

Financial Crisis Analysis and Prediction in Africa using Machine Learning and Statistical Learning Methods

Junda Chen*

SDU-ANU Joint Science College, Shandong University, Weihai, China

*Corresponding author: 202100700003@mail.sdu.edu.cn

Abstract. Africa has been prone to financial crises that hinder its development. Understanding the causes and predicting the financial crisis for Africa is important for investors and governments. This article aims to analyze those financial ratios that affect the crises and find the best machine learning model to predict possible financial crises in Africa. 11 different machine learning and statistical learning methods are applied to the African systemic crisis data for 13 African countries from 1860 to 2014 to find the most suitable model for financial crisis prediction. The financial ratios being considered are exchange rate, domestic debt, sovereign external debt, GDP, inflation rate, independence, currency crises and inflation crises. After visualizing various machine learning and statistical learning models, this essay concludes that support vector classifier with the “rbf” kernel is the best model after comparing the performance of those models with metrics that have the priority as follows: accuracy, f1 score, recall and precision. Finally, this article uses the support vector classifier to find that systemic crisis is the financial ratio that have the biggest effect on the African banking crises after shuffling the values of each ratio and analyzing the influence on the model's performance.

Keywords: Financial crisis; machine learning; support vector classifier; permutation importance.

1. Introduction

Africa, despite its vast economic potential, has been prone to financial crises that impede its development trajectory [1]. From currency devaluations to banking sector collapses, these crises have significant socioeconomic implications [2]. Understanding the root causes and predicting such crises is crucial for policymakers, investors, and stakeholders. This essay tries to examine the dynamics of financial crises in Africa, aiming to contribute to the body of knowledge on crisis analysis and prediction.

A number of models have been created to predict whether countries are in danger of bankruptcy and whether they should be estimated to have a good or bad credit risk [3]. Previous work has illustrated the importance of machine learning while those methods often attain a very high sample-fit [4]. Some warning system had been used to predict the banking crisis but not all the countries over the world are involved in the prediction [5]. Other applications of machine learning methods lie in the detection of rare events and achieved good prediction performance [6]. Similarly, some machine learning models were used to predict economic recession using the more precise algorithm [7]. Various people had utilized the online machine learning methods to forecast crisis in economic [8]. Some economists have realized that they should apply the AI and machine learning methods to financial prediction as the pioneering work during the early year of 21st century had suggested the potential of AI and machine learning algorithm [9]. In the method of using different machine learning classifiers, some researchers and combined more than one classifier to compare their performances although the number of the classifiers that were compared are small [10]. With the advent of neural networks, researchers had utilized the artificial neural networks to predict the economic crisis to the Turkish economy [11].

However, early models used in financial crisis prediction are not suitable for predicting the results that have much more automated manner and intelligence, so it's still important to develop new statistical method and provide more precise machine learning models [10]. All the global financial crisis has revealed the need for machine learning method, so in general machine and statistic learning

methods has performed well and should be used in predicting the financial crisis in other countries that are not included in the previous machine learning events like African countries [11].

This paper endeavors to apply various machine learning and statistical methods to the African systemic crisis data for 13 African countries from 1860 to 2014 to find the best machine learning model for financial crisis prediction in Africa. This essay first goes through 11 different machine learning and statistical learning models to find the suitable model for the African systemic crisis data after analyzing the distribution of 8 financial ratios in the data including exchange rate, domestic debt, sovereign external debt, GDP, inflation rate, independence, currency crises and inflation crises, then using different kernel for support vector machine to select the best model. In the end, this essay tries to predict the banking systemic crisis in Africa countries and find which financial ratio has the biggest effect on banking crisis.

2. Methods

2.1. Data Source

The data this paper use is a subset of global financial stability dataset which specifically put its emphasis on debt, banking, inflation, systemic crises and financial in 13 African countries from year 1860 to year 2014. The African countries included are Algeria, Angola, Central African Republic, Ivory Coast, Egypt, Kenya, Mauritius, Morocco, Nigeria, South Africa, Tunisia, Zambia and Zimbabwe. As shown in Table 1, the datasets illustrate 8 financial ratios considering the countries' domestic state.

Table 1. Financial ratios considering domestic state

Financial ratios	Numerical Explanation
Exch usd	The exchange rate of the country in USD per year
Domestic debt default	"0" means that no sovereign domestic debt occurred in the year, "1" means that a sovereign domestic debt happened
Sovereign external debt default	"0" means that no sovereign external debt occurred in the year and "1" means that a sovereign external debt happened
GDP weighted default	The total debt in default vis-a-vis the GDP
Inflation annual cpi	The annual CPI Inflation rate
Independence	"0" means independence and "1" means no independence
Currency crises	"0" means that no crises occurred in the year and "1" means that a crisis happened.
Inflation crises	"0" means that no crises occurred in the year and "1" means that a crisis happened.

At the last column, there states whether the country has a banking crisis or not in a certain year. This is the binary condition for the classification model using machine learning and statistical learning methods. The heatmap of all the financial ratios is shown in Figure 1:

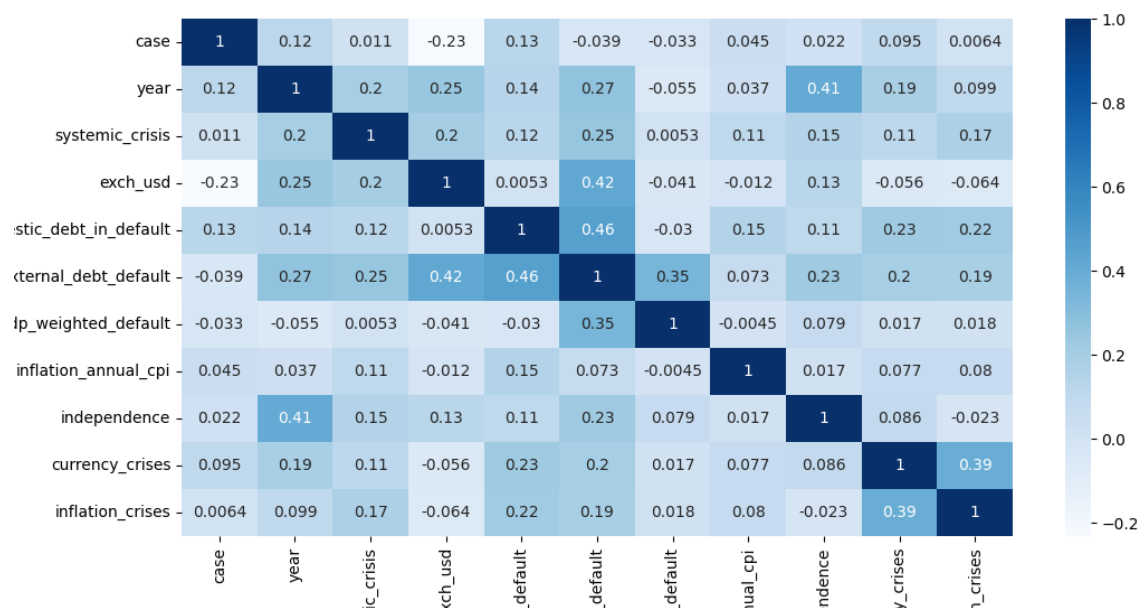


Fig. 1 The heatmap of African financial stability dataset

2.2. Data Structure and Description

The models this essay uses includes logistic regression, decision tree classifier, random forest classifier, extra trees classifier, k neighbors classifier, support vector classifier, bagging classifier, gradient boosting classifier, lgb.LGBM classifier, gaussian NB classifier and MLP classifier.

The evaluation metrics for the performance of models includes: accuracy, precision, recall, f1 score, specificity, balanced accuracy, matthews correlation coefficient and the roc score. Accuracy is an important evaluation metric that intuitively reflects the overall predictive accuracy of the classifier. Precision measures the proportion of true positive samples among those predicted as positive while recall evaluates the ability of the classifier to correctly identify positive samples. The f1 score is the harmonic mean of precision and recall, emphasizing both metrics' performance. Specificity measures the proportion of true negative samples among those predicted as negative. The Matthews correlation coefficient considers all four values of the confusion matrix so it's a comprehensive evaluation metric. The roc score measures the performance of the classifier at different thresholds which helps to evaluate the classifier's predictive ability on different classes.

2.3. Method Introduction

First, this essay separates the data into two parts: the banking crisis data which illustrates whether a certain country have banking crisis or not in the certain year and others datasets. Then this essay imports the machine learning and statistical learning models and separate the datasets onto two parts: training data which takes up 65 percent of the total data and the test data which contains the other 35 percent of dataset.

After applying the model on the data, this essay draws the heatmap for the four metrics of each model: true positive, false negative, false positive and true negative, then this essay illustrates the other metrics it has mentioned in the part 2.2, sorting to compare the performance of models with metrics that have the priority as follows: accuracy, f1 score, recall and precision as accuracy comes first when people want to precisely predict the banking crisis.

Eventually, this essay gets the best model for predicting the banking crisis in Africa and use this model to predict the banking crisis with the dataset. The process of model application is shown in Figure 2.

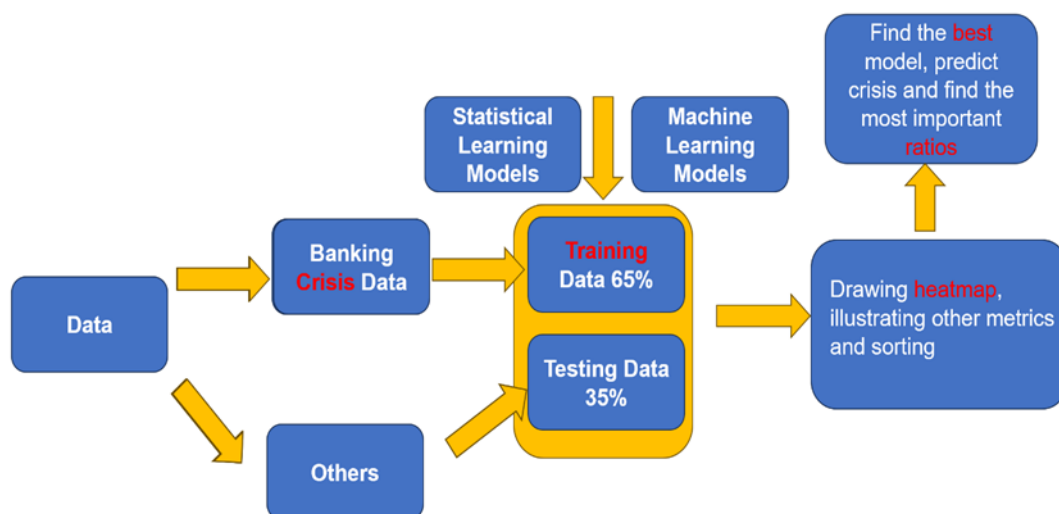


Fig. 2 The process of model application

3. Results and Discussion

3.1. Visualization of Model Performance

Figure 3 shows the roc curves for 9 machine and statistical learning models for 65 percent of training data, the roc curves for every picture are the curves that are blue. In addition, a red diagonal is also drawn in every picture to compare with the roc curve. It's illustrated that nearly all the machine learning models have good performance as the roc curves lie in the top left corner of the picture.

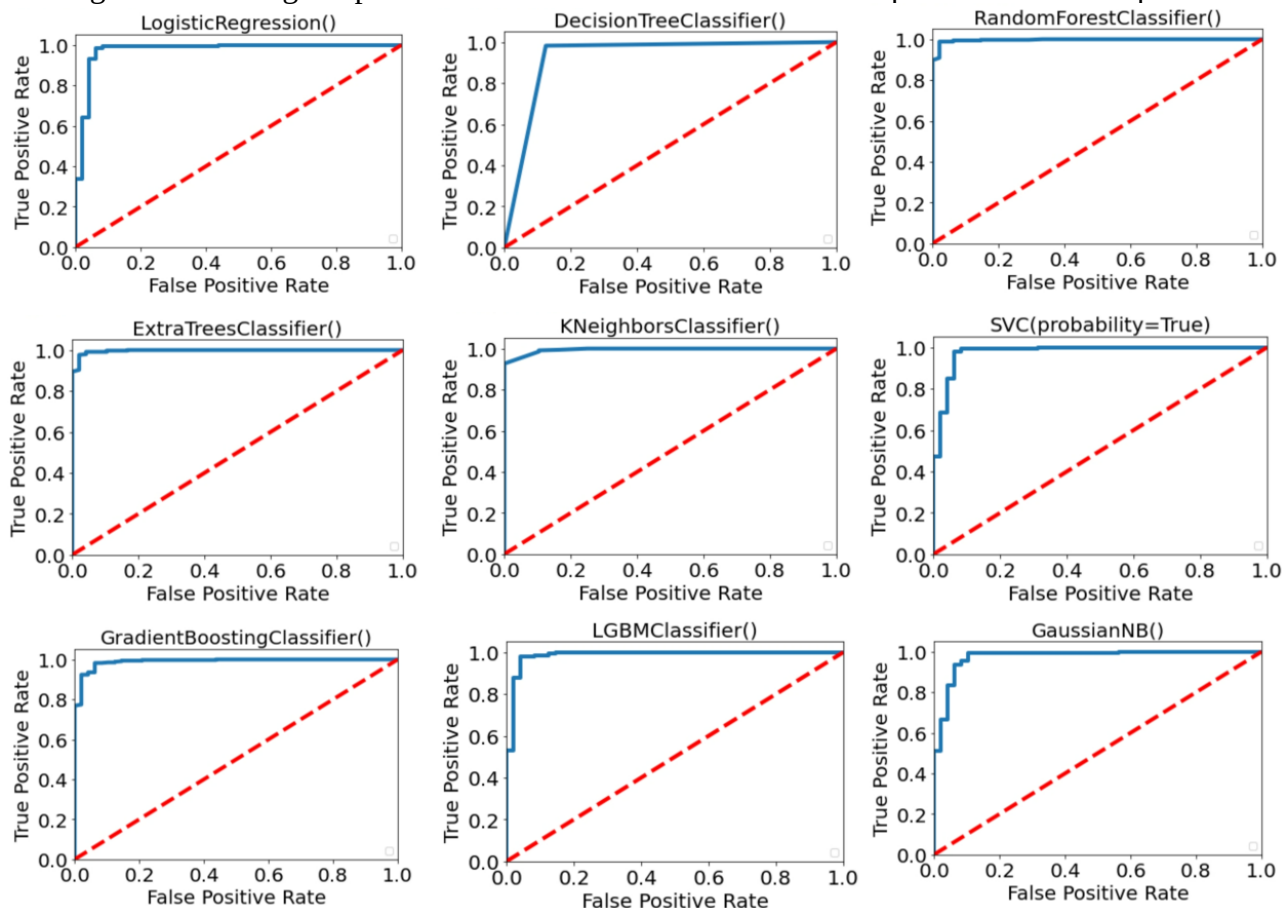


Fig. 3 The roc curves for 9 machine and statistical learning models for 65 percent of training data

The performance of the machine learning model this essay has applied on the datasets may include: accuracy, precision, recall, f1 score, specificity, balanced accuracy, matthews correlation coefficient and the roc score. This essay also draws the roc curves and the heatmaps to make the results more intuitive and change the proportion of the training data to take up 70 percent of the total data and eventually find that nearly all the models has increased their performances in various metrics like accuracy and precision.

However, if changing the proportion of the training data to 70 percent, the roc curves of the decision tree classifier and Gaussian NB verge to the diagonal of the picture, implying that those two models are not so good at predicting the financial crisis, as shown in Figure 4 and Figure 5.

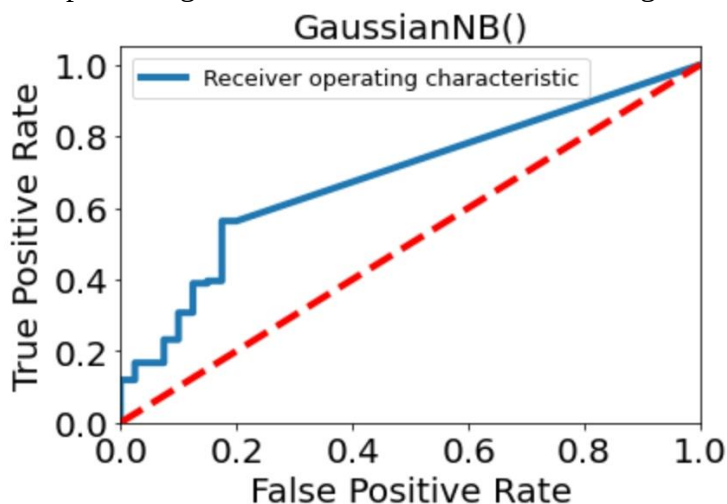


Fig. 4 Gaussian NB for 70% test data

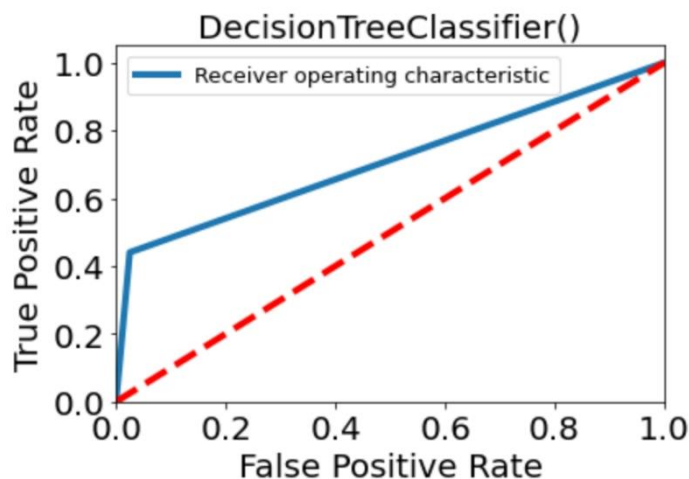


Fig. 5 Decision Tree Classifier for 70% test data

3.2. Sorting and Prediction Results

After visualizing and reporting various machine learning and statistic learning model, this essay sorts to compare the performance of models with metrics that have the priority as follows: accuracy, f1 score, recall and precision as accuracy comes first when people want to precisely predict the banking crisis. The results are shown in Table 2.

As it's illustrated in Table 2, support vector classifier and k neighbors classifier have the best accuracy which is 0.982, in addition, they have the best f1 score which is 0.990. While the recall of the support vector classifier is larger than k neighbors classifier, the k neighbors classifier has the largest precision of 0.991. Performances of logistic regression classifier and random forest classifier are nearly the same. The accuracy of the two models is 0.979, which is slightly lower than the support vector classifier and k neighbors classifier while their recall is the same as support vector classifier,

which is 0.994. All the metrics in these two models imply that they have good performances and can be utilized to predict African financial crisis accurately.

Table 2. Final sorting results

	models	Accuracy	Precision	Recall	F1 Score	True positive
0	SVC	0.982	0.985	0.994	0.990	339.0
1	KNeighbors	0.982	0.991	0.988	0.990	337.0
2	Logistic Regression	0.979	0.983	0.994	0.988	339.0
3	Random Forest	0.979	0.983	0.994	0.988	339.0
4	Bagging	0.977	0.980	0.994	0.987	339.0
5	Extra Trees	0.977	0.983	0.991	0.987	338.0
6	Gradient Boosting	0.972	0.974	0.994	0.984	339.0
7	LGBM	0.972	0.974	0.994	0.984	339.0
8	MLP	0.964	0.961	1.000	0.980	341.0
9	Decision Tree	0.961	0.974	0.982	0.978	335.0
10	Gaussian NB	0.925	0.991	0.924	0.956	315.0

Table 2 also shows the classification results of the bagging classifier and extra tree classifier. They have the same accuracy of 0.977 and f1 score of 0.987, while the recall of bagging classifier is slightly lower than extra trees classifier. Both the models have good performance as the four fundamental metrics: accuracy, precision, recall and f1 score are all larger than 0.95. However, the Gaussian NB and decision tree classifier have the slightly low performance because they are at the bottom of the sorting results.

The accuracies of other models are nearly the same while the Gaussian NB has the worst accuracy: 0.925. In general, nearly all the models have good performance while Gaussian NB and decision tree classifier performed slightly low.

Table 3. Prediction results

	Country	Year	Banking crisis	Predicted
423	Egypt	2000	No crisis	No crisis
427	Egypt	2004	No crisis	No crisis
23	Algeria	1947	No crisis	No crisis
332	Egypt	1909	No Crisis	No crisis
361	Egypt	1938	No crisis	No crisis
6	Algeria	1876	No crisis	No crisis
934	Zambia	1980	No crisis	No crisis
86	Angola	1922	No crisis	No crisis
674	Nigeria	1980	No crisis	No crisis
971	Zimbabwe	1923	No crisis	No crisis

Then this essay gets the best model for predicting the banking crisis in Africa and use this model to predict the banking crisis with the dataset. From the results illustrated in Table 3, it's clear that all the predictions are correct.

3.3. Permutation Importance of Financial Ratios

Eventually, this essay uses the support vector classifier to find which financial ratio has the biggest effect on banking crisis. As shown in Table 4, after shuffling the values of each feature one at a time and measuring the effect on the model's performance, this essay finds that systemic crisis has the

highest permutation importance. In addition, the domestic debt also has some effects on the African banking crisis while other financial ratios nearly have no influence on the crisis.

Table 4. Permutation importance of the financial ratios

Weight	Feature
0.1815 ± 0.0220	Systemic crisis
0.0108 ± 0.0021	Domestic debt
0 ± 0.0000	Inflation crises
0 ± 0.0000	Currency crises
0 ± 0.0000	Independence
0 ± 0.0000	Inflation annual cpi
0 ± 0.0000	Gdp weighted
0 ± 0.0000	Sovereign external debt
0 ± 0.0000	Exch usd
0 ± 0.0000	Year
0 ± 0.0000	Country
0 ± 0.0000	CC3
0 ± 0.0000	Case

4. Conclusion

In conclusion, it's clear that support vector machine is the best model. In addition, this essay changes the sorting sequence to be f1 score, accuracy recall and precision and find that support vector classifier is also the best machine learning model to predict the financial crisis in Africa. After utilizing the support vector classifier to predict the banking crisis in Africa, this essay gets the expected results that all the predictions are correct as performance of the models are tested to be exemplary. In addition, after changing all the kernels for the svc, this essay has found that the "rbf" kernel may be the most suitable. Finally, this essay uses the support vector classifier to find the financial ratio that has the biggest effect on African banking crises and gets that systemic crisis is the most important financial ratio. For further investigation, this model can be applied to different countries that are not just African countries and applied for other use such as predicting the currency of ocean.

References

- [1] Massa I, Willem D. The global financial crisis: will successful African countries be affected? World Development, 2008.
- [2] Of E. Devaluation and inflation after currency crises. Journal of Manipulative and Physiological Therapeutics, 2002, 588-590.
- [3] Lin W Y, Hu Y H, Tsai C F. Machine learning in financial crisis prediction: a survey. IEEE Transactions on Systems, 2012, 42(4): 421-436.
- [4] Beutel J, List S, von Schweinitz G. Does machine learning help us predict banking crises? Journal of Financial Stability, 2019, 45: 100693.
- [5] Casabianca E J, et al. An early warning system for banking crises: from regression-based analysis to machine learning techniques. Marco Fanno Working Papers, 2019.
- [6] Coffinet J, Kien J N. Detection of rare events: a machine learning toolkit with an application to banking crises. Journal of Finance and Data Science, 2019, 5(4): 183-207.
- [7] Jeremy F, Michael H, Hélène R, Vania S. Answering the queen: machine learning and financial crises. Angewandte Chemie International Edition, 2014, 6(11): 22-31.
- [8] Ramli N A, Ismail M T, Wooi H C. Measuring the accuracy of currency crisis prediction with combined classifiers in designing early warning system. Machine Learning, 2015, 101: 85-103.

- [9] Aydin A D, Cavdar S C. Prediction of financial crisis with artificial neural network: an empirical analysis on Turkey. *International Journal of Financial Research*, 2015, 6(4).
- [10] Christy S A, Arunkumar D R. Machine learning based classification models for financial crisis prediction. *International Journal of Recent Technology and Engineering (IJRTE)*, 2019, 8(4): 4887-4893.
- [11] Samitas A, Kampouris E, Kenourgios D. Machine learning as an early warning system to predict financial crisis. *International Review of Financial Analysis*, 2020, 71: 101507.