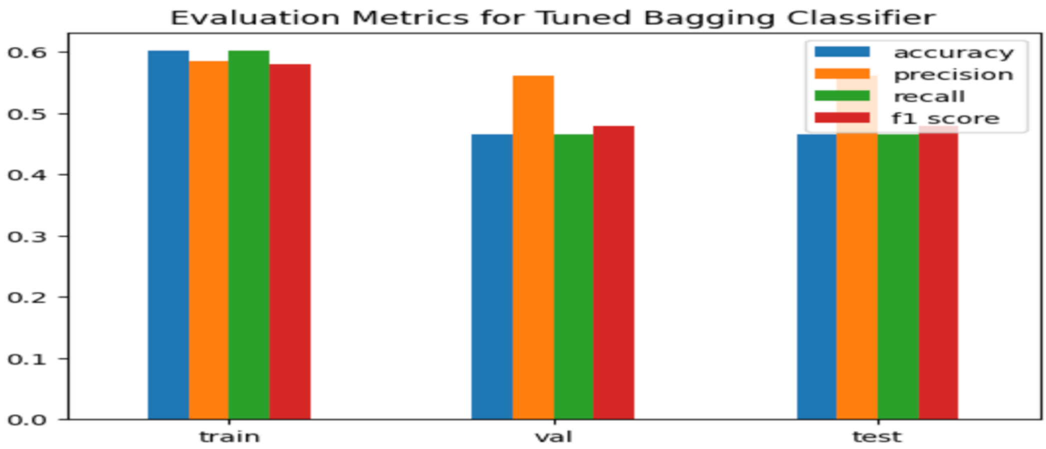
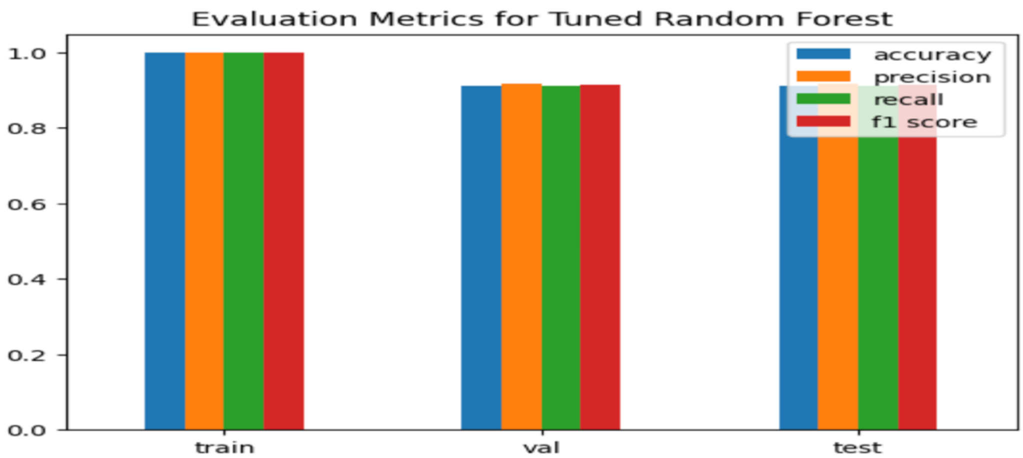


Figure 6: Evaluation metrics after hyper-parameter tuning

Panel A: Bagging



Panel B: Random Forest



Panel C: Gradient Boost

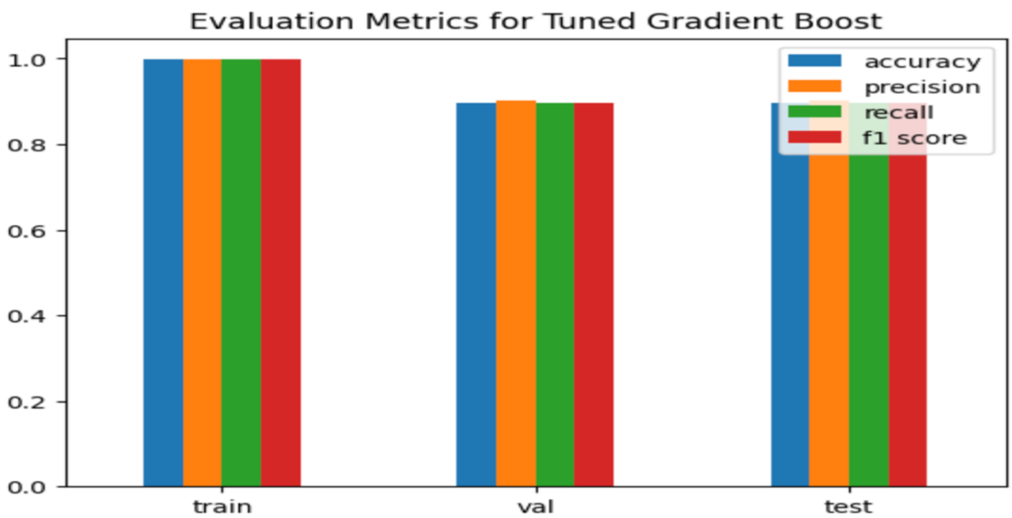


Figure 7: Feature importances: full sample based on Fitch credit ratings

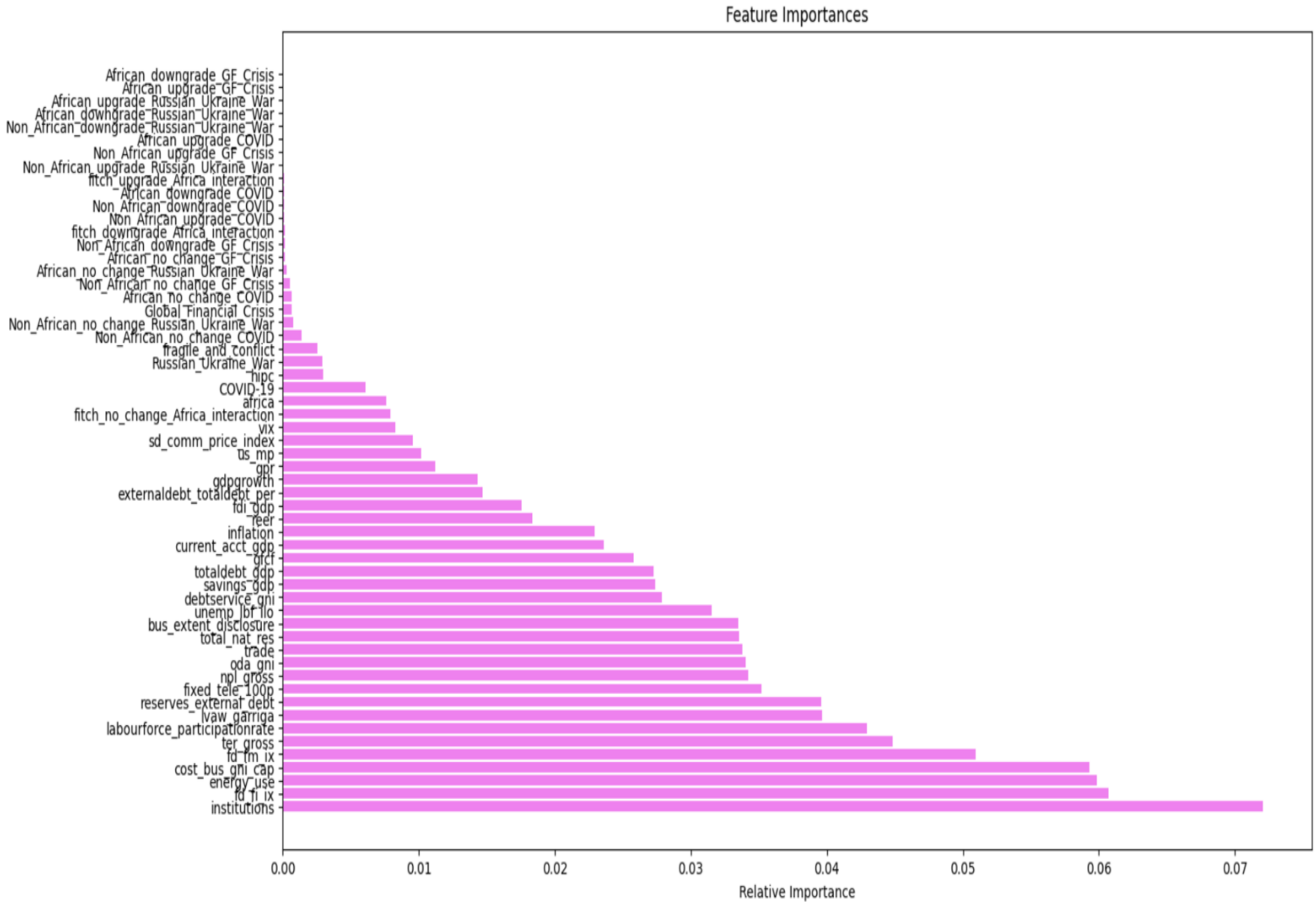


Figure 8: Feature Importances for sub-sample (2015 – 2023)

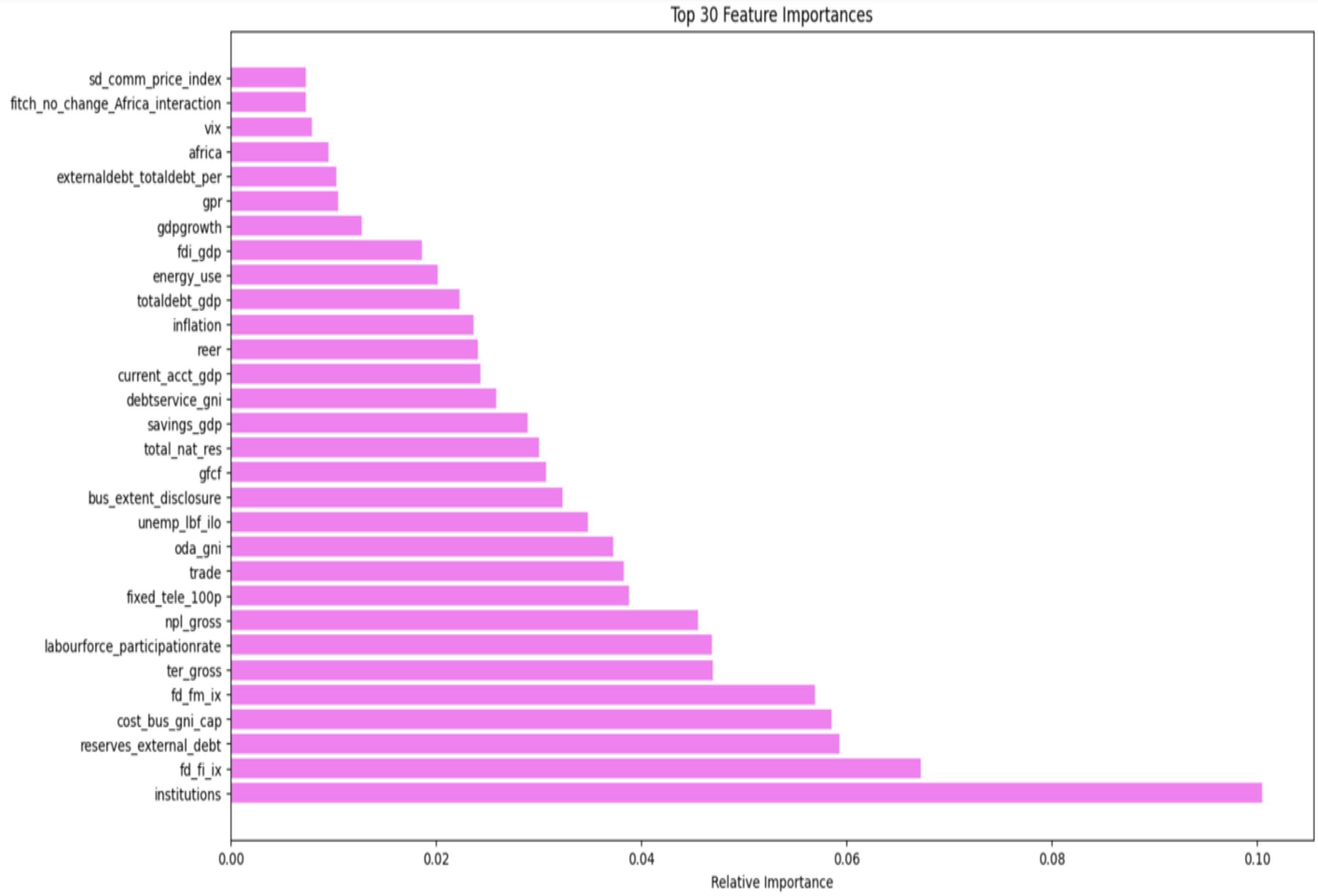


Figure 9: Feature Importance using Moodys credit ratings

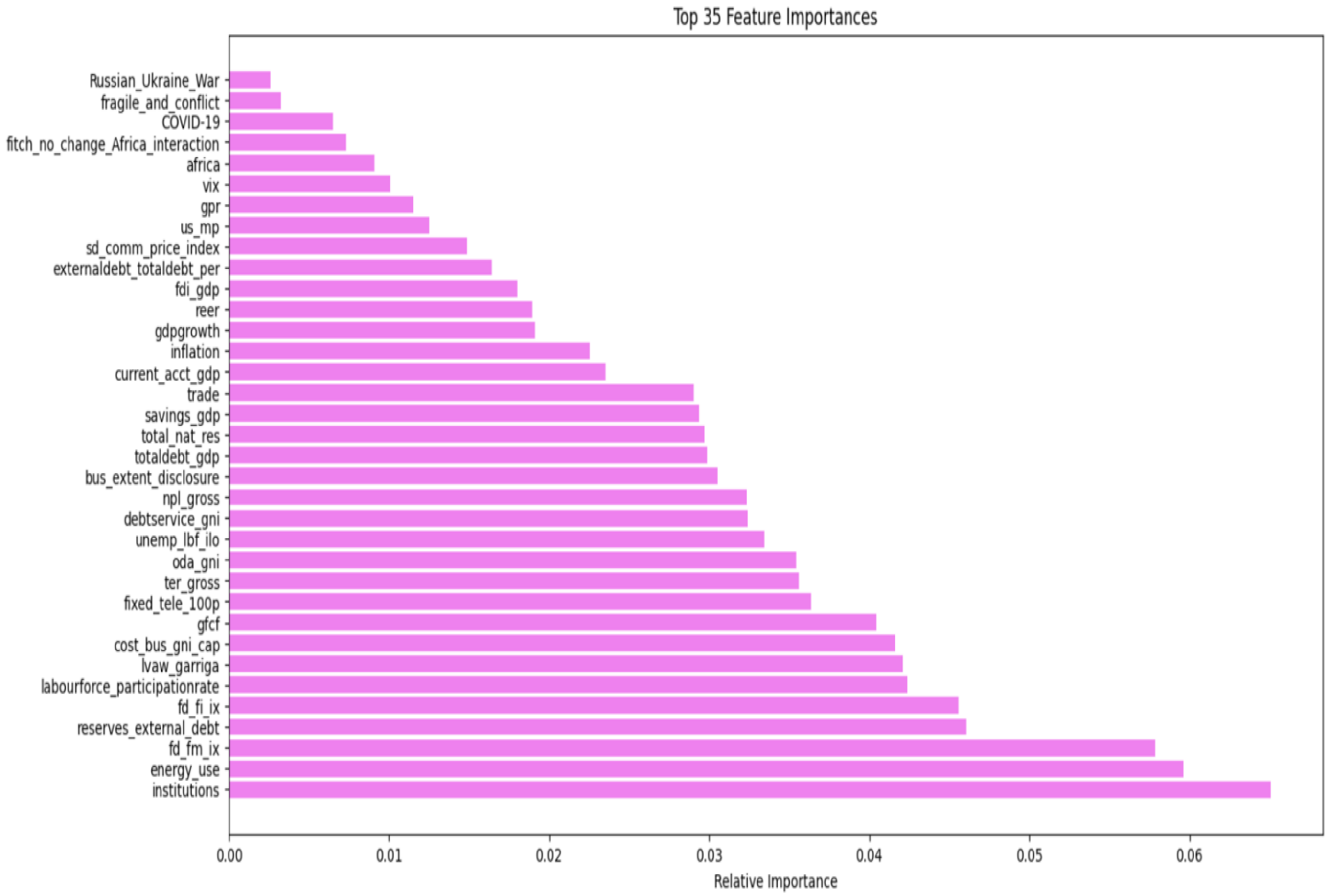
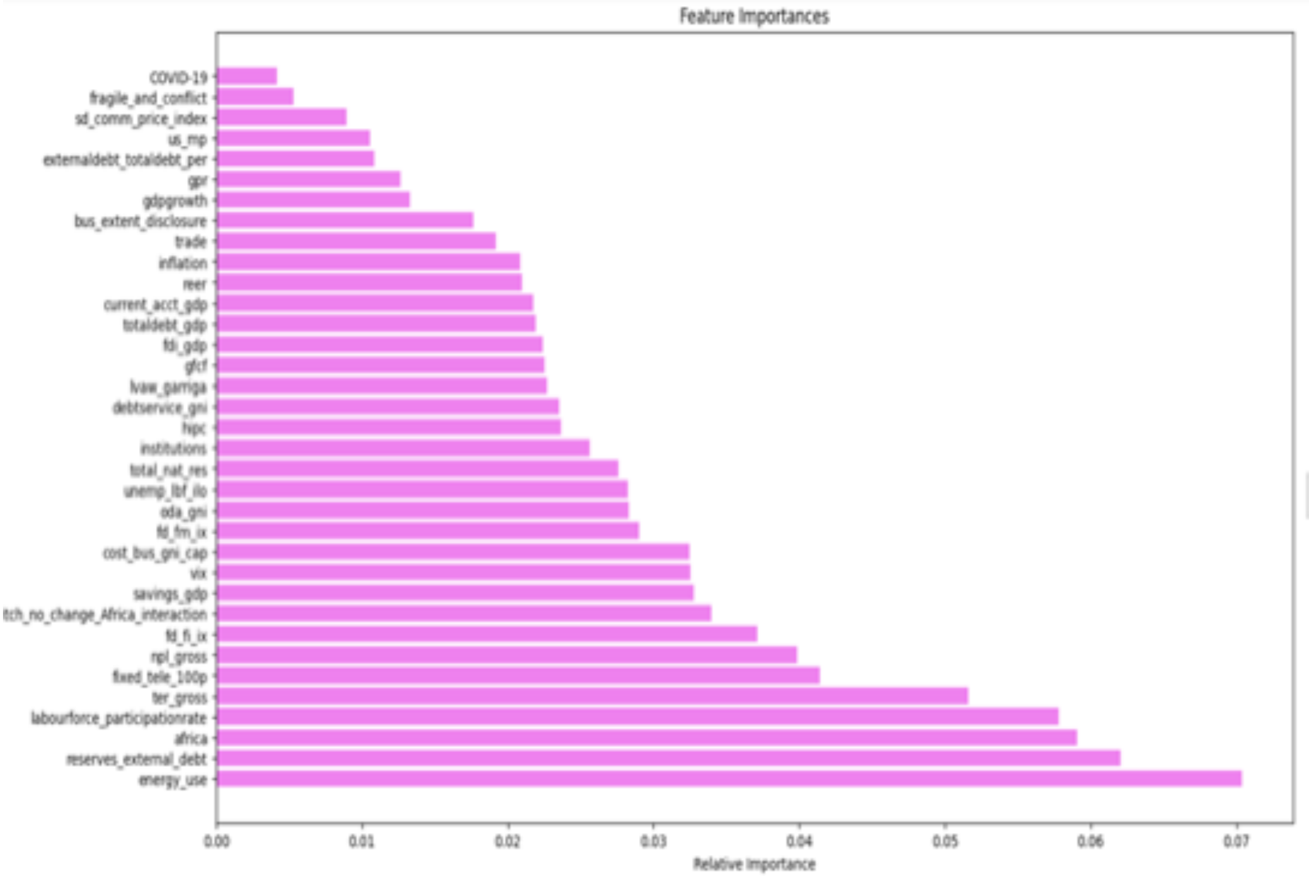


Figure 10: Feature Importances after randomly selecting non-African countries



Tables

Table 1: Fitch ratings and categorization

Rating Value	Original Fitch Category	Meaning in Our Dataset
7	AAA and Aaa	Prime
6	AA+, AA and AA-	High Grade
5	A+, A and A-	Upper Medium Grade
4	BBB+, BBB and BBB-	Lower Medium Grade
3	BB+, BB and BB-	Non-Investment Grade
2	B+, B and B-	Highly Speculative
1	CCC+, CC	Substantial Risk of Default
0	CCC-, CC and C	Imminent Defaultt with little Prospect for Recovery
-1	D, SD	Default

Table 2: Downgrades of African and non-African countries during the COVID-19/Russia-Ukraine war period

	Count of Downgraded	Proportion of Downgraded
African Countries	8	0.074
Non-African Countries	13	0.031
Test of Difference (*p*-value)		0.04

Table 3: Upgrades of African and non-African countries during the COVID-19/Russia-Ukraine war period

	Count of Upgraded	Proportion of Upgraded
African Countries	3	0.028
Non African Countries	8	0.019
Test of Difference (*p*-value)		0.571

Table 4: No changes in ratings of African and non-African countries during the COVID-19/Russia-Ukraine war period

	Count of No Change	Proportion of No Change
African Countries	97	0.898
Non African Countries	399	0.95
Test of Difference (*p*-value)		0.044

Table 5: Training, validation and test data characteristics

	Full Sample	Training Data	Validation Data	Test Data
Shape (Rows/Columns)	3168/57	1900/57	634/57	634/57
Class 0	0.014	0.015	0.013	0.011
Class 1	0.04	0.036	0.065	0.03
Class 2	0.278	0.267	0.289	0.298
Class 3	0.19	0.201	0.166	0.185
Class 4	0.178	0.182	0.194	0.151
Class 5	0.124	0.125	0.101	0.145
Class 6	0.141	0.137	0.144	0.151
Class 7	0.034	0.037	0.030	0.028

Note: Class 0 - Imminent Default with little Prospect for Recovery; Class 1 - Substantial Risk of Default; Class 2 - Highly Speculative; Class 3 - Non-Investment Grade; Class 4 - Lower Medium Grade; Class 5 - Upper Medium Grade; Class 6 - High Grade; Class 7 – Prime.

Table 6: Comparison of data before and after the application of SMOTE

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Original Data	37	109	691	486	469	302	351	89
After SMOTE	691	691	691	691	691	691	691	691

Table 7: Hyper-parameter tuning: best parameters

Model	Bagging	Random Forest	Gradient Boost
Learning rate			0.2
N Estimators	200	300	125
Max Depth	3		
Minimum Sample Leave		1	
Max Samples	0.5	0.6	
Max Features	0.5	sqrt	0.5
Sub-Samples			0.7