INTERNATIONAL JOURNAL OF ECONOMICS AND FINANCE STUDIES
Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102
Received: 23.11.2023 | **Accepted:** 28.12.2023 | **Published Online:** 02.01.2024

-RESEARCH ARTICLE-

COMPARING THE PREDICTION PERFORMANCE OF MACHINE LEARNING ALGORITHMS AGAINST THE LOGISTIC REGRESSION MODEL IN FORECASTING SOVEREIGN DEBT RATINGS

Oliver Takawira

Senior Lecturer: Department of Finance and Investment Management (DFIM), College of Business and Economics (CBE) - University of Johannesburg, South Africa, and Visiting Professor (Researcher): School of Management – University at Buffalo – New York, USA.
ORCID ID: https://orcid.org/0000-0001-7515-1733
Email: otakawira@uj.ac.za

─Abstract─

This study examines the performance of fourteen (14) Machine Learning (ML) techniques in comparison to the traditional statistical methodology, Logistic Regression (LR), for analysing and predicting future sovereign debt ratings (SDR). The negative impact of speculation and pessimistic expectations regarding sovereign ratings can have detrimental effects on macroeconomic indicators and lead to disruptions in the monetary or fiscal system, ultimately resulting in financial instability. This situation highlights the importance for developing nations to anticipate and forecast changes in sovereign debt ratings to mitigate the negative consequences of a downgrade. By analysing macroeconomic variables and SDRs in South Africa's Quarterly data from 1999 to 2022, this study aims to determine the model with higher precision and superior analytical capacity. The data for South Africa's sovereign debt rating was obtained from the three major debt rating agencies (DRAs), including Moody Ratings, Fitch Ratings, and Standard & Poor's. Macroeconomic indicators were obtained from Thomson Reuters, the South African Reserve Bank (SARB), Statistics South Africa (Stats SA), and Quantec Easy Data. The data was divided into a test set and a training set, with a ratio of 75:25, respectively. The comparison between statistical and ML techniques was conducted using performance measurements such as accuracy, sensitivity, specificity, precision, and the area under the curve (AUC). The

Citation (APA): Takawira, O. (2024). Comparing The Prediction Performance of Machine Learning Algorithms Against the Logistic Regression Model in Forecasting Sovereign Debt Ratings. *International Journal of Economics and Finance Studies, 16*(01), 23-61. doi: 10.34109/ijefs.202416102

study discovered that ML techniques exhibit superior precision and outperform traditional statistical models. However, the effectiveness of these techniques is contingent upon the specific analysis and data employed. ML techniques offer a wide range of possibilities, but the excessive reliance on assumptions in traditional models can hinder their performance. The study highlighted that DRAs employ various methodologies and variables when evaluating sovereigns.

Keywords: Sovereign Debt Rating, Machine Learning Algorithms, Logistic Regression and Macroeconomic Indicators.

JEL Classifications: C51, C52, C53, C58, G17, G24.

INTRODUCTION

Econometric models are specifically crafted to analyse and draw conclusions about the connections between different variables. On the other hand, machine learning (ML) models are primarily focused on accurately categorising and making predictions. Artificial intelligence involves the use of software applications to improve accuracy in classifying inputs and forecasting outputs without explicit programming (Takawira, 2022). ML techniques use historical data to predict future values, while econometric modelling formalises relationships between data variables using mathematical equations. Despite their shared focus on the question of how we learn from data, both fields approach it from different perspectives. However, machine learning is a relatively new discipline that does not depend on traditional statistical methods to analyse the relationships between variables. The primary focus of the study is to address the issue of determining the most suitable approach for predicting sovereign debt ratings (SDRs) by comparing traditional statistical models with new machine learning models.

Murphy (2012) defined "machine learning as the set of systems that can automatically detect data patterns, and then utilize the uncovered patterns to forecast future output values, or to perform other kinds of decision making under uncertainty -such as planning how to collect more data." One drawback of the econometric model compared to machine learning is the requirement to have a thorough understanding of various factors, such as the type and source of data, statistical indicators like p-values, unbiased estimators, estimator properties, the distribution of the underlying population, and the expectations properties when conducting repeated experiments. This study compares an older statistical method with machine learning in analysing the determinants of sovereign debt ratings. Machine learning focuses on optimisation, handling high-dimensional data, and performance, while statistics is more concerned with inference and low-dimensional data. Forecasting future SDRs can help governments implement policies that enhance macroeconomic performance and

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

facilitate rating upgrades.

An academic perspective on sovereign debt ratings (SDR) would highlight that these ratings serve as a measure of a government's capacity to fulfil its payment obligations, whether they are related to domestic or international financial debt. Debt Rating Agencies (DRAs) analyse a country's fundamental information and financial behaviour to determine its ability to meet its obligations or repay its debts. Chee et al. (2015) emphasised the importance of DRAs in the credit market, particularly for the purposes of debtor monitoring and information provision. The most widely recognised DRAs are Standard & Poor’s (S&P), Moody’s, and Fitch Ratings. These rating agencies are highly regarded in the industry for their expertise in rating sovereigns, organisations, and companies (Chee et al., 2015).

Investors, creditors, and other economic agents closely monitor credit ratings for both sovereigns and companies as a crucial measure of creditworthiness. Despite some researchers and market participants contesting the weaknesses in rating activities, such as insignificance, inconsistency, lack of oversight authority, and integrity, credit ratings remain a significant factor in decision-making (Afik et al., 2014; Baum et al., 2016; Bedendo et al., 2018; Dallara, 2008). Countries that are rated have limited influence over the ratings they receive, leaving them susceptible to potential exploitation by rating agencies. Downgrades in Africa often seem to be influenced by political events, which can lead African governments to perceive a bias in how ratings reflect a country's economic and financial standing. The connection between ratings and political events can lead to the perception that rating agencies are meddling in the politics of the countries they rate. Opponents of rating agencies and their practices have suggested the creation of a rating agency by the BRICS[1] union. This agency would aim to address the perceived biases and hidden agendas of existing rating agencies (Kräussl, 2005). Regrettably, rating agencies and their services are inevitable, and countries cannot afford to disregard them as investors and other economic participants consider published ratings when making investment decisions (Gu et al., 2018; Iannotta et al., 2013).

In their studies, Bellotti et al. (2011) highlighted the difficulties faced by DRAs due to the lack of clarity in their methodologies and their inability to predict major global financial crises, such as the 1990 Asian crisis. In their research, Bellotti et al. (2011) highlighted the tendency of rating agencies to prioritise historical events and information over prospects when assigning ratings. The 2008 financial crisis was a result of the excessive reliance and unwavering trust placed by economic agents on rating agencies and sovereign debt ratings (Ozturk et al., 2016). Rating agencies play

[1] BRICS countries:1- Brazil, 2-Russia, 3-India, 4-China, and 5-South Africa 6-Egypt, 7-Ethiopia, 8-Iran, and 9-the United Arab Emirates.

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

a crucial role in fulfilling a specific financial need, with sovereign debt ratings often serving as a benchmark for other ratings.

Typically, local firms receive lower ratings compared to their country's credit rating (Iyengar, 2010). In a study conducted by Ozturk et al. (2016), it was observed that sovereign debt ratings have a significant impact on the interest rates of assets and serve as benchmark indicators for measuring credit risk in other assets. As a result, they have a direct influence on the volume and diversity of investment assets. Debates surrounding the accuracy of sovereign ratings have led to an increased focus on developing internal credit scoring systems to decrease dependence on rating agencies and their debt rating services (Ben Mim et al., 2023; Ozturk et al., 2016). There has been a growing focus by governments on sovereign debt ratings to enhance access to international lenders or capital markets and reduce borrowing costs. Sovereign debt ratings evaluate the capacity of both governments and other borrowers in a country to fulfil their financial obligations (Iyengar, 2010). Hence, sovereign ratings offer an overview of the default risk associated with a borrowing country.

Governments now rely on sovereign debt rate forecasting to assess the perceived risk of default by lenders, funders, and investors (Kabadayi & Çelik, 2015). The debt crisis was utilised by rating agencies to increase borrowing costs, as evidenced by the publication of rating outlooks. This, in turn, led to financial instability (Polito & Wickens, 2015). Developing nations are compelled to engage in credit ratings forecasting activities to mitigate the negative consequences of downgraded sovereign debt ratings. South Africa, being an emerging and developing market with a volatile political climate, is susceptible to the negative consequences of sovereign debt rating downgrades. Speculation and negative expectations regarding sovereign ratings can have a detrimental impact on macroeconomic indicators and disrupt the monetary or fiscal system, leading to financial instability. Government entities, central banks, and policymakers actively work to prevent financial instability by promoting sustainable financial stability (FS). The potential for financial instability arises from the downgrades in sovereign debt ratings (Ozturk et al., 2016).

This study applied logistic regression and ML techniques to analyse and forecast SDR in their real ordinal format contradicting previous literature that converted ratings to numerical values (Bennell et al., 2006; Kräussl, 2005; Kumar & Haynes, 2003). The models applied in this study includes the Gradient Boosting Classifier (gbc), Decision Tree Classifier (dt), K-Nearest Neighbors Classifier (knn), , Random Forest Classifier (rf), Extra Trees Classifier (et), Extreme Boosting Gradient (xgboost), Ada Boost Classifier (ada), Support Vector Machine (svm), Linear Kernel, Logistic Regression (lr), Light Gradient Boosting Machine (lightgbm), Naïve Bayes (nb), Ridge Classifier (ridge), Quadratic Discriminant Analysis (qda), Dummy Classifier (dummy), and the Linear Discriminant Analysis (lda). This study aimed to

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

identify a model that accurately predicts SDRs to help governments adjust policies and systems. The goal is to prevent rating downgrades and promote rating upgrades, ultimately restoring financial stability (Bennell et al., 2006; Cantor & Packer, 1994; Kräussl, 2005; Kumar & Haynes, 2003). This study provides valuable insights for academics seeking to compare the performance of machine learning models with traditional econometric and statistical models, specifically the logistic regression model.

Evolution of Debt Rating

According to Kabadayi and Çelik (2015) "the first debt reporting service was established by Louis Tappan in 1841 aiming to examine the ability of traders in paying back their financial obligations." In 1890, John Moody assessed the performance of railway corporations in the United States following his foray into evaluating industrial bonds. In 1916, Poor's Company introduced the first credit or debt rating, which was later followed by Fitch and Standard Statistic Company in 1924 with their initial ratings. In 1941, the merger of Standard Statistic Company and Poor's Company resulted in the formation of the Standard and Poor's (S&P) Corporation. The emergence of international ratings by rating agencies coincided with the expansion of finance companies into global markets (Cantor & Packer, 1994). In 1949, the United States received its first sovereign rating as a country from Moodys. S&P followed suit in 1975 by publishing its first sovereign debt rating for the USA, and Fitch joined in 1994 by rating its first sovereign (Takawira, 2022). "Sovereign debt or credit ratings determine the capacity of the government to meet its financial debt obligations and the DRAs basically calculate the sovereign's capacity to pay back loans and award an ordinal score. Debt Rating Agencies (DRAs) consider various factors, such as the political situation, financial status, infrastructure, economy, previous debt, country's production capacity, and other relevant information that assess the country's risk profile and its ability to repay its financial debts on time (Saadaoui et al., 2022). Dunne et al. (2007) have shown that sovereign assets play a crucial role in valuing other financial assets, known as price discovery. Additionally, their ratings serve as a benchmark for institutions operating within the economy.

The Three Big Credit Rating Agencies

Moody's Investors Services

Moody's Corporation specialises in bond and debt rating through its division known as Moody's Investors Services, commonly referred to as Moody's. Pİrdal (2017) states that Moody's is the second largest DRA globally, founded by John Moody in 1900. It offers comprehensive research on international finance and bonds issued by

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

corporations or governments. In July 2019, Moody's downgraded South Africa to a rating just above junk status and expressed concerns about the country's poor performance and government interference at Eskom, a government-owned entity responsible for supplying power and electricity.

Fitch Rating Inc.

Fitch Ratings Inc., established in 1914 by John Knowles Fitch, is one of the three major debt rating agencies, albeit the smallest one (Pİrdal, 2017). Following a cabinet reshuffle by the President of South Africa in 2017, Fitch, like Moody's, downgraded South Africa's sovereign debt rating to junk status.

Standard & Poor’s Financial Services LLC

Standard & Poor's Financial Services LLC (S&P), a division of Standard & Poor’s Global, is an American company that specialises in conducting financial research and analysis for stocks, bonds, and commodities. In the 1860s, S&P Global was established by Henry Varnum Poor (Pİrdal, 2017). S&P is the largest credit rating agency and provides debt ratings for both public and private debt, including government entities and parastatal organisations. S&P provides debt ratings for both the interim and long-term periods. S&P downgraded SA to junk status following the Cabinet reshuffle of the former President of South Africa in 2017.

Rating Terminologies

Rating agencies such as S&P, Moody's, and Fitch utilise similar terminology, though not the same, when they publish ordinal bond rating scores. The ordinal rating scores are represented using different combinations of letters, including letters with positive or negative signs and letters with numerical figures. The top rating for S&P and Fitch is represented by 'AAA', while for Moody's it is 'Aaa'. Rating classifications such as 'AAA' to 'BBB' are commonly referred to as investment grade, while 'BB' to 'D' are classified as speculative grade. The table provided in the appendices presents the terminologies utilised by rating agencies.

Background

The rating market is dominated by three major rating agencies: S&P, Moody's, and Fitch. This has led to an oligopoly market structure, which some market participants criticise for its lack of competition and transparency. The impact of sovereign debt rating agencies on emerging markets is widely recognised, as they play a crucial role in determining the availability of loans and the stability of these markets (Kräussl, 2005). In a thought-provoking inquiry, Osobajo and Akintunde (2019), and Overes and van der Wel (2023) have delved into the factors that credit rating agencies (DRAs) employ to assess debt ratings. Interestingly, these agencies have assigned varying credit ratings to the same organisation in certain instances. DRAs are often criticised for using the issuer paying approach to downplay risks and cater to the desires of issuers to boost their ratings. This has resulted in a situation where ratings are artificially inflated and incidents such as mis-rating practices have brought rating agencies under scrutiny (Vu et al., 2022).

In a study conducted by Mutize and Nkhalamba (2020), it was found that long-term bond investors in South Africa tend to consider debt rating downgrades. These downgrades are primarily attributed to the structural issues within the country's economy. According to Mahomed Karodia and Soni (2014), "continuous downgrading by international rating agencies of South Africa's sovereign debt rating coupled with downward economic growth adjustments for several months became a persistent pattern". A market like South Africa, which has experienced rating downgrades and financial instability in the past, is particularly vulnerable to fluctuations in loaning. This volatility has hindered the country's economic growth (Mahomed Karodia & Soni, 2014). In a recent study, it was found that in 2014, South Africa experienced a significant downgrade in its sovereign debt ratings by all three rating agencies. This had a detrimental impact on the country's economy and reputation. Rating downgrades were observed for corporate firms and banks because of rating ceilings (Darwis et al., 2023). To prevent downgrades, it is crucial for emerging and developing countries to thoroughly evaluate the factors that influence ratings and make accurate predictions about their future sovereign debt ratings. African countries heavily rely on international borrowings, making sovereign debt ratings crucial for their access to the international bond market. The impact of downgrades or negative outlooks can have significant consequences on African countries and other emerging markets.

A sovereign debt rating downgrade to junk status indicates that the bonds of that sovereign are considered non-investment grade or speculative. If there is a downgrade in the foreign currency debt aspect, it will result in an increase in the global cost of borrowing money. A potential debt rating downgrade could result in ongoing inflationary pressure and currency depreciation. Consequently, the reserve bank may be compelled to increase the repurchase rate, leading to higher costs for home loan repayments (Meyer & Mothibi, 2021; Mutize & Nkhalamba, 2021). There is a direct correlation between higher interest rates and property values. When interest rates are high, the demand for property decreases, which in turn hinders the growth of property values. Downgrading a sovereign's debt rating may result in a significant outflow of capital from government bonds. When a country's foreign debt is downgraded to the 'junk status,' the borrowing cost in the international lending

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

markets tends to increase (De Villiers et al., 2020; Meyer & Mothibi, 2021; Mutize & Nkhalamba, 2021; Slabbert et al., 2019; Weyers & Eliott, 2017). The downgrades in sovereign asset ratings have a detrimental impact on the performance of the stock market, as highlighted in recent studies (Saadaoui et al., 2022). Afonso et al. (2011), Cantor and Packer (1994), Ozturk et al. (2016), and Ferri et al. (1999) propose that sovereign debt ratings are influenced by macroeconomic variables. On the other hand, Kume (2012) presents a contrasting view, suggesting that changes in debt ratings are not solely driven by macroeconomic factors.

Metz and Tudela (2015) established a connection between financial stability and sovereign debt ratings. They suggested that DRAs can amplify the spread of negative effects and contribute to instability in the financial sector. Rating agencies often take their time to respond to new information and changes in the market (Metz & Tudela, 2015). Caruana and Avdjiev (2012) argue that a strong partnership between the banking sector and the government is crucial for maintaining financial stability. “Sovereign creditworthiness represents the ultimate source of insurance for the financial sector and gives a solid foundation for the pricing of assets, by supplying a risk-free asset; therefore, a sound banking system ensures the ease flow of credit to the economy as well as solid income and funding for the government” as noted by (Caruana & Avdjiev 2012). In a study conducted by Kiff et al. (2012), it was discovered that changes in DRA views have a significant impact on the cost of funding for borrowers. As a result, movements in ratings are closely monitored due to their potential effects on the stability of the financial sector. Research has shown that financial stability can be negatively impacted by downgrades in sovereign debt ratings (Caruana & Avdjiev, 2012; Kiff et al., 2012). The study aims to offer recommendations to African states that depend on sovereign debt ratings to access the international bond market or are impacted by increasing borrowing costs caused by rating downgrades.

This study seeks to analyse and compare the effectiveness of various machine learning techniques (naïve bayes, decision trees, KNN, random forest, and SVM) and logistic regression in predicting sovereign debt ratings. The focus is on utilising macroeconomic factors for this analysis. This study seeks to address the question of which techniques yield more accurate forecasts for sovereign debt ratings using macroeconomic variables: machine learning techniques or traditional statistical models? The primary objective of model comparison is to help new researchers select suitable models for conducting research analysis. This study stands out for its use of fourteen (14) new machine learning models to classify, assess, and predict sovereign debt ratings. In contrast to traditional statistical models like logit, probit, multinomial, and panel regression (Matthews & Mokoena, 2020; Valencia, 2020), this research takes a fresh approach. ML techniques were selected based on research conducted by

(Meharie & Shaik 2020; Arora & Kaur 2020; Phoenix et al., 2021); and (Ly et al., 2021). These studies have shown that these models outperform artificial neural networks (ANN) in the analysis, classification, and forecasting of large data sets. Therefore, the hypothesis being examined is whether ML techniques provide more accurate forecasts compared to traditional statistical models.

LITERATURE REVIEW

This literature review provides a concise overview of the theoretical framework and prior research that has explored the factors influencing SDR. It is interesting to observe that most previous studies have examined the factors influencing SDR by converting it into an arithmetic scale. These studies have focused on macroeconomic indicators as explanatory factors, as demonstrated by (Butler & Fauver, 2006; Archer et al., 2007; Ferri et al., 1999; Cantor & Packer, 1994; Mora, 2006), and (Ratha et al., 2011). In this section, we will explore the literature pertaining to the subject matter. We will begin by examining the theoretical literature, followed by an analysis of the empirical literature on the same topic.

Theoretical Literature on Determinants of Sovereign Debt Rating

The theory behind sovereign debt rating in monetary policy involves the use of the open market operation system for managing sovereign debt. Government securities or bonds are sold internationally to borrow funds by central banks from foreign creditors and investors. These transactions have been extensively studied by (Takawira, 2022) and (Chee et al., 2015). However, the theory of fiscal policy revolves around effectively managing government debt to prevent sudden economic disruptions, reduce risks, and avoid unfavourable tax changes for the government (Heryán & Tzeremes, 2017). A substantial tax base signifies the government's significant capacity to repay loans. As stated by Takawira (2022), the debt overhang theory examines how the current external debt of a borrowing government impacts its ability to repay its debts. Governments with significant debt face a heightened risk of default and may find it difficult to obtain new credit (Demmou et al., 2021). Based on Mellios and Paget-Blanc (2006), nations may struggle to repay debt due to insolvency or a lack of liquidity.

Empirical Literature Review

In a study conducted by Afonso et al. (2011), the factors influencing sovereign debt ratings were examined using rating codes from the top three global DRAs between 1995 and 2005. The study utilised the linear regression method and a random effect technique with ordered probit response effects. They utilised a distinct methodology to categorise the impacts of various fiscal and macroeconomic factors on a country's sovereign debt rating, distinguishing between long-term and short-term effects. Afonso et al. (2011) found that certain factors had a short-term impact on a country's credit rating, including variations in per capita GDP, economic growth, government liability, and government reserves. On the other hand, long-run sovereign ratings were influenced by factors such as effective governance by the government, foreign debt, foreign reserves, and historical default record. In Liu (2002) study, logistic regression was utilised to develop a framework for the data mining application process in credit rating. In a comprehensive study conducted by Bolton et al. (2012), logistic regression and its practical application in credit rating were thoroughly examined. The study delved into the intricate statistical aspects that underpin credit rating. Liu (2002) and Bolton et al. (2012) have found that the logistic model is more effective than newer models. In her study, Novotná (2012) employed various statistical methods to forecast debt ratings based on financial data from European companies. Novotná (2012) expressed a preference for the Discriminant analysis and the logistic regression model, highlighting their ease of use and application, as well as their strong ability to classify sovereign debt ratings.

According to Cantor and Packer (1994), Ferri et al. (1999), Afonso et al. (2011), and Ozturk et al. (2016), macroeconomic factors affect sovereign debt ratings. Pacelli and Azzollini (2011) predicted Indonesian Islamic bank defaults using macroeconomic indices as GDP growth, inflation, stock exchange prices, exchange rates, and money supply. It was found that stock exchange prices reveal future economic or financial concerns. Pacelli and Azzollini (2011) assessed and managed default risk using an ANN. The researchers observed that standard arithmetic models and neural networks have different assumptions, strengths, and limitations that must be evaluated. In a study conducted by Iyengar (2010), the author compared the debt rating of organisations to the assessment of their ability to fulfil their debt obligations in a timely and efficient manner. This assessment applies to various entities such as banks, financial or nonfinancial institutions, and business organisations. In a scholarly analysis, Kume (2012) pointed out that the information provided by ratings agencies may have limited value. This is because there is often a strong correlation between ratings and actual defaults. In many cases, ratings agencies only downgrade ratings after new information have already been made public or reported.

In their study, Gogas et al. (2014) employed the ordered probit model to predict the debt rating of banks. They utilised publicly available quantitative information and data from corporate financial statements. The results showed that the debt ratings of banking corporations are heavily influenced by historical data, causing them to respond slowly to financial issues that have already become public knowledge. In their study, Gurný and Gurný (2013) conducted a comparative analysis of debt rating models. They specifically examined the estimation of default probability for US financial institutions using various regression models, such as linear discriminant, multinomial, logit, and probit regression. The researchers also assessed the statistical significance of the estimated parameters. It was discovered that the control sample yielded the best results when using the logit model. This finding emphasises the suitability of the logit model for predicting banks' default risk. Furthermore, it highlights significant differences between the probit and logit models. However, the LDA model appears to be inadequate for predicting the bank's loan repayment failure (Gurný & Gurný, 2013). In their study, Ramayah et al. (2010) presented an example of the Discriminant analysis method, which is typically employed when the data follows a normal distribution. On the other hand, logistic regression is utilised when the data does not exhibit a normal distribution. Tabachnick et al. (2013) state that discriminant analysis can address similar inquiries as regression analysis.

León-Soriano and Muñoz-Torres (2012) utilised artificial neural networks to model sovereign debt ratings in the European Union. It has been observed that there is significant criticism surrounding DRAs, primarily regarding the lack of transparency in their rating processes and the potential impact of their published ratings, particularly in the realm of sovereign debt. In a study conducted by Gallo (2005), financial modelling was performed using artificial neural networks. The findings revealed that the purchasing behaviour of buyers can be accurately described using a model that incorporates behavioural inferencing, such as Artificial Neural Networks (ANN). These models were found to be superior to traditional statistical or econometric models. Niaki and Hoseinzade (2013) made predictions for the S&P 500 index using ANN and created experimental models. The results of using the ANN demonstrated that this methodology utilises highly effective features to accurately predict the daily movements of the S&P 500, surpassing the capabilities of the traditional logit analytical model. Artificial Neural Networks (ANN) are crucial in financial modelling for analytical processes such as pattern classification, recognition, causation, forecasting, and time series projection (Nazari & Alidadi, 2013).

A study conducted by Almeida et al. (2017) examined the real-world consequences of downgrading sovereign debt ratings. The findings revealed that such downgrades result in increased borrowing costs and decreased investment, including a reduction in firms' financial leverage. Valle and Marín (2005) emphasised the importance of a particular variable in influencing ratings as perceived by DRAs. This variable is GDP, which provides an estimate of the government's revenue generated through tax collection. Having a strong GDP enables the government to meet its debt obligations promptly. A higher GDP, coupled with a growth projection, improves the credit rating of a country. The authors presented conflicting information regarding the various factors and their impact on sovereign debt ratings. Further investigations were conducted to explore the intricacies of the evaluation methods employed by DRAs in analysing changes in sovereign debt ratings. The study utilises the primary economic indicators commonly used in academic research. However, the existing literature reveals a lack of agreement on the models or variables to employ when analysing SDR. Additionally, there is ongoing debate regarding the performance of traditional statistical models versus more recent machine learning algorithms.

RESEARCH METHODOLOGY

The research centred on sovereign debt ratings and other macroeconomic indicators. The original objective was to find a model capable of capturing and predicting sovereign debt ratings by utilising macroeconomic indicators. Information was gathered for South Africa from a range of sources including Trading Economics website, Statistics South Africa (Stats SA), Thomson Reuters, Quantec Easy data, and the South African Reserve Bank (SARB), in quarterly form, spanning from 1999 to 2022. The approach utilised a multidisciplinary environment, employing data mining and econometric models within machine learning to analyse time series data and identify solutions to financial challenges.

The sovereign debt ratings used were the sovereign debt ratings released by the three major DRAs that is S&P, Fitch, and Moody's rating companies. The sovereign debt ratings in categorical form were modelled and forecasted using the Gradient Boosting Classifier (gbc), K Nearest Neighbors Classifier (knn), Decision Tree Classifier (dt), Random Forest Classifier (rf), Ada Boost Classifier (ada), Extra Trees Classifier (et), Extreme Boosting Gradient (xgboost), Support Vector Machine (svm) - Linear Kernel, Light Gradient Boosting Machine (lightgbm), Logistic Regression (lr), Naive Bayes (nb), Ridge Classifier (ridge), Linear Discriminant Analysis (lda), Dummy Classifier (dummy), and Quadratic Discriminant Analysis (qda). Sovereign debt ratings were grouped into two classes namely upgrade and downgrade.

Correlation tests were conducted to analyse the data and eliminate variables that exhibited multicollinearity. Prior to the analysis, the independent variables underwent testing for stationarity and multicollinearity using the Variance Inflation Indicator (VIF). As per the research conducted by Hair et al. (2006), multicollinearity refers to the extent to which one factor can be accounted for by another factor in the analysis or equation. The data of the selected variables used in the models were examined for multicollinearity using the variance inflation factor (VIF). It is worth noting that calculated VIF values greater than 10 indicate a high level of multicollinearity (Greene, 2012). The VIF values calculated for all the variables analysed were below 10, suggesting that the selected variables did not exhibit significant issues with multicollinearity. The VIF results for all variables can be found in the appendices section. After careful consideration, a final list of

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

independent variables has been selected for classifying and analysing sovereign debt ratings. These variables include REER, PIR, HDDIR, UR, GDPpc, BOP, CAB, FDGDP, and CPIH.

Unit root tests were conducted, and the data was transformed into first differences to achieve stationarity. The data variables were divided into two sections using cross-validation, a Machine Learning tool. The datasets were divided into a test set and a train set, with a ratio of 75:25. The models were trained using the provided data to identify trends and patterns. After capturing these trends and patterns, the models were used to predict the values of the dependent variables in the test set by inputting the independent variables from the test set data. The study conducted a comparison of the performance of ML models and LR in predicting ratings from various Credit Rating Agencies.
The model used is shown below:

$$y_{jt} = \beta_0 + \sum_{i=1}^{9} x_{it}\, \beta_i + \varepsilon_t \quad (1)$$

where coefficients β_0 and β_j are parameters that areunknown whilst the ε is the stochastic error term alleged to be identically distributed and independent with a mean of zero (0) and a variance that is constant. y_{jt} is the responsive variable at time ‘t’, from rating agent ‘j’; ‘j’ are SDR notes (category) that can either be Fitch Rating (for example, “AAA”, “BB”, “D”, etc.), Moody’s rating (for example, “Aaa”, “Baa”, “Ca”, etc.) or Standard & Poor’s ratings (for example, “AAA”, “CCC”, “B”, etc.) being either downgrades or upgrades; x_{jt} are the macroeconomic indicator ‘i’ at time ‘t’, and ε_t *is the stochastic error term*. The list of macroeconomic variables include[2] REER, PIR, HDDIR, UR, GDPpc, CPIH, FDGDP, BOP and CAB. Hence equation [1] above becomes:

$$SCR_{jt} = \beta_0 + \mathrm{REER}_t\beta_1 + \mathrm{PIR}_t\beta_2 + \mathrm{HDDIR}_t\beta_3 + \mathrm{UR}_t\beta_4 + \mathrm{GDPpc}_t\beta_5 + \mathrm{CPIH}_t\beta_6 + \mathrm{FDGDP}_t\beta_7 + \mathrm{BOP}_t\beta_8 + \mathrm{CAB}_t\beta_9 + \varepsilon_t \quad (2)$$

Models

Logistic Regression

According to Romadhon and Kurniawan (2021) the logistic model is a type of regression that connects one or more independent factors with the responsive factor of a classification type; can either be 0 or 1, high or low, true, or false, upgrade or downgrade, big or small etc. The encryption was “1” for upgrade and “0” for downgrade. The equation is as follows:

2 1-real effective exchange rates (REER), 2-prime interest rates (PIR), 3-household debt to disposable income (HDDIR), 4-unemployment rate (UR), 5-gross domestic product percentage change (GDPpc), 6-consumer price index headline (CPIH), 7-foreign debt to gross domestic product (FDGDP), 8-balance of payments (BOP) and 9-current account balance (CAB)

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

$$Ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{j=1}^{n} \beta_j X_j \quad (3)$$

Where:
β_0 = constant
β_j= the coefficient of every data factor
The p value or probability (Y=1) can be deduced from the equation:

$$p = \frac{e^{\left(\beta_0 + \sum_{j=1}^{n} \beta_j X_j\right)}}{1+e^{\left(\beta_0 + \sum_{j=1}^{n} \beta_j X_j\right)}} \quad (4)$$

Naïve Bayes

The purpose of a Naive Bayes technique as adopted from the Bayes rule is to gauge the conditional probability of an incident with a feature vector $B_1, B_2, \cdots, B_n$ belonging to a particular cluster A, the Bayes Theorem equation is shown below:

$$P(A|B) = \frac{P(X|A).P(A)}{P(B)} \quad (5)$$

with
B data statistic with unknown cluster
A The hypothesis B in each cluster.
$P(A|B)$ The possibility of the A hypothesis in reference to B
$P(A)$ Probability of the hypothesis A (preceding probability)
$P(B|A)$ Probability B in hypothesis of A
$P(B)$ Probability B

Decision Trees

In a bivariate classification problem, a decision tree poses a question that requires a Yes/No response and subsequently utilises the answer to divide the trees into subtrees. In a study conducted by Ramesh et al. (2022), decision trees were found to be a key component in a process that involves the growth of a tree-like model. This model considers various assertions and their potential outcomes, as well as factors such as occurrence inferences, cost features, and performance characteristics. The Decision Tree Classifier divides the divisions into sub-sets, starting from the root and branching out to left and right outputs. When calculating information gain, the process relies on "Entropy," which is a measure of the disorder, impurity, or randomness of the dataset (Bel et al., 2005; Charbuty & Abdulazeez, 2021; Hastie et al., 2009; Mathuria et al., 2013; Steinberg & Cardell, 1998; Zacharis, 2018). The anticipated quantity of information when evaluating the output of an X random variable:

$$H(X) = E\big(I(X)\big) = \sum_i p(x_i)I(x_i) = -\sum_i p(x_i)log_2 p(x_i) \quad (6)$$

Entropy Conditional information has a formula as shown below:

$$H(X|Y) = -\sum_j p(y_j)H\big(X|Y = y_j\big) = -\sum_j p(y_j) \sum_i p(x_i \mid y_j)log_2 p(x_i \big| y_j\big) \quad (7)$$

Information Gain (IG) and Gain Ratio are widely used feature selection methods in the field of decision trees. Information gain is a key factor in selecting the characteristic that will serve as the initial node. The focus is on the decrease in the entropy value. Gain ratio is useful for analysing various characteristics in the dataset, while the IG method measures the changes in entropy values when splitting the dataset. The calculation for information gain is demonstrated in equation 8 below (*note*- Gini index is shown in appendix B):

$$\text{IG}\ (X, Y) = H(X) - H(X|Y) \quad (8)$$

Random Forest

A Random Forest is a combination of multiple decision trees. Random forest algorithms utilise multiple individual decision trees to create a forest consisting of numerous trees. The trees are combined and undergo a training process that involves random attribute selection. This process utilises bootstrapping, an internal method that prevents interaction between decision trees.

Random Forest algorithm development according to Hastie et al. (2009), requires taking $b = 1\ to\ B$ and creates a bootstrap sample Z^* of size N from the data used for the model training, then draw a random-forest combined trees T_b to the data bootstrapped by reiteratively repeating the process in stages for every terminal node of each tree until the process reaches the minimum node size n_{min}; Continuing the process, a certain number of variables are randomly chosen from a larger set. The best split-point among these selected variables is used to divide the node into two separate nodes. This process is repeated multiple times, resulting in a collection of trees $\{T_b\}_1^B$.

To predict the new point x using regression:

$$\hat{f}_{rf}^B\ (x) = \frac{1}{B} \sum_{b=1}^B T_b\ (x) \quad (9)$$

This study applies the classification method and so in using classification, thus let $\hat{C}_b(x)$ be the cluster prediction of the b^{th} random-forest tree.

Usually values for m are $\sqrt{p}$ or even as lower than 1. After B, such trees $\{T(x;\ \Theta_b)\}_1^B$ are developed, the random forest technique as a regression predictor is therefore illustrated below:

$$\hat{f}_{rf}^B\ (x) = \frac{1}{B} \sum_{b=1}^B T(x;\ \Theta_b) \quad (10)$$

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

Where: Θ_b illustrates the bth random forest tree in terms of split data, cut points of every node, and values of the terminal node.

Support Vector Machine

Support Vector Machine (SVM) is a powerful tool in the field of machine learning. It is used to classify data and find the best separating hyperplane for data points. SVM is a numerical classifier that creates a single decision boundary to optimise the margin between binary points, such as positive and negative. The vectors are then utilised to calculate the maximal margin, as depicted in the diagram above. The optimal separating hyperplane is positioned at the centre of the maximal margin to ensure a fair separation of the data points. When provided with a set of observations (x_i, y_i), SVM regression generates a linear model. The covariates, represented by xi, and the actual number outcomes, denoted by yi, are used in this process. The SVM model used the approach applied by (Okasha, 2014) as shown below:

$$y(x) = w^T \varphi(x) + b \quad (11)$$

Where w denotes weights of vector covariates and b resembles the intercept. The $\varphi(x)$ denotes the higher dimension feature space and so the coefficients $w\ and\ b$ are estimated by minimising the following function:

$$Min\left(\frac{1}{2}\,w^T w\right) \quad (12)$$

Subject to the following constraints:

$$y_t - w^T\varphi(x_i) - b \leq \varepsilon \qquad \text{or } w^T\varphi(x_i) + b - y_t \leq \varepsilon \quad (13)$$

K-Nearest Neighbour (kNN)

K-Nearest Neighbour (kNN) is a technique used in machine learning that involves a supervised learning process. Supervised learning is centred around discovering fresh features or patterns in the dataset through the examination of existing data patterns in relation to new ones. The kNN model classifies by evaluating the distance between data points. The kNN principle is that, in samples of feature space, similar samples should be selected most of the time. In the kNN model, Xu et al. (2020) found that all selected samples were classified accurately. However, the determination of the required category only considered a few data samples that were close to their position. kNN utilises either the Euclidean or Manhattan distance to classify an object, based on k training datasets located near the object being analysed. There are certain conditions that need to be met for the properties to be set. One condition is that the value of k cannot exceed the amount of training data. Additionally, the value of k must be an odd number greater than one (1). The kNN algorithm is commonly employed to categorise new objects based on data characteristics, proximity of datapoints, and training samples. The results of the new test samples are derived from the classification of datasets, primarily based on the categories established by the kNN algorithm used for classifying the training samples.

In their study, Lubis and Lubis (2020) mention that distance can be standardised by dividing each attribute distance by the range obtained through subtracting the minimum value from the maximum value on the attributes, as shown in equation 14. This is done to ensure that the values for each point are within a standardised range of 0 to 1. Here is a formula for normalising data, where the data is scaled to a range between 0 and 1.

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (14)$$

Where x is value of the data, y is value of normalization, x_{min} represent minimum value i.e. 0 and x_{max} represent maximum value thus 1.

The proximity value is calculated from the data after normalisation. The Euclidean distance formula is utilised in this calculation process to determine predictions. The equation for calculating proximity between two case points is shown as:

$$similarity\ (T, S) = \frac{\sum_{i=1}^{n} f(\,T_i, S_i\,)x}{w_i} \qquad (15)$$

where: t is new case; s is the value of the closeness of the case in storage; n is number of features in every case; i is individual qualities between 1 to n ; f is similarity feature function of i between case T and case S and w is the weight given to the feature i.

The other ML models applied in this study include the Gradient Boosting Classifier (gbc), Ada Boost Classifier (ada), Extra Trees Classifier (et), Extreme Boosting Gradient (xgboost), Light Gradient Boosting Machine (lightgbm), Ridge Classifier (ridge), Linear Discriminant Analysis (lda), Dummy Classifier (dummy), and Quadratic Discriminant Analysis (qda).

Comparison of Model Performance on Prediction

To evaluate the efficiency of forecasting on the above-mentioned modelling techniques, this study compares them to against one another as adopted from (Lalwani et al., 2022; Pranckevičius & Marcinkevičius, 2017; Ramesh et al., 2022; Romadhon & Kurniawan, 2021; Van der Heide et al., 2019). In this study, various performance metrics were utilised to evaluate the prediction capacity of each model. Testing parameters such as false positive (FP), false negative (FN), true positive (TP), and true negative (TN) were considered. The parameters are utilised to calculate various metrics such as model accuracy, precision, specificity, sensitivity, F1 score, and area under the curve. These metrics effectively demonstrate the models' prediction capacity and performance.

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-61) **Doi:** 10.34109/ijefs.202416102

Accuracy

The fraction of incidences in the whole dataset that has been correctly forecasted out of all possible occurrences, as shown by the equation below:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (16)$$

Sensitivity / Recall

The portion of exactly projected elements in the dataset that are positive and is calculated as illustrated below:

$$\text{Recall or Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (17)$$

Precision

It assesses the accuracy of positive outcome predictions in forecasting. The precision determines the level of accuracy of a dataset for a specific minority class. The calculation involves subtracting the number of positive predictions in productive instances from the total number of expected sample instances and the formula is as shown below:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (18)$$

Specificity

Specificity and sensitivity are inversely related thus as specificity increases, sensitivity tends to decrease, and vice versa.

$$S\text{pecificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (19)$$

F1-Score

To calculate the F1-score the precision and recall values are merged in a complex harmonic mean of reciprocals. This technique assesses the model's accuracy, and the formula is as shown below:

$$F1 - \text{Score} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision+Recall}}\right) \quad (20)$$

AUC (Area under the curve)

In the realm of academia, a probability function is employed to assess the likelihood of selecting a positive case over a randomly chosen negative case. This is achieved through the implementation of a ranking technique within the area under the curve (AUC) framework. The shaded region beneath the curve represents the receiver needed to operate characteristic curves, which are plotted to visualise components utilised for assessing the models' efficacy. The AUC provides values that represent the level of prediction. A value close to 1 indicates perfect prediction, while a value of 0.5 suggests no difference from random chance, represented by a transverse line. Therefore, the best fit is achieved when the curve is significantly above the transverse line.

Matthews Correlation Coefficient (MCC)

MCC assesses the performance of a classification model and ranges between negative one (−1) and positive one (+1).

$$\mathrm{MCC} = \frac{\mathrm{TP*TN} - \mathrm{FP*FN}}{\sqrt{(\mathrm{TP+FP})(\mathrm{TP+FN})(\mathrm{TN+FP})(\mathrm{TN+FN})}} \quad (21)$$

Kappa Statistic

A metric that compares expected versus observed accuracy thus P_o observed probability and P_e expected probability.

$$\mathrm{Kappa} = \frac{\mathrm{P_o} + \mathrm{P_e}}{1-\mathrm{P_e}} \quad (22)$$

Where $\mathrm{P_o} = (\mathrm{TP} + \mathrm{TN}) / (\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})$ and

$$\mathrm{P_e} = ((\mathrm{TP} + \mathrm{FN}) * (\mathrm{TP} + \mathrm{FP}) + (\mathrm{FP} + \mathrm{TN}) * (\mathrm{FN} + \mathrm{TN})) / (\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2$$

Findings and Empirical Results

Machine Learning techniques excel at efficiently and effectively modelling, identifying patterns or trends, and forecasting the movement of debt ratings. Nevertheless, DRAs employ various methodologies and distinct economic variables when assessing the creditworthiness of sovereigns. The data of the selected explanatory variables used in the models were examined for stationarity and multicollinearity using appropriate tests. The VIF values obtained from the analysed variables indicate that the selected variables do not exhibit significant issues of multicollinearity, as all values were below 10. To check the model specification a Ramsey regression equation specification error test (RESET) was conducted. According to Griffith, and Chun, (2016:1) “the Ramsey RESET test furnishes a diagnostic for omitted variables in a linear regression model specification (i.e., the null hypothesis is no omitted variables).” The results of the Ramsey RESET test, as presented in Appendix C, demonstrate that the three models for Fitch ratings, Moody's ratings, and Standard & Poor's Ratings, when incorporating macroeconomic variables, were found to be free from any specification errors. This conclusion is supported by the fact that the p-values, t-statistic, F-statistic, and Likelihood ratio for all three models were greater than 0.05.

In a similar vein, the study revealed that ML techniques effectively captured the fluctuations in sovereign debt ratings by utilising economic indicators. As a result, these techniques can be employed to classify and analyse debt ratings. Machine learning algorithms outperformed the logistic regression model in accurately predicting values, showcasing their superior precision. Machine Learning algorithms have been found to have superior predictive and classification capabilities compared to traditional econometric models. However, the effectiveness of these algorithms is contingent upon the specific data and analysis being conducted. Machine Learning approaches such as the Gradient Boosting Classifier (gbc), K Nearest Neighbors Classifier (knn), Decision Tree Classifier (dt), Random Forest Classifier (rf), Ada Boost Classifier (ada), Extra Trees Classifier (et), Extreme Boosting Gradient (xgboost), Support Vector Machine (svm)– Linear Kernel, Light Gradient Boosting Machine (lightgbm), Logistic Regression (lr), Naive Bayes (nb), Ridge Classifier (ridge), Linear Discriminant Analysis (lda), Dummy Classifier (dummy), and Quadratic Discriminant Analysis (qda) can categorise, analyse, model, recognise patterns or trends, and forecast credit scores efficiently and effectively.

The models successfully categorised the given variables into two scenarios: upgrade and downgrade. Variables categorised as stable indicate that when they improve, the ratings become more favourable, while less stable variables are used when there are unfavourable rating movements. The model emphasised the factors that impact ratings during stable situations and those that contribute to negative rating movements or downgrades. The findings suggest that to prevent rating downgrades, it is important for sovereigns to manage household debt, keep inflation in check, address exchange rate risks, and consistently promote GDP growth. The results validate the conclusions reached by various researchers in the field, including Kumar and Haynes (2003), Bennell et al. (2006), Kräussl (2005), and Cantor and Packer (1994). These studies have consistently shown that sovereign debt ratings serve as a comprehensive and valuable tool in assessing credit risk, complementing the information provided by macroeconomic indicators. As a result, there is a strong correlation between these ratings and market-determined credit spreads.

Comparison of Prediction and Classification Accuracy

This study utilised Machine Learning (ML) algorithms and logistic regression to categorise, analyse, and predict SDR from the major DRAs, including Fitch, Moodys, and S&P, by utilising macroeconomic indicators. The classification under machine learning effectively captured the relationship between SDR and economic indicators by accurately identifying the key factors influencing SDR. The sovereign debt ratings were sourced from the major rating agencies, including Fitch ratings, Moody's ratings, and Standard & Poor's rating agent, on the Trading Economics website. The sovereign debt ratings were categorised as the response variable, while numerical macroeconomic indicators were used as explanatory variables. The performance results of the models utilised are displayed in Tables 1, 2, and 3.

The metrics for prediction performance on the Moodys ratings using macroeconomic variables are illustrated in Table 1 above. The results indicate the performance of each model, with the top-performing model listed first and the least effective model listed last. In terms of accuracy, the gbc model achieved the highest ranking of 0.9857, indicating its superiority over other models in predicting Moody's future ratings. After that, knn, dt, rf, ada, et, and xgboost achieved a score of 0.9714, while svm scored 0.9571 and lightgbm scored 0.9548. The model of interest, lr, ranked number 10 with a score of 0.9262, which is equal to nb, ridge, and lda. It is evident that 9 ML models outperformed lr, while 3 models achieved similar results. However, it is worth noting that 2 models (dummy and qda) exhibited poorer performance compared to lr when predicting Moody's future sovereign debt ratings.

Among the models tested, including rf, et, xgboost, lda, and lr, all achieved a perfect score of 1 on the AUC metric, indicating excellent relationship and prediction capabilities. Several models achieved scores above 0.9, including Gbc, knn, dt, ada, lightgbm, and nb. Out of the 12 models tested, including lr, one model stood out with a score of 1, indicating exceptional performance in predicting Moody's ratings. In terms of precision metrics, gbc achieved the highest ranking with a score of 0.9750. It was closely followed by eight other models (knn, dt, rf, ada, et, xgboost, svm, and lightgbm) which all achieved a score of 0.95. The LR model ranked 10th with a score of 0.8750, which was on par with NB, Ridge, and LDA. However, the weakest models for forecasting Moody's ratings were Dummy and QDA.

The F1 score as per Moodys indicated that the gbc model performed the best with a score of 0.9857. It was followed by knn, dt, rf, ada, and et with scores of 0.9714, and svm and lightgbm with scores of 0.9514. The lr ranked 10th with a score of 0.9286, like nb, ridge, and lda. The Cohen Kappa statistics revealed that the top-performing model for predicting Moody's rating was the gbc, achieving an impressive score of 0.9720. Following closely behind were the knn, dt, rf, ada, et, and xgboost models, all achieving a score of 0.9440. Following were the svm and lightgbm models, both achieving scores of 0.9136 and 0.9107 respectively. The lr model, with a score of 0.8547, ranked 10th, on par with nb, ridge, and lda.

Table 1: Model Prediction Performance Results for Moody.

Models	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
Gradient Boosting Classifier (gbc)	0.9857	0.9875	1.0000	0.9750	0.9857	0.9720	0.9750
K Neighbours Classifier (knn)	0.9714	0.9875	1.0000	0.9500	0.9714	0.9440	0.9500
Decision Tree Classifier (dt)	0.9714	0.9750	1.0000	0.9500	0.9714	0.9440	0.9500
Random Forest Classifier (rf)	0.9714	1.0000	1.0000	0.9500	0.9714	0.9440	0.9500
Ada Boost Classifier (ada)	0.9714	0.9792	1.0000	0.9500	0.9714	0.9440	0.9500
Extra Trees Classifier (et)	0.9714	1.0000	1.0000	0.9500	0.9714	0.9440	0.9500
Extreme Boosting Gradient (xgboost)	0.9714	1.0000	1.0000	0.9500	0.9714	0.9440	0.9500
SVM – Linear Kernel (svm)	0.9571	0.0000	0.9667	0.9500	0.9514	0.9136	0.9230
Light Gradient Boosting Machine (lightgbm)	0.9548	0.9917	0.9667	0.9500	0.9514	0.9107	0.9207
Logistic Regression (lr)	0.9262	1.0000	1.0000	0.8750	0.9286	0.8547	0.8707
Naive Bayes (nb)	0.9262	0.9333	1.0000	0.8750	0.9286	0.8547	0.8707
Ridge Classifier (ridge)	0.9262	0.0000	1.0000	0.8750	0.9286	0.8547	0.8707
Linear Discriminant Analysis (lda)	0.9262	1.0000	1.0000	0.8750	0.9286	0.8547	0.8707
Dummy Classifier (dummy)	0.5500	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
Quadratic Discriminant Analysis (qda)	0.4500	0.0000	1.000	0.4500	0.6200	0.0000	0.0000

Source: By Author

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

The MCC exhibits a similar pattern to the Kappa metric in Moodys rating, with the gbc model emerging as the top performer in predicting ratings, while the lr model ranks tenth. The least effective models in predicting Moody's sovereign debt ratings were dummy and qda, as they exhibited poor performance metrics.

Table 2 above presents the performance metrics for forecasting future sovereign debt ratings of Standard & Poor's. The top-performing models across all prediction metrics, including accuracy, AUC, recall/sensitivity, precision, F1 score, and MCC, were lr, nb, dt, svm, ridge, rf, ada, et, xgboost, and lightgbm. These models consistently achieved the highest scores of 1. Next, knn achieved an impressive score of 1 on AUC and recall, along with scores above 0.9400 on all other performance metrics. The lda exhibited strong performance across all performance metrics in predicting S&P ratings, with scores ranging from 0.8189 to 0.9958. Nevertheless, the dummy and qda models exhibited limited effectiveness in predicting S&P ratings.

The findings of predicting Fitch ratings using machine learning algorithms are highlighted in Table 3, in comparison to logistic regression. Among the models, nb, ridge, rf, gbc, lda, et, and xgboost achieved the highest accuracy score of 0.9690. Afterwards, lr achieved a score of 0.9548, placing it in the 8th position. Out of the 5 ML models (dt, ada, lightgbm, knn, and svm), their scores were 0.9405, 0.9405, 0.9405, 0.9262, and 0.9119 respectively. However, all these models had scores lower than the lr model's score. Models such as rf, et, lr, and lightgbm achieved a score of 1 on AUC, indicating exceptional prediction performance. The models nb, gbc, lda, xgboost, dt, ada and knn achieved AUC scores of 0.9708, 0.9708, 0.9583, 0.9889, 0.9375, 0.9917 and 0.9833 respectively, indicating strong predictive capabilities. However, the AUC scores of models such as ridge, svm, dummy, and qda were not satisfactory. Out of the 9 models tested, including nb, ridge, rf, gbc, lda, et, and xgboost, lr and dt, all achieved an impressive recall or sensitivity score of 0.9750. This indicates their excellent performance in accurately forecasting Fitch ratings. Ada and knn achieved a score of 0.9500, while lightgbm and svm obtained scores of 0.9250 and 0.9000, respectively, on the recall metric.

In terms of precision, 9 models (nb, ridge, rf, gbc, lda, et, xgboost, lightgbm, and lr) achieved an impressive score of 0.9750, indicating excellent prediction performance. Afterwards, ada, svm, dt, and knn achieved scores of 0.9550, 0.9500, 0.9350, and 0.9350, respectively. The dummy and qda performed poorly in terms of precision when predicting Fitch ratings. Among the models tested, nb, ridge, rf, gbc, lda, et, and xgboost outperformed lr on the F1 metric, with a score of 0.9714 compared to lr's score of 0.9524. The other models, such as dt, ada, lightgbm, knn, and svm, performed well, although they were not able to outperform lr. Their scores ranged from 0.9492 down to 0.9095. The kappa and MCC metrics on Fitch Ratings prediction performance showed a similar pattern to the precision metrics, with 7 models outperforming the LR model.

Table 2: Model Prediction Performance Results for Standard & Poors.

Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
Logistic Regression (lr)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Naive Bayes (nb)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Decision Tree Classifier (dt)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SVM – Linear Kernel (svm)	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Ridge Classifier (ridge)	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Random Forest Classifier (rf)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Ada Boost Classifier (ada)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Gradient Boosting Classifier (gbc)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Extra Trees Classifier (et)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Extreme Boosting Gradient (xgboost)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Light Gradient Boosting Machine (lightgbm)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
K Neighbours Classifier (knn)	0.9714	1.0000	1.0000	0.9550	0.9746	0.9416	0.9480
Linear Discriminant Analysis (lda)	0.9095	0.9958	0.8833	0.9550	0.9060	0.8189	0.8395
Dummy Classifier (dummy)	0.5357	0.5000	1.000	0.5357	0.6964	0.0000	0.0000
Quadratic Discriminant Analysis (qda)	0.4643	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: By Author

Table 3: Classification Results for Fitch.

Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
Naive Bayes (nb)	0.9690	0.9708	0.9750	0.9750	0.9714	0.9387	0.9457
Ridge Classifier (ridge)	0.9690	0.0000	0.9750	0.9750	0.9714	0.9387	0.9457
Random Forest Classifier (rf)	0.9690	1.0000	0.9750	0.9750	0.9714	0.9387	0.9457
Gradient Boosting Classifier (gbc)	0.9690	0.9708	0.9750	0.9750	0.9714	0.9387	0.9457
Linear Discriminant Analysis (lda)	0.9690	0.9583	0.9750	0.9750	0.9714	0.9387	0.9457
Extra Trees Classifier (et)	0.9690	1.0000	0.9750	0.9750	0.9714	0.9387	0.9457
Extreme Boosting Gradient (xgboost)	0.9690	0.9889	0.9750	0.9750	0.9714	0.9387	0.9457
Logistic Regression (lr)	0.9548	1.0000	0.9500	0.9750	0.9524	0.9128	0.9255
Decision Tree Classifier (dt)	0.9405	0.9375	0.9750	0.9350	0.9492	0.8778	0.8918
Ada Boost Classifier (ada)	0.9405	0.9917	0.9500	0.9550	0.9460	0.8802	0.8937
Light Gradient Boosting Machine (lightgbm)	0.9405	1.0000	0.9250	0.9750	0.9429	0.8827	0.8957
K Neighbours Classifier (knn)	0.9262	0.9833	0.9500	0.9350	0.9302	0.8520	0.8715
SVM – Linear Kernel (svm)	0.9119	0.0000	0.9000	0.9500	0.9095	0.8288	0.8505
Dummy Classifier (dummy)	0.5357	0.5000	1.0000	0.5357	0.6964	0.0000	0.0000
Quadratic Discriminant Analysis (qda)	0.4643	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: By Author

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

The models that showed subpar performance across various metrics of prediction under the Fitch sovereign debt ratings were dummy and qda.

In general, the performance of Machine Learning models, including logistic regression, was strong in accurately predicting and classifying sovereign debt ratings. However, the dummy classifier and quadratic Discriminant analysis did not perform as well in this regard. DRAs utilise distinct variables, data, and various models to assess a nation's sovereign credit rating. This leads to similarities in performance between machine learning algorithms and traditional econometric or statistical models, such as Logistic Regression. The findings align with previous studies that have also explored the performance of traditional statistical models compared to Machine Learning models. These studies include Takawira (2022), Novotná (2012), Pacelli and Azzollini (2011), Tabachnick et al. (2013), Gallo (2005), Gurný and Gurný (2013), and Ramayah et al. (2010).

However, the Covid 19 pandemic may have impacted the accuracy of sovereign debt rating predictions in the first and second quarters of 2020. Subjective factors play a significant role in determining Sovereign debt ratings (SDRs). These factors include political events, social events, global events, governance, international relations, and other unquantifiable aspects. It is these factors that lead to different ratings being assigned to the same institution or country by different rating agencies, as they apply different weights to these variables when determining SDRs.

CONCLUSION

This study utilised Machine Learning techniques, which have been shown to accurately predict variables with great precision. Testing the level of accuracy for each model on prediction was aided by cross-validation. The performance results from both ML techniques and the traditional statistical model of Logistic regression were found to be nearly identical, with minimal variations. Nevertheless, ML techniques outperformed the traditional statistical model, although the effectiveness of these techniques is contingent upon the specific data, variables, and analysis employed, leading to potential variations in performance.

To enhance productivity, efficiency, and effectiveness within the economic system and financial sector, it is crucial to ensure proper supervision, monitoring, regulation, and transparency in the services and involvement of DRAs in credit facilitation. It would be beneficial for emerging economies, such as countries in Africa, to consider borrowing from other countries in Africa and institutions like the African Union (AU), Economic Community of West African States (ECOWAS), Southern African Development Community (SADC), Economic Community of Central African States (ECCAS), East African Community (EAC), Arab Mahgreb Union (AMU) and so on. It is crucial to provide adequate funding, strengthen, and integrate these African

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

regional organisations with financial institutions. Additionally, effective management is essential to ensure the accumulation of funds and the provision of credit to African countries. As African countries and regional organisations continue to develop and collaborate, it becomes crucial to establish Sovereign debt rating Agencies in Africa. This will help prevent manipulation in the oligopolistic rating market and ensure proper supervision and regulation of rating agencies.

The study's findings may be limited due to the specific information used from one nation. For example, South Africa's unique conditions or characteristics may differ from those of other developing countries. It is important to note that cross-validation in Machine Learning algorithms does not always guarantee the mitigation of overfitting and misclassification problems. ML models typically require a substantial amount of data to achieve optimal effectiveness, which is often attainable. Further research is necessary to enhance the existing findings. More research should be conducted through the inclusion of variables like governance index, political instability indicator, a measure for corruption, and asset prices like gold. Future studies could be improved by incorporating non-quantitative indicators like political instability, measure of corruption and governance, including effective management of governmental institutions. Further studies must consider the impact of Covid 19 in the analysis to try and see its impact on sovereign debt ratings and financial stability.

REFERENCES

Afik, Z., Feinstein, I., & Galil, K. (2014). The (un) informative value of credit rating announcements in small markets. *Journal of Financial Stability*, *14*, 66-80. doi: https://doi.org/10.1016/j.jfs.2014.08.001

Afonso, A., Gomes, P., & Rother, P. (2011). Short-and long-run determinants of sovereign debt credit ratings. *International Journal of Finance & Economics*, *16*(1), 1-15. doi: https://doi.org/10.1002/ijfe.416

Almeida, H., Cunha, I., Ferreira, M. A., & Restrepo, F. (2017). The real effects of credit ratings: The sovereign ceiling channel. *The Journal of Finance*, *72*(1), 249-290. doi: https://doi.org/10.1111/jofi.12434

Archer, C. C., Biglaiser, G., & DeRouen, K. (2007). Sovereign bonds and the “democratic advantage”: Does regime type affect credit rating agency ratings in the developing world? *International organization*, *61*(2), 341-365. doi: https://doi.org/10.1017/0S0020818307070129

Arora, N., & Kaur, P. D. (2020). A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. *Applied Soft Computing*, *86*, 105936. doi: https://doi.org/10.1016/j.asoc.2019.105936

Baum, C. F., Schäfer, D., & Stephan, A. (2016). Credit rating agency downgrades and the Eurozone sovereign debt crises. *Journal of Financial Stability*, *24*, 117-131. doi: https://doi.org/10.1016/j.jfs.2016.05.001

Bedendo, M., Cathcart, L., & El-Jahel, L. (2018). Reputational shocks and the information content of credit ratings. *Journal of Financial Stability*, *34*, 44-60. doi: https://doi.org/10.1016/j.jfs.2017.12.003

Bel, L., Laurent, J. M., Bar-Hen, A., Allard, D., & Cheddadi, R. (2005). A spatial extension of CART: application to classification of ecological data. In P. Renard, H. Demougeot-Renard, & R. Froidevaux (Eds.), *Geostatistics for Environmental Applications* (pp. 99-109). Springer Berlin Heidelberg. doi: https://doi.org/10.1007/3-540-26535-X_9

Bellotti, T., Matousek, R., & Stewart, C. (2011). A note comparing support vector machines and ordered choice models' predictions of international banks' ratings. *Decision Support Systems*, *51*(3), 682-687. doi: https://doi.org/10.1016/j.dss.2011.03.008

Ben Mim, S., Nouira, R., & Mabrouk, F. (2023). Non-Linear Determinants of Developing Countries' Sovereign Ratings: Evidence from a Panel Threshold Regression (PTR) Model. *Sustainability*, *15*(4), 3390. doi: https://doi.org/10.3390/su15043390

Bennell, J. A., Crabbe, D., Thomas, S., & Ap Gwilym, O. (2006). Modelling sovereign credit ratings: Neural networks versus ordered probit. *Expert systems with applications*, *30*(3), 415-425. doi: https://doi.org/10.1016/j.eswa.2005.10.002

Bolton, P., Freixas, X., & Shapiro, J. (2012). The credit ratings game. *The Journal of Finance*, *67*(1), 85-111. doi: https://doi.org/10.1111/j.1540-6261.2011.01708.x

Butler, A. W., & Fauver, L. (2006). Institutional environment and sovereign credit ratings. *Financial Management*, *35*(3), 53-79. doi: https://doi.org/10.1111/j.1755-053X.2006.tb00147.x

Cantor, R., & Packer, F. (1994). The credit rating industry. *Quarterly Review*, *19*(Sum), 1-26. Retrieved from https://www.newyorkfed.org/medialibrary/media/research/quarterly_review/1994v19/v19n2article1.pdf

Caruana, J., & Avdjiev, S. (2012). Sovereign creditworthiness and financial stability: an international perspective. *Banque de France Financial Stability Review*, *16*(April), 71-85. Retrieved from https://publications.banque-france.fr/sites/default/files/medias/documents/financial-stability-review-16_2012-04.pdf

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(1), 20-28. doi: https://doi.org/10.38094/jastt20165

Chee, S. W., Fah, C. F., & Nassir, A. M. (2015). Macroeconomics determinants of sovereign credit ratings. *International Business Research*, *8*(2), 42. doi: https://doi.org/10.5539/IBR.V8N2P42

Dallara, C. (2008). Structure of regulation: Lessons from the crisis. A view from the Institute of International Finance (IIF). *Journal of Financial Stability*, *4*(4), 338-345. doi: https://doi.org/10.1016/j.jfs.2008.09.011

Darwis, F., Wiryono, S. K., & Prasetyo, A. D. (2023). Modelling Sovereign Ratings in Indonesia: Analysis the Driving Factors Using the Panel Data Regression. *Migration Letters*, *20*(8), 1124-1137. doi: https://doi.org/10.59670/ml.v20i8.5769

De Villiers, C., Cerbone, D., & Van Zijl, W. (2020). The South African government's response to COVID-19. *Journal of Public Budgeting, Accounting & Financial Management*, *32*(5), 797-811. doi: https://doi.org/10.1108/JPBAFM-07-2020-0120

Demmou, L., Franco, G., Calligaris, S., & Dlugosch, D. (2021). *Liquidity shortfalls during the COVID-19 outbreak: Assessment and policy responses* (No. 1647). OECD Publishing. doi: https://doi.org/10.1787/581dba7f-en

Dunne, P. G., Moore, M. J., & Portes, R. (2007). Benchmark status in fixed-income asset markets. *Journal of Business Finance & Accounting*, *34*(9-10), 1615-1634. doi: https://doi.org/10.1111/j.1468-5957.2007.02039.x

Ferri, G., Liu, L. G., & Stiglitz, J. E. (1999). The procyclical role of rating agencies: Evidence from the East Asian crisis. *Economic notes*, *28*(3), 335-355. doi: https://doi.org/10.1111/1468-0300.00016

Gallo, C. (2005). *Artificial Neural Networks in Finance Modelling*. University Library of Munich, Germany. Retrieved from https://econwpa.ub.uni-muenchen.de/econ-wp/exp/papers/0509/0509002.pdf

Gogas, P., Papadimitriou, T., & Agrapetidou, A. (2014). Forecasting bank credit ratings. *The Journal of Risk Finance*, *15*(2), 195-209. doi: https://doi.org/10.1108/JRF-11-2013-0076

Greene, W. H. (2012). *Econometric Analysis*. New York: Pearson Addison Wesley.

Gu, X., Kadiyala, P., & Mahaney-Walter, X. W. (2018). How creditor rights affect the issuance of public debt: The role of credit ratings. *Journal of Financial Stability*, *39*, 133-143. doi: https://doi.org/10.1016/j.jfs.2018.11.001

Gurný, P., & Gurný, M. (2013). Comparison of credit scoring models on probability of default estimation for us banks. *Prague Economic Papers*, *22*(2), 163-181. doi: https://doi.org/10.18267/j.pep.446

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer. doi: https://doi.org/10.1007/978-0-387-21606-5

Heryán, T., & Tzeremes, P. G. (2017). The bank lending channel of monetary policy in EU countries during the global financial crisis. *Economic Modelling*, *67*, 10-22. doi: https://doi.org/10.1016/j.econmod.2016.07.017

Hill, P., Brooks, R., & Faff, R. (2010). Variations in sovereign credit quality

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

assessments across rating agencies. *Journal of Banking & Finance*, *34*(6), 1327-1343. doi: https://doi.org/10.1016/j.jbankfin.2009.11.028

Iannotta, G., Nocera, G., & Resti, A. (2013). Do investors care about credit ratings? An analysis through the cycle. *Journal of Financial Stability*, *9*(4), 545-555. doi: https://doi.org/10.1016/j.jfs.2012.11.006

Iyengar, S. (2010). Are Sovereign Credit Ratings Objective and Transparent? *IUP Journal of Financial Economics*, *8*(3), 7-22. Retrieved from https://ssrn.com/abstract=1697494

Kabadayi, B., & Çelik, A. A. (2015). Determinants of Sovereign Ratings in Emerging Countries: Qualitative Dependent Variables Panel Data Analysis. *International Journal of Economics and Financial Issues*, *5*(3), 656-662. Retrieved from https://www.econjournals.com/index.php/ijefi/article/view/1198/pdf

Kiff, M. J., Nowak, S., & Schumacher, M. (2012). *Are rating agencies powerful? An investigation into the impact and accuracy of sovereign ratings*. International Monetary Fund. Retrieved from https://www.elibrary.imf.org/view/journals/001/2012/023/article-A001-en.xml

Kräussl, R. (2005). Do credit rating agencies add to the dynamics of emerging market crises? *Journal of Financial Stability*, *1*(3), 355-385. doi: https://doi.org/10.1016/j.jfs.2005.02.005

Kumar, K., & Haynes, J. D. (2003). Forecasting credit ratings Using ANN and statistical techniques. *International journal of business studies*, *11*(1), 91-108. Retrieved from http://epublications.bond.edu.au/business_pubs/347

Kume, O. (2012). *Determinants of US corporate credit spreads* [Doctoral dissertation, Robert Gordon University]. Retrieved from http://hdl.handle.net/10059/735

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, *104*(2), 271-294. doi: https://doi.org/10.1007/s00607-021-00908-y

León-Soriano, R., & Muñoz-Torres, M. J. (2012, May). Using neural networks to model sovereign credit ratings: Application to the European Union. In *International Conference on Modeling and Simulation in Engineering, Economics and Management* (pp. 13-23). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: https://doi.org/10.1007/978-3-642-30433-0_3

Liu, Y. (2002). *A framework of data mining application process for credit scoring* (Arbeitsbericht Nr. 01/2002). Institut für Wirtschaftsinformatik. Retrieved from https://webdoc.sub.gwdg.de/ebook/lm/arbeitsberichte/2002/01.pdf

Lubis, A. R., & Lubis, M. (2020). Optimization of distance formula in K-Nearest Neighbor method. *Bulletin of Electrical Engineering and Informatics*, *9*(1), 326-338. doi: https://doi.org/10.11591/eei.v9i1.1464

Ly, H.-B., Nguyen, T.-A., & Pham, B. T. (2021). Estimation of soil cohesion using machine learning method: A random forest approach. *Advances in civil* *engineering*, *2021*, 1-14. doi: https://doi.org/10.1155/2021/8873993

Mahomed Karodia, A., & Soni, D. (2014). South African economic woes: Poor political leadership and rating downgrades hampers growth and development. *Management Studies and Economic Systems*, *1*(1), 51-66. doi: https://doi.org/10.12816/0006205

Mathuria, M., Bhargava, N., Sharma, G., & Bhargava, R. (2013). Decision Tree Analysis on J48 Algorithm for Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(6), 1114-1119. Retrieved from https://www.academia.edu/download/31975379/V3I6-0408.pdf

Matthews, M., & Mokoena, B. A. (2020). The influence of service quality dimensions on customer satisfaction within visa facilitation centres in South Africa. *International Journal of eBusiness and eGovernment Studies*, *12*(2), 122-135. doi: https://doi.org/10.34111/ijebeg.202012203

Meharie, M. G., & Shaik, N. (2020). Predicting highway construction costs: comparison of the performance of random forest, neural network and support vector machine models. *Journal of Soft Computing in Civil Engineering*, *4*(2), 103-112. doi: https://doi.org/10.22115/SCCE.2020.226883.1205

Mellios, C., & Paget-Blanc, E. (2006). Which factors determine sovereign credit ratings? *The European Journal of Finance*, *12*(4), 361-377. doi: https://doi.org/10.1080/13518470500377406

Metz, A., & Tudela, M. (2015). The Price Impact of Sovereign Rating Announcements. In *Emerging Markets and Sovereign Risk* (pp. 275-292). Springer. doi: https://doi.org/10.1057/9781137450661_15

Meyer, D. F., & Mothibi, L. (2021). The effect of risk rating agencies decisions on economic growth and investment in a developing country: The case of South Africa. *Journal of Risk and Financial Management*, *14*(7), 288. doi: https://doi.org/10.3390/jrfm14070288

Mora, N. (2006). Sovereign credit ratings: Guilty beyond reasonable doubt? *Journal of Banking & Finance*, *30*(7), 2041-2062. doi: https://doi.org/10.1016/j.jbankfin.2005.05.023

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press. Retrieved from https://mitpress.mit.edu/9780262018029/machine-learning/

Mutize, M., & Nkhalamba, M. M. (2020). The impact of sovereign credit rating changes on government bond yields in South Africa. *International Journal of Sustainable Economy*, *12*(1), 81-100. doi: https://doi.org/10.1504/IJSE.2020.107860

Mutize, M., & Nkhalamba, M. P. (2021). International credit rating agencies in Africa: perceptions, trends and challenges. *International Journal of Sustainable Economy*, *13*(1), 55-71. doi: https://doi.org/10.1504/IJSE.2021.113303

Nazari, M., & Alidadi, M. (2013). Measuring credit risk of bank customers using artificial neural network. *Journal of Management Research*, *5*(2), 1-17. doi: https://doi.org/10.5296/jmr.v5i2.2899

Niaki, S. T. A., & Hoseinzade, S. (2013). Forecasting S&P 500 index using artificial neural networks and design of experiments. *Journal of Industrial Engineering International*, *9*, 1-9. doi: https://doi.org/10.1186/2251-712X-9-1

Novotná, M. (2012, October). The use of different approaches for credit rating prediction and their comparison. In *Proceedings of the 6th International Conference on Managing and Modelling of Financial Risks* (pp. 448-457). Retrieved from https://ssrn.com/abstract=2867849

Okasha, M. K. (2014). Using support vector machines in financial time series forecasting. *International Journal of Statistics and Applications*, *4*(1), 28-39. doi: https://doi.org/10.5923/j.statistics.20140401.03

Osobajo, O. A., & Akintunde, A. E. (2019). Determinants of sovereign credit ratings in emerging markets. *International Business Research*, *12*(5), 142-166. doi: https://doi.org/10.5539/ibr.v12n5p142

Overes, B. H., & van der Wel, M. (2023). Modelling sovereign credit ratings: Evaluating the accuracy and driving factors using machine learning techniques. *Computational Economics*, *61*(3), 1273-1303. doi: https://doi.org/10.1007/s10614-022-10245-7

Ozturk, H., Namli, E., & Erdal, H. I. (2016). Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. *Economic Modelling*, *54*, 469-478. doi: https://doi.org/10.1016/j.econmod.2016.01.012

Pacelli, V., & Azzollini, M. (2011). An artificial neural network approach for credit risk management. *Journal of Intelligent Learning Systems and Applications*, *3*(02), 103-112. doi: https://doi.org/10.4236/jilsa.2011.32012

Phoenix, P., Sudaryono, R., & Suhartono, D. (2021). Classifying promotion images using optical character recognition and Naïve Bayes classifier. *Procedia Computer Science*, *179*, 498-506. doi: https://doi.org/10.1016/j.procs.2021.01.033

Pİrdal, B. (2017). A Comparative analysis of sovereign credit rating methods and credit default swaps. *Bolu Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *17*(4), 107-124. Retrieved from https://dergipark.org.tr/en/pub/basbed/issue/38799/459070

Polito, V., & Wickens, M. (2015). Sovereign credit ratings in the European Union: a model-based fiscal analysis. *European Economic Review*, *78*, 220-247. doi: https://doi.org/10.1016/j.euroecorev.2015.05.009

Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, *5*(2), 221. doi: https://doi.org/10.22364/bjmc.2017.5.2.05

Ramayah, T., Ahmad, N. H., Halim, H. A., & May-Chiun, S. R. M. Z. (2010). Discriminant analysis: An illustrated example. *African Journal of Business*

Management, *4*(9), 1654-1667. Retrieved from https://academicjournals.org/journal/AJBM/article-full-text-pdf/1BF418E31373.pdf

Ramesh, T., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, 132-148. doi: https://doi.org/10.22452/mjcs.sp2022no1.10

Ratha, D., De, P. K., & Mohapatra, S. (2011). Shadow sovereign ratings for unrated developing countries. *World development*, *39*(3), 295-307. doi: https://doi.org/10.1016/j.worlddev.2010.08.006

Romadhon, M. R., & Kurniawan, F. (2021, April). A comparison of naive Bayes methods, logistic regression and KNN for predicting healing of Covid-19 patients in Indonesia. In *2021 3rd east Indonesia conference on computer and information technology (eiconcit)* (pp. 41-44). IEEE. doi: https://doi.org/10.1109/EIConCIT50028.2021.9431845

Saadaoui, A., Elammari, A., & Kriaa, M. (2022). Credit rating announcement and bond liquidity: the case of emerging bond markets. *Journal of Economics, Finance and Administrative Science*, *27*(53), 86-104. doi: https://doi.org/10.1108/JEFAS-08-2020-0314

Slabbert, A., Keeton, G., & Cattaneo, N. (2019). *Investment-grade or "junk" status: Do sovereign credit ratings really matter* [Masters Thesis, Rhodes University; Faculty of Commerce, Economics and Economic History]. Retrieved from http://hdl.handle.net/10962/97067

Steinberg, D., & Cardell, N. S. (1998). The hybrid CART-Logit model in classification and data mining. *Salford Systems White Paper*, *42*, 1-7. Retrieved from https://www.researchgate.net/publication/265631242

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (Vol. 6). pearson Boston, MA. Retrieved from https://www.pearsonhighered.com/assets/preface/0/1/3/4/0134790545.pdf

Takawira, O. (2022). *An assessment of the impact of SRMs on the South African economy* [Doctoral dissertation, University of Johannesburg]. Retrieved from https://hdl.handle.net/10210/503100

Tangirala, S. (2020). Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*. *International Journal of Advanced Computer Science and Applications, 11*. https://doi.org/10.14569/ijacsa.2020.0110277

Trading Economics, T. (2017). China GDP annual growth rate. *Online at: http://www. tradingeconomics. com/china/gdp-growth-annual. Last visit: February*, *28*(2017), 11. Retrieved from https://tradingeconomics.com/china/gdp-growth-annual

Valencia, G. A. D. (2020). Forms of informal financing of informal traders in Colombia Cases: Cúcuta, Ibagué and Villavicencio. *Cuadernos de Economía*, *43*(123), 259-274. doi: https://doi.org/10.32826/cude.v43i123.126

Valle, C. T., & Marín, J. L. M. (2005). Sovereign credit ratings and their determination by the rating agencies. *Investment management and financial innovations*, *2*(4), 159-173. Retrieved from https://www.researchgate.net/publication/265187211

Van der Heide, E., Veerkamp, R., Van Pelt, M., Kamphuis, C., Athanasiadis, I., & Ducro, B. (2019). Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *Journal of dairy science*, *102*(10), 9409-9421. doi: https://doi.org/10.3168/jds.2019-16295

Vu, H., Alsakka, R., & ap Gwilym, O. (2022). Does competition improve sovereign credit rating quality? *Journal of International Financial Markets, Institutions and Money*, *76*, 101478. doi: https://doi.org/10.1016/j.intfin.2021.101478

Weyers, K., & Eliott, A. (2017). Checkmate: South Africa's credit rating downgraded to “junk” status. *Without Prejudice*, *17*(4), 10-13. Retrieved from https://hdl.handle.net/10520/EJC-7fe320f11

Xu, H., Przystupa, K., Fang, C., Marciniak, A., Kochan, O., & Beshley, M. (2020). A combination strategy of feature selection based on an integrated optimization algorithm and weighted k-nearest neighbor to improve the performance of network intrusion detection. *Electronics*, *9*(8), 1206. doi: https://doi.org/10.3390/electronics9081206

Zacharis, N. Z. (2018). Classification and regression trees (CART) for predictive modeling in blended learning. *International Journal of Intelligent Systems and Applications*, *3*(1), 1-9. doi: https://doi.org/10.5815/ijisa.2018.03.01

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

Appendices

A1: S&P, Moody's and Fitch rating.

Table A1: S&P, Moody's and Fitch Rating Systems.

	S&P	Moody's	Fitch
Prime	AAA	Aaa	AAA
High grade	AA+	Aa1	AA+
	AA	Aa2	AA
	AA-	Aa3	AA-
Upper medium grade	A+	A1	A+
	A	A2	A
	A-	A3	A-
Lower medium grade	BBB+	Baa1	BBB+
	BBB	Baa2	BBB
	BBB-	Baa3	BBB-
Non-investment grade speculative	BB+	Ba1	BB+
	BB	Ba2	BB
	BB-	Ba3	BB-
Highly speculative	B+	B1	B+
	B	B2	B
	B-	B3	B-
Substantial risks Extremely speculative - In default with little prospect for recovery	CCC+	Caa1	CCC+
	CCC	Caa2	CCC
	CCC-	Caa3	CCC-
	CC	Ca	CC; C
In Default	SD; D	C	DDD; DD; D

Source: Chee et al. (2015); Afonso et al. (2011), Mellios and Paget-Blanc (2006), Hill et al. (2010) and Trading Economics (2017).

B1: Gini Index

According to Tangirala, (2020) the Gini index determines the purity of a specific class after splitting along a particular attribute. The best split increases the purity of the sets resulting from the split. At node t the Gini index is defined:

$$\mathrm{Gini}(t) = 1 - \sum_{j=0}^{1} \left(\frac{n(j|t)}{n(t)} \right)^2 \quad (9)$$

where j is a class of target variable (in this study $j = 0$ means failure and $j = 1$ denotes success), $n(j|t)$ is the number of records of node t belonging to class j, and $n(t)$ is the total record number in node t. When the data in a node are equally distributed between all classes, the Gini index attains its maximum impurity value .5. In the case where all data belong to the same class, the node has minimum impurity and the Gini index is 0. In order to decide which attribute to split upon, the tree growing algorithm calculates the weighted average of the Gini index for the descended nodes.

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

C1: Ramsey RESET Test Results

Table C1: Fitch Ratings.

Ramsey RESET Test				
Equation: UNTITLED				
Specification: FR GDPPC UR REER PIR HDDI FDGDP CAB BOP				
CPIH C				
Omitted Variables: Squares of fitted values				
	Value	df	Probability	
t-statistic	0.337029	69	0.7371	
F-statistic	0.113588	(1, 69)	0.7371	
Likelihood ratio	0.131588	1	0.7168	
F-test summary:				
	Sum of Sq.	df	Mean Squares	
Test SSR	0.023387	1	0.023387	
Restricted SSR	14.23011	70	0.203287	
Unrestricted SSR	14.20672	69	0.205895	
Unrestricted SSR	14.20672	69	0.205895	
LR test summary:				
	Value	df		
Restricted LogL	-44.44842	70		
Unrestricted LogL	-44.38262	69		
Unrestricted Test Equation:				
Dependent Variable: FR				
Method: Least Squares				
Sample: 1999Q1 2022Q4				
Included observations: 96				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDPPC	0.168459	0.054328	3.100773	0.0028
UR	-0.075636	0.049672	-1.522706	0.1324
REER	0.003097	0.008940	0.346408	0.7301
PIR	0.072707	0.033866	2.146930	0.0353
HDDI	-0.019945	0.008941	-2.230579	0.0290
FDGDP	0.045979	0.023866	1.926544	0.0582
CAB	-1.02E-05	6.76E-06	-1.512249	0.1350
BOP	2.10E-06	2.34E-06	0.900330	0.3711
CPIH	0.006140	0.007156	0.858022	0.3939
C	0.296169	1.897190	0.156109	0.8764
FITTED^2	0.097706	0.289904	0.337029	0.7371
R-squared	0.479548	Mean dependent var		-0.156250
Adjusted R-squared	0.404120	S.D. dependent var		0.587818
S.E. of regression	0.453756	Akaike info criterion		1.384566
Sum squared resid	14.20672	Schwarz criterion		1.712094
Log likelihood	-44.38262	Hannan-Quinn criter.		1.515881
F-statistic	6.357699	Durbin-Watson stat		1.125815
Prob(F-statistic)	0.000001			

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

Table C2: Standard & Poors' Ratings.

Ramsey RESET Test				
Equation: UNTITLED				
Specification: SPR GDPPC UR REER PIR HDDI FDGDP CAB BOP				
CPIH C				
Omitted Variables: Squares of fitted values				
	Value	df	Probability	
t-statistic	0.238628	69	0.8121	
F-statistic	0.056943	(1, 69)	0.8121	
Likelihood ratio	0.065994	1	0.7973	
F-test summary:				
	Sum of Sq.	df	Mean Squares	
Test SSR	0.008260	1	0.008260	
Restricted SSR	10.01671	70	0.143096	
Unrestricted SSR	10.00845	69	0.145050	
Unrestricted SSR	10.00845	69	0.145050	
LR test summary:				
	Value	df		
Restricted LogL	-30.40419	70		
Unrestricted LogL	-30.37119	69		
Unrestricted Test Equation:				
Dependent Variable: SPRO				
Method: Least Squares				
Sample: 1999Q1 2022Q4				
Included observations: 96				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDPPC	0.101051	0.056017	1.803938	0.0756
UR	-0.012827	0.041358	-0.310140	0.7574
REER	0.011169	0.007563	1.476824	0.1443
PIR	0.080435	0.032390	2.483368	0.0154
HDDI	-0.014174	0.008196	-1.729458	0.0882
FDGDP	0.030613	0.019139	1.599533	0.1143
CAB	1.72E-06	5.53E-06	0.310489	0.7571
BOP	-1.82E-06	1.96E-06	-0.925996	0.3577
CPIH	0.007273	0.006658	1.092339	0.2785
C	-1.991371	1.628096	-1.223129	0.2254
FITTED^2	-0.085530	0.358423	-0.238628	0.8121
R-squared	0.518825	Mean dependent var		-0.300000
Adjusted R-squared	0.449089	S.D. dependent var		0.513119
S.E. of regression	0.380854	Akaike info criterion		1.034280
Sum squared resid	10.00845	Schwarz criterion		1.361809
Log likelihood	-30.37119	Hannan-Quinn criter.		1.165595
F-statistic	7.439887	Durbin-Watson stat		0.856795
Prob(F-statistic)	0.000000			

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

Table C3: Moody's Ratings

Ramsey RESET Test				
Equation: UNTITLED				
Specification: MR BOP REER PIR CAB UR GDPPC HDDI FDGDP CPIH				
C				
Omitted Variables: Squares of fitted values				
	Value	df	Probability	
t-statistic	1.672573	69	0.0989	
F-statistic	2.797500	(1, 69)	0.0989	
Likelihood ratio	3.179452	1	0.0746	
F-test summary:				
	Sum of Sq.	df	Mean Squares	
Test SSR	0.243194	1	0.243194	
Restricted SSR	6.241554	70	0.089165	
Unrestricted SSR	5.998360	69	0.086933	
Unrestricted SSR	5.998360	69	0.086933	
LR test summary:				
	Value	df		
Restricted LogL	-11.48319	70		
Unrestricted LogL	-9.893461	69		
Unrestricted Test Equation:				
Dependent Variable: MR				
Method: Least Squares				
Sample: 1999Q1 2022Q4				
Included observations: 96				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
BOP	-7.23E-07	1.62E-06	-0.447685	0.6558
REER	0.013681	0.006058	2.258446	0.0271
PIR	0.083019	0.021311	3.895523	0.0002
CAB	-2.15E-06	4.55E-06	-0.473101	0.6376
UR	-0.040607	0.032521	-1.248626	0.2160
GDPPC	0.046987	0.026519	1.771876	0.0808
HDDI	-0.005695	0.005589	-1.019065	0.3117
FDGDP	0.059143	0.014705	4.022003	0.0001
CPIH	-0.000198	0.004434	-0.044580	0.9646
C	-1.720097	1.220706	-1.409101	0.1633
FITTED^2	-0.554352	0.331437	-1.672573	0.0989
R-squared	0.655706	Mean dependent var		-0.150000
Adjusted R-squared	0.605809	S.D. dependent var		0.469611
S.E. of regression	0.294844	Akaike info criterion		0.522337
Sum squared resid	5.998360	Schwarz criterion		0.849865
Log likelihood	-9.893461	Hannan-Quinn criter.		0.653652
F-statistic	13.14103	Durbin-Watson stat		0.797210
Prob(F-statistic)	0.000000			

Vol: 16 **No:** 01 **Year:** 2024 **ISSN:** 1309-8055 (Online) (**pp.** 23-62) **Doi:** 10.34109/ijefs.202416102

$$\mathrm{Gini}(t)_{\mathrm{split}} = \frac{n(t_L)}{n(t)}\mathrm{Gini}(t_L) + \frac{n(t_R)}{n(t)}\mathrm{Gini}(t_R) \quad (10)$$

Where $t_L$and t_R are the left and right child nodes of node t_R . The attribute that minimizes the $\mathrm{Gini}(t)_{\mathrm{split}}$ is chosen to split the node.

D1: Variance Inflation Factors

Table D1: SDRI.

Variance Inflation Factors			
Sample: 1999Q1 2022Q4			
Included observations: 96			
Coefficient	Uncentered	Centered	
Variable	Variance	VIF	VIF
BANKI	9.00E-05	1.263441	1.263395
BOP	1.23E-14	1.947874	1.947497
CAB	4.33E-14	2.350851	2.350331
FDGDP	0.001477	1.764344	1.721669
GDPPC	9.46E-06	1.240412	1.240412
GOVI	0.005955	1.908278	1.245219
HDDI	0.021752	1.211979	1.180613
ORI	5.98E-05	1.261796	1.261796
CPIH	5.64E-05	1.291963	1.291249
UR	0.003481	1.269367	1.267253
CONFI	4.14E-05	1.338219	1.337851
PIR	0.002625	1.180411	1.143670
REER	0.004315	1.794452	1.794420
ALBI	0.001284	1693.749	338.1902
ALSI	0.012400	1590.032	326.1298
BCO	0.000229	107.2134	21.70515
FBYGB	0.411215	409.3130	16.86921
FD	1.00E-10	153.8753	38.43782
FIPI	1.93E-07	1395.925	324.6765
GDPMP	1.68E-10	14908.58	2923.345
GDS	2.38E-10	542.2223	101.8114
HPDIR	0.002336	438.4545	17.18411
LIQ	2.23E-10	1506.495	570.2119
M3	4.84E-11	2136.188	479.9926
TT	0.021120	1885.876	38.18472
C	1.09E-05	1.816609	NA