PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS

Reza Darmawan
Inspektorat Jenderal Kementerian Keuangan

Hermulia Hadie Putra
Inspektorat Jenderal Kementerian Keuangan

Rinto Fridayanto
Inspektorat Jenderal Kementerian Keuangan

Alamat Korespondensi: reza.darmawan@kemenkeu.go.id

ABSTRACT

The purpose of developing this predictive model is to predict the potential occurrence of risks in the implementation of projects financed through government loans. It uses data on Indonesian sovereign loans from the MoF DMFAS in the period between 1998 and 2019, consisting of 1930 government loans. Government loan performance is represented by two qualities: timeliness of loan disbursement and rate of loan realization. In the development of the model, the ensemble learning methodology was applied with the voting classifier algorithm. The algorithms used to make predictions for voting classifier include k-NN Classifier, Random Forest, and Logistic Regression. The predictive model developed can accurately predict 73.14% of observations on the rate of disbursement of government loans and is able to predict numerous amendments to drawing limit correctly as much as 89.69% of observations.

Tujuan dari pengembangan model prediktif ini adalah sebagai sarana untuk memprediksi potensi keterjadian risiko dalam pelaksanaan proyek yang dibiayai melalui pinjaman pemerintah. Pengembangan model prediktif ini menggunakan data pinjaman pemerintah RI yang bersumber dari aplikasi DMFAS DJPPR dalam rentang waktu antara tahun 1998 sampai dengan tahun 2019 dan terdiri atas 1930 pinjaman pemerintah. Kinerja pinjaman pemerintah direpresentasikan oleh dua kualitas: ketepatan waktu pencairan pinjaman dan tingkat realisasi pinjaman. Dalam pengembangan model, diaplikasikan metodologi *ensemble learning* dengan algoritma *voting classifier*. Model prediktif yang dikembangkan bisa memprediksi secara tepat 73,14% observasi atas tingkat pencairan pinjaman pemerintah dan mampu memprediksi amandemen berulang kali atas pinjaman pemerintah secara tepat sebanyak 89,69% observasi.

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **14**

1. INTRODUCTION

1.1. Research Backgrounds

Based on Indonesian National Budget (APBN) realization data as of June 30, 2021, the portion of loans compared to the composition of total outstanding sovereign debt was 11.63% with a value of Rp. 751.75 trillion, consisting of domestic loans of Rp. 12.6 trillion and foreign loans of Rp. 739.15 trillion. Although the portion is relatively small and the amount tends to decrease in the composition of total government debt, the need to procure loans in debt management has not ceased for the benefits that loans wield uniquely when compared to the financing obtained from the issuance of Government Securities (SBN).

Loans, especially foreign loans, exist not solely for the objective of deficit coverage (financing gap), but there are several considerations that form the basis for using loans, including:

- as part of government borrowing costs and risks management,
- to increase government capacity, especially for infrastructure financing, while encouraging the role of SOEs and the private sector,
- as an endeavor to develop project/activity models through replication of foreign loan projects/activities, and
- as an instrument of development cooperation.

With the decreasing portion of loans in government debt, the management and utilization of loans needs to be optimized, one of which is by increasing the effectiveness of loan procurement.

Based on data from the 2021 Quarter II Loan project realization Performance Report, it is known that of the 201 activities financed through active loans (domestic and foreign) there are 88 activities (44.0%) which are categorized as behind schedule and 28 activities (14.0%) are categorized at -risk. The status indicates that the activities financed through the loan experience delays in implementation from the schedule planned by the Ministry/Agency as the loan executing Agency.

The problem of delays in loan project realization oftenly became a spotlight (finding) by the Supreme Audit Agency (BPK) in various audits. Most recently, the BPK revealed this problem in the Central Government Debt Management Performance Audit Report published on December 31, 2019. Delays and even cancellation of projects financed by loans resulted in delayed output of activity results so that they could not immediately provide a multiplier effect on national development and eventually increased commitment fee for the loan.

The problem of delays and even cancellation of projects financed by the loan can be minimized through careful identification at the planning stage of loan acquisition before it is signed. The Minister of Finance has the authority ensure the readiness of the executing agency in implementing the project to be financed through the loan. Based on this, this research initiates the creation of a predictive model that is built based on historical data on government loans which is expected to help predict the potential occurrence of problems (risks) in the implementation of projects financed through government loans. The historical data used as predictors are taken from the Debt Management and Financial Analysis System (DMFAS) application managed by the Directorate General of Financing and Risk Management (DJPPR) containing loan characteristics, lender characteristics, and Executing Agency characteristics.

The purpose of making this predictive model is to predict the potential occurrence of problems (risks) in the implementation of projects financed through government loans. Based on these objectives, in this study, the analysis targets for this predictive model have been set, namely the possibility of delays in project realization based on the value of disbursement (loan project realization) and the possibility of amendments to loan agreements that have the potential to increase the burden of government spending.

The analysis results from this predictive model are expected to assist business process owner (DJPPR) in considering the acquisition of a loan. In addition, this predictive model is also expected to assist management in preparing a monitoring and evaluation plan for the performance of the implementation of activities financed by government loans in a more measurable manner after considering the potential for future problems based on the prediction results.

2. LITERATURE REVIEW

One of the instruments in managing government debt is a loan. Loans can be defined as any financing that is bound in a loan agreement and must be repaid with certain conditions. Loans consist of domestic loans and foreign loans.

Domestic loans are used for certain activities of ministries/agencies in the context of empowering domestic industry and infrastructure development as well as to finance the activities of Regional Governments and SOEs in the scope of infrastructure development for public services and investment activities that generate revenue. In accordance with the laws and regulations, domestic loans can come from State-Owned Enterprises, Regional Governments and Regional Companies. Until 2021, creditors who have provided domestic loans consist of 6 SOEs, including Bank Mandiri, BNI, BRI, Bank Jabar and Banten, Bank DKI, and Bank Jateng.

In addition to being used for financing the national budget deficit, foreign loans are also used to finance priority activities of Ministries/Institutions,

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **15**

manage debt portfolios and can also be loaned or transferred to Regional Governments, and SOEs.

Government sovereign foreign loans can be acquired from various sources:

- **Bilateral Creditors**, governments or institutions acting in behalf of governments (e.g. Official Development Assistance-ODA) that provide loans to other governments.
- **Multilateral Creditors**, international financial institutions consisting of several countries that provide loans to the Government.
- **Foreign Private Creditors (KSA)**, that provide loans to the Government based on an unsecured loan agreement.
- **Export Credit Guarantee Agency (LPKE),** which is an institution appointed by a foreign country to provide guarantees, insurance, direct loans, interest subsidies, and financial assistance to increase exports of the aforementioned country. Usually, the largest part of the funds provided is specifically used to purchase goods/services from the partner country.

In the loan management process, there is an imminent risk of delays in loan project realization, thereby increasing the financing costs. The delay in loan project realization, may results from a variety of causes including:

- Ministries/Agencies budget revision.
- Cancellation of the activity leads thus in turn leads to the cancellation of the loan.
- Loan refinancing due to the time constraints regarding the completion of the project that is estimated to be exceeding the loan closing date.

The risk of delays and even failure of project realization that is financed through the loan has actually been attempted to be mitigated through trilateral meetings between the Ministry of Finance, Bappenas, and Ministries/Agencies as executing agencies in endeavor to examine the proposed project budget plans and loan disbursement. In addition, a review is carried out on the disbursement plan made by the executing agency. As well as loan monitoring and evaluation activities that are routinely carried out by DJPPR.

However, it is necessary to have other control instruments that function as an early warning system for the government, especially the DJPPR as the manager of government loans, regarding the risk of failure to implement activities financed through loans. One of them is through the development of a predictive model of loan realization performance based on historical data on the implementation of activities financed through loans. If a project is then predicted to potentially encounter a realization problem based on the predictive model, either in terms of disbursement and timeliness of realization, mitigation can be carried out before the loan is negotiated with lenders. For example, an in-depth

check on the readiness of the Ministry/Agency as the executing agency for the loan. Mitigating these problems from the upstream side of loan procurement, would be better than if the Government had to bear the financial risk in the form of additional costs that must be incurred when the loan was realized late or canceled.
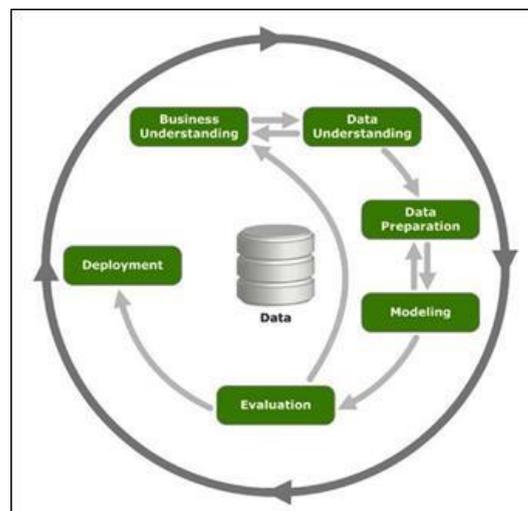
3. METHODOLOGY

Figure 1. CRISP-DM Methodology Commonly Applied in Machine Learning

The methodology commonly used in data mining projects is the Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM divides the machine learning model development stages into a cycle consisting of six processes.

The stage in CRISP-DM starts from understanding business processes and available data. Next is the data preparation process for later model development. After the model is developed, an evaluation of the accuracy of the model will be carried out. After that, then the model is deployed.

In the development of this predictive model, a loan with a risky withdrawal behavior will be predicted based on consideration of the factors (features), as a predictive variable. Some of the features that will be used are:

- **Lender Characteristics**. There are three characteristics of lenders used in the development of predictive models. The first characteristic of lenders is to identify whether the loan is sourced from abroad or from within the country. The second characteristic is the type of lender, which consists of multilateral institutions, foreign countries (bilateral), foreign private creditors (KSA), or domestic state-owned banks. In addition, the characteristics of lenders are also constructed from clustering analysis of

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **16**

the amount of loans granted and the frequency of lending to the Government of the Republic of Indonesia.

- **Characteristics of Executing Agency (EA)**. The characteristics of the EA are identified based on the codification of the tasks and functions of the EA in the realm of public policy as well as the results of a cluster analysis of the history of projects that have been carried out and the intensity of the EA in carrying out loan-financed projects.
- **Loan amount** converted into USD.
- **Availability period**. It is calculated from the time the loan is declared effective until the deadline for disbursing the loan.
- **Types of projects financed through loan**s. Codified based on the objectives of the project being financed. Namely the purchase of goods, government programs, or infrastructure development.

The development of this predictive model uses Indonesian government loan data sourced from the DJPPR DMFAS application. The dataset is formed using loan population data, with instrument limits signed from 1998 to 2019. This is to ensure that there are no significant differences in characteristics between each test data.

Based on these criteria, a dataset consisting of 1930 government loans was formed. In terms of quantity, 1284 loans (92.4%) were loans with the executing agency of the Central Government Ministries/Agencies. The remaining 106 loans (7.63%) are loans with the executing agency State-Owned Enterprises (BUMN).

The predictive model in this study was developed to predict the delay in the realization of projects financed through government loans. Delay in project realization, for the means of model development, is determined to be proxied as two target variables (classes): the disbursement category and numerous drawing limit amendments.

The disbursement category variable represents the amount of the loan that has been disbursed from the total value of the loan commitment agreed in the loan agreement, and is divided into 3 (three) categories:

- Loans with disbursement below 50% of the commitment value (DISBURSEMENT UNDER 50%)
- Loans with disbursement between 50-90% of the commitment value (DISBURSEMENT UNDER 90%)
- Loans with disbursement of more than 90% of the commitment value (FULLLY DISBURSED)

Meanwhile, numerous drawing limit amendments is a variable with binary values which represents whether the loan has been amended several times over the disbursement deadline. The criterion for this variable is as follows:

- True, if the disbursement period of the loan has been amended more than twice
- False, if the disbursement period of the loan is amended a maximum of two times

Because there are two types of classes to be predicted, two predictive models have been developed. Both predictive models are included in the supervised learning family, namely classification. The model will make predictions by grouping the test data, in this case the instrument or loan agreement, into the predefined categories.

In developing the model, the ensemble learning methodology was applied. Ensemble learning uses multiple machine learning models to make better predictions over the dataset. Through this method, there is a set of different models applied to a single dataset, and each model makes its predictions individually. Predictions from each model are combined in a model ensemble which is used to make the final prediction. It is hoped that the resulting model ensemble will have better accuracy when compared to individual models with different algorithms previously applied.
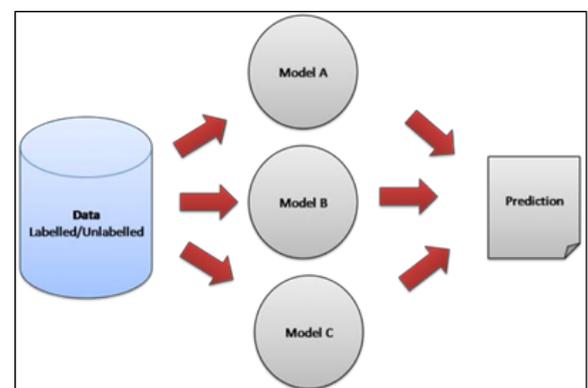

Figure 2. Ensemble Learning Illustration

Each model has its own strengths and weaknesses. Ensemble models can be used to help hide the weaknesses of each model. In the development of this predictive model, the Voting Classifier is used, which is an ensemble model that makes the final prediction using the most votes. For example, if three models are used each predicting [1, 0, 1] for the target variable, the predicted model will be 1, because two of the three models predict 1.

The machine learning algorithms used to form the ensemble model in this study are 3 classification algorithms. Includes Random Forest, kNN Classifier, and Logistic Regression.

4. RESULT AND ANALYSIS

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **17**

4.1 Data Understanding

Based on the data used, in terms of quantity, 1284 loans (92.4%) were loans with the executing agency of the Central Government Ministries/Agencies. The remaining 106 loans (7.63%) are loans with the executing agency State-Owned Enterprises (BUMN).

Based on the results of data processing (Graph 1), it is known that there are 5 (five) executing agencies with the highest number of loans, namely the Ministry of Defense (495 loans), the National Police Republic of Indonesia/Polri (192 loans), the Ministry of Civil Works and Public Housing (131 loans), the Ministry of Finance (100 loans) and PT PLN (72 loans). Meanwhile, in terms of numbers, the Ministry of Defense and the National Police are also the executing agencies with the largest loan value. The Ministry of Defense, in 1998-2019 has carried out activities financed through loans of USD 115,662 trillion. Meanwhile, National Police's loan amounted to USD 40,887.66 trillion.

In terms of lenders, foreign institutions are the parties that provide the most loans to the Government of Indonesia, amounting to 1138 loans. Foreign Private Creditors were the lenders with the largest number of loans, with a total of 448 loans signed (32.2%). Next is the Multilateral Institution with a total of 368 loans (26.5%) and the State (Bilateral) as many as 240 loans (17.3%). The rest are foreign loans originating from the Export Credit Guarantee Agency (LPKE) with a total of 82 loans signed (5.9%). From domestic sources, there were loans from state-owned banks with 187 loans (13.5%) and domestic private banks with 65 loans (4.68%).

However, when viewed in terms of the amount (Graph 2), most of these loans were originated from domestic creditors. Amounting to USD 125,760 trillion (80%) were loans originating from Domestic State-Owned Banks and USD 30,766.86 trillion (19.6%) were loans originating from Domestic Private Banks. This is because the Ministry of Defense and the National Police, as the executing agency with the most loans, preferred using domestic banks as a source of loans for the procurement of defense equipment.

In the other hand, taking a closer look on the 5 largest lenders in terms of number (Graph 3), JICA is the creditor with the largest number of loans (108 loans). Next are the International Bank of Reconstruction and Development (IBRD) and the Asia Development Bank (ADB) with 107 loans. And the last one is Bank Mandiri and Bank Negara Indonesia with 84 loans.

However, in terms of value (Graph 4), the 5 lenders with the largest loan value are all domestic banking institutions. Bank Mandiri is the lender with the largest loan value of USD 58,535.61 trillion. Next is

Bank Negara Indonesia with a loan value of USD 57,946.13 trillion. The rest are Bank Jabar Banten (BJB) with a loan value of USD 14,494.17 trillion, Bank DKI for USD 13,838.54 trillion and Bank BRI for USD 9,278.57 trillion.

Graph 5. Loan Project Realization Performance based on Numerous Drawing Limit Amendment criterion
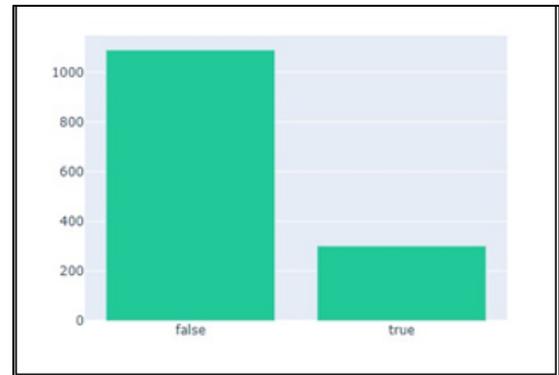
As already mentioned, there are 2 measures that are used as proxies for the performance of government loans, the timeliness of loan project realization and the percentage of disbursement which shows the realization of projects financed through these loans. The timeliness of loan disbursement can be seen from the amendments to the drawing limit. It is known that 1089 loans did not undergo more than 3 times amendments to the drawing limit (Graph 5). However, there were 301 loans that experienced more than 3 amendments to the drawing limit.

With regards to the amendment to the drawing limit, the executing agency that has made the most amendments to the loans drawing limit for more than 3 (three) times (Graph 6) is the Ministry of Defense (112 loans) and the National Police (53 loans). Next is the Ministry of Civil Works and Public Housing (24 loans), the Ministry of Education and Culture (15 loans), and the Ministry of Health (13 loans).

Graph 7. Loan Project Realization Performance based on Disbursement Ratio criterion
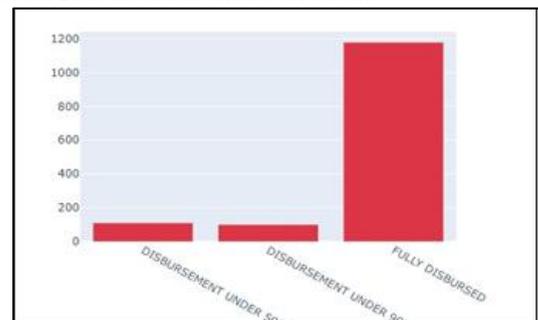
In terms of loan disbursement performance, most of the loans (1178 loans) were fully disbursed, or in this measurement defined as having a disbursement rate of more than 90% of the loan commitment value (Graph 7). Meanwhile, there are 101 loans that have a Disbursement status of Under

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **18**

90% or which in this measurement is defined as loans that have a realization rate between 50-90% of the commitment value. The remaining 111 loans are loans with low disbursement, or below 50% of the commitment value.

If analyzed further (Graph 8), the 5 (five) executing agencies historically have the most loans with low disbursement, the first being the Ministry of Defense (27 loans). Next are the Police (17 loans), the Ministry of Civil Works and Public Housing (15 loans), PT PLN (12 loans), and the Ministry of Transportation (8 loans).

4.2 Data Preparation

4.2.1 Feature Encoding

Predictors (features) in predictive models can be divided into two types based on their shape: continuous features and categorical features. Continuous features are predictors that are numerical. Meanwhile, categorical features are discrete predictors. Data preparation is carried out by transforming predictors that are categorical, namely project type, loan type, EA characteristics, and lender characteristics.

There are various ways to encode categorical features. One of them, the most widely used is One Hot Encoder. One Hot Encoder converts each value in a categorical feature into individual features. The goal is that machine learning algorithms do not mistranslate these categorical variables as numeric variables.

4.2.2 Clustering Analysis

In addition, there are several features that are the result of cluster analysis. Namely features to reflect the characteristics of lenders and executing agencies. Therefore, at this feature engineering stage, a cluster analysis of lenders and executing agencies is also carried out using the K-Means clustering algorithm. In conducting cluster analysis, libraries such as Pandas, NumPy, Matplotlib Pyplot, and Seaborn are used.

In determining the number of clusters used to divide lenders and executing agencies, the elbow method analysis is used. The Elbow method uses the within-custer sum of squares (WCSS) to determine the optimal number of clusters for the analysis.

4.2.2.1 Lender Clusters

Graph 9. Elbow aAnalysis of Lender Data
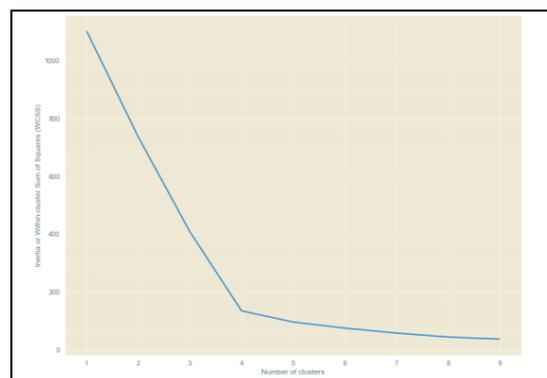
The optimal number of clusters is located at the elbow point on the elbow method visualization diagram (Graph 9). Based on the result of the elbow method, it is decided that cluster analysis will be optimal if using 4 clusters. Furthermore, cluster analysis can be carried out using the K-Means algorithm which produces labels to provide identity for each cluster.

The four clusters that are formed will show the characteristics of lenders based on the size of the loan and the frequency of lending to the Government of the Republic of Indonesia. The amount of the loan granted to the Indonesian government can show the size of the lender's capacity in providing financing to the government. Meanwhile, the number of loans signed by the government with these lenders shows the close relationship between the two entities. The combination of these two characteristics will be reflected by each cluster.
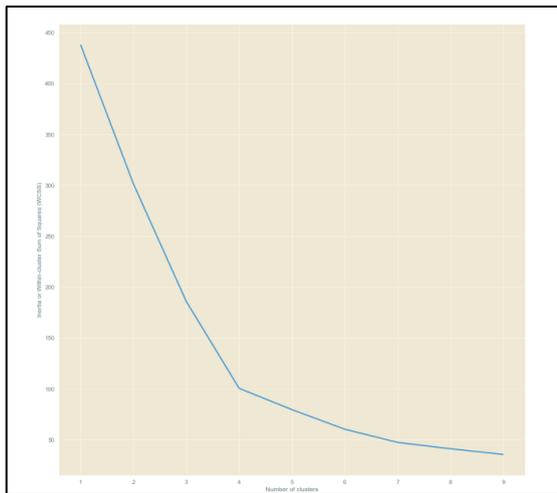
The results of the cluster analysis are presented in Graph 10. In cluster 1, cluster 2, and cluster 3, there will be lenders with a frequency of lending to the Indonesian government under 200 times. Cluster 1 is for lenders with an average loan commitment value of under USD 6 billion. Next, cluster 2 consists of 4 lenders with a low frequency of lending and a low average loan commitment value, including Credit Agricole and the Asian Infrastructure Investment Bank (AIIB).

Meanwhile, cluster 3 is for lenders with a low frequency of loans to the Indonesian government, which is below 90 times, but with a very high average loan commitment value of at least USD 6.4 billion. Domestic state-owned banks. Bank Mandiri, BNI, and BRI are included in this cluster.

Meanwhile, cluster 4 is a lender with a very frequent loan frequency to the Indonesian government, which is above 200 loan commitments. There are 6 lenders in this cluster. Namely IBRD, JICA (Japan), ADB, KfW (Germany), OECF, and BFCE (France).

4.2.2.2 Executing Agency Clusters

Graph 11. Elbow Analysis of Executing Agency

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **19**

The optimal number of clusters for the executing agency is located at the elbow point on the elbow method visualization diagram (Graph 11). Based on the result of the elbow method, it is decided that cluster analysis will be optimal if using 4 clusters. Furthermore, labeling is carried out to provide identity for each cluster.

The four clusters formed will show the characteristics of the executing agency based on the size of the loan received and the frequency with which the executing agency obtains a loan. The amount of the loan received by the executing agency can indicate the capacity of the executing agency to carry out projects financed by government loans. Meanwhile, the historical number of loans that have been the responsibility of the executing agency shows the entity's experience in managing projects financed by government loans. The combination of these two characteristics will be reflected by each cluster.

The results of the cluster analysis are presented in Graph 12. In cluster 1, cluster 2, and cluster 3, the executing agency will be found with the frequency of obtaining Indonesian government loans below 300 times. Cluster 1 is for executing agencies with a low frequency of obtaining government loans, under 300 times, but with a very high average loan commitment value of at least USD 45 billion. The Ministry of Finance, the National Police, the Ministry of Health, as well as several SOEs such as PT PGN and PT Pertamina are included in this cluster.

Meanwhile, cluster 2 is for executing agencies that have received less than 10 times government loans with an average loan commitment value of less than USD 700 million. Next, cluster 3 consists of executing agencies with a low frequency of obtaining government loans and a low average value of loan commitments, namely BPPT, BP Batam, and the National Library.

Meanwhile, cluster 4 is the executing agency with a very frequent frequency of obtaining Indonesian government loans, which is above 400 loan commitments. There are 4 executing agencies in this cluster. Namely the Ministry of Civil Works and

Public Housing, PT PLN (Persero), the Ministry of Transportation, and the Ministry of Defense.

4.3 Data Modelling

The next step is data modeling with stages as presented in Figure 4. The working papers represented in this section can be obtained in this link: https://s.id/loan-prediction-working-papers.
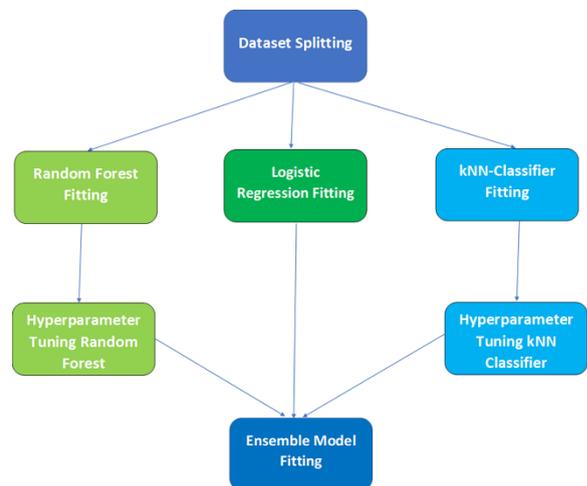

Figure 4. Stages of Data Modelling

4.3.1 Dataset Splitting

To develop the model, the dataset must be divided into two, so that the resulting predictive model can be evaluated. This method is known as the holdout method, where the dataset is divided into training datasets and testing datasets (Geron: 2019, Raschka & Mirjalili: 2017). The training dataset is the dataset that will be processed by the algorithm to form a predictive model. Meanwhile, dataset testing will be used to evaluate the accuracy of the predictive model formed using dataset testing (Albon: 2018). The training and testing datasets must be mutually exclusive so that the resulting predictive models are not overfitting.

Splitting the training and testing datasets using the train_test_split library from scikit-learn. In separating the datasets, the test-size and stratify parameters are set. The test-size parameter is used to define the portion size of the dataset that will be used as a testing dataset. In the development of the predictive model this time, the test-size parameter is set to 0.3, meaning that 30% of the dataset becomes the testing dataset, and 70% becomes the training dataset.

4.3.2 Individual Algorithm Fitting

In developing the model, the ensemble learning methodology was applied. Ensemble learning uses several machine learning models to make better predictions over the dataset. However, before developing the ensemble model, an individual model

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **20**

was first built using the Random Forest algorithm, kNN Classifier, and Logistic Regression.

4.3.2.1 Random Forest Fitting

Random forest is a learning ensemble with bootstrap aggregation (bagging) type. The use of decision tree-based algorithms such as random forest, instead of using linear regression algorithms, is based on the consideration that this predictive model has categorical features.

Bootstrap means ensemble learning will be done by taking random sampling with replacement. Through this method, the training set that has been generated through the splitting process will be re-selected by sampling to produce several new datasets, which in turn will allow for several observations.

Bagging is a general procedure to reduce variance in algorithms that have high variance such as decision trees (Raschka & Mirjalili: 2017). Bagging makes each generated model run independently and then aggregates the results without preference for each model (Geron: 2019).

The Random Forest Regression algorithm will produce several decision tree models (Shwartz & David: 2014, Geron: 2019, Raschka & Mirjalili: 2017, Muller & Guido: 2017)). There is no interaction between each of the decision tree models. The need for a random forest algorithm is because the decision tree algorithm is very sensitive to the training set data used in forming the model.

Consequently, if the training set is changed, the resulting decision tree will also be different so that it will produce different predictions. This is referred to as overfitting risk. Random forest regression is used as an effort to overcome the overfitting risk (Shwartz & David: 2014, Muller & Guido: 2017).
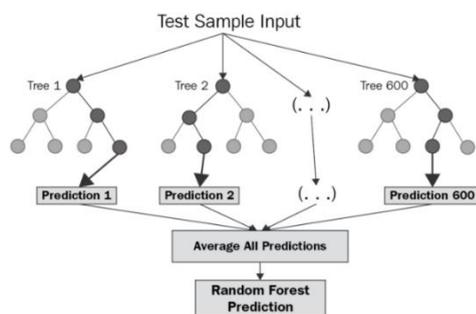

Figure 5. *Random Forest Classification Concept* (sumber: *Towards Data Science*)

The random forest model is a meta-estimator, meaning that it combines the results of several predictions, which in this case are generated by several decision trees. In the regression concept using a random forest, the final prediction generated from the model is the average of the individual predictions of each decision tree.

4.3.2.2 kNN-Classifier Fitting

k-Nearest Neighbors (kNN) is a non-parametric algorithm that can be used for both classification and regression. This algorithm works by taking one data point and looks at the nearest 'k' data points. The data points are then given a label (assigned) based on the majority labels of the closest 'k' data points (Zhang, 2016).
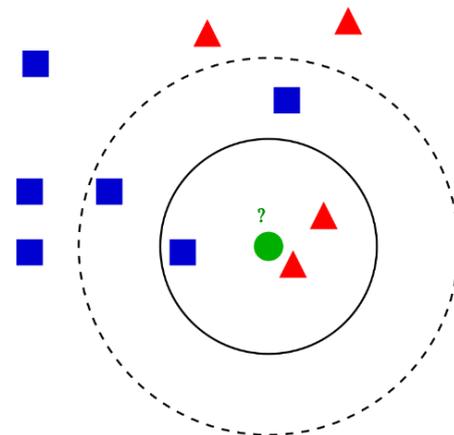

Figure 6. kNN Classifier Concept Illustration (credit: Antti Ajanki)

kNN classifier could be described from Figure 6. For example, take the green dot at the epicenter as an unclassified data point. Thus, we would like to classify the green dot as red triangle or blue rectangle based on the nearest data points classification. If 3 data points is selected as the 'k' in implementing the algorithm, thus the green dot would be classified as red triangle, as the majority classification of 3-nearest data points are classified as red triangle.

It would be a completely different story when 5 data points is selected to be the number of nearest neighbors (thus k=5). Under this circumstance, the green dot would be classified as blue rectangle. For as can be comprehended from the illustration, the majority classification of 5-nearest data points is classified as blue rectangle.

And it goes on and on based on the number of nearest neighbor parameter we feed into the algorithm. kNN classifier would then decide a certain data point's classification based on the majority classification of k-nearest data points.

4.3.2.3 Logistic Regression Fitting

Logistic regression is one type of regression algorithm. The difference lies within the objective that logistic regression predicted the outcome or target variable in a dichotomic scale. It has similarity from linear regression that logistic regression algorithm may include single or multiple predictors, even though multiple predictors is favored for this may lead to the

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **21**

revelation of each predictor's unique contributions, thereby making it more informative (Stoltzfus, 2011).

When compared to linear regression, logistic regression holds some differences in which logistic regression omits several of the key assumptions that linear and general linear models commonly require (Schreiber-Gregory et.al.,2018). These assumptions are:

- logistic regression does not require a linear relationship between the dependent and independent variables,
- the error terms (residuals) do not need to be normally distributed,
- homoscedasticity is not required, and
- in logistic regression, the dependent or target variable is not measured on an interval or ratio scale.

But even so, implementing logistic regression algorithm still require the data to have little to no multicollinearity among the independent variables (Schreiber-Gregory et.al.,2018). Therefore, pearson-correlation test was then conducted, and the result is there was no multicollinearity among predictors as presented in a heatmap visual on Graph 13.

4.3.3. Hyperparameter Tuning

In developing predictive models, each algorithm has a set of hyperparameters that can be tuned. Parameter tuning is a series of processes to find the most optimal parameters to improve the accuracy of the resulting predictive model.

For tuning, we used the Cross Validation module from scikit-learn. The main way of working of cross validation is to divide the training dataset randomly into several groups as determined. From these groups, one group will be played as a test-set, and the rest will be used as a training-set. The model will then be trained using the training-set and assessed using the test-set. The process is then repeated until each group has been used as a test-set (Geron: 2019, Raschka & Mirjalili, 2017, Muller & Guido: 2017). Therefore, after the dataset training fit was carried out on the Random Forest classifier and knn-classifier algorithms, one additional step was taken to take the model with the best parameter estimator.

In the development of this model, a search cross validation grid is used. Grid search works by repeatedly training the developed model for the specified parameters (Geron: 2019, Albon: 2018). In this way, the model being developed can be tested using each hyperparameter value to then find the optimal value in endeavour to acquire the model with the best estimator.

For the purposes of cross validation, there are several hyperparameters of the random forest algorithm and the kNN classifier that can be set as a grid to be tested. For the random forest algorithm, a hyperparameter is used in the form of the number of decision trees that will be used by the random forest algorithm (n_estimators). Meanwhile, for the kNN classifier, a hyperparameter is used in the form of the number of neighbors being inspected (n_neighbors). The hyperparameter settings are shown in Table 1.

Tabel 1. Random Forest and kNN Classifier Algorithm Hyperparameter Setting under Grid Search Cross Validation

Algorithm	Hyperparameter	Value
kNN-Classifier	n_neighbors	Range 1-25
Random Forest	n_estimators	50, 100, 200

Based on the tuning results using Grid Search Cross Validation, it is known that the parameters that produce the best accuracy are the random forest model with 100 decision trees, and the kNN-classifier model with 24 inspected neighbors (n_neighbors).

4.3.4 Ensemble Model Fitting

After fitting the algorithm and hyperparameter tuning, then the ensemble model is formed using the Voting Classifier algorithm. As already mentioned, the Voting Classifier is an ensemble model that makes the final prediction using the most votes.

The voting classifier algorithm fitting is done by using the scikit-learn library. The trick, the three individual models that have been developed previously are put into an array named 'estimators'. After that, the VotingClassifier() function is initiated, with two inputs. The first is the 'estimator' array variable. The second is setting the voting parameter to 'hard', which represents the voting classifier model that will make decisions based on the majority vote (majority vote).

4.4 Model Evaluation

Model evaluation is done by calculating the accuracy of the model. Accuracy is a metric that represents the fraction of correct model predictions. So accuracy is measured by comparing the number of correct predictions with the total number of predictions. For each ensemble model built using these 3 different algorithms, the accuracy value is checked and presents a confusion matrix for each model.

4.4.1 Predictive Model Evaluation - Disbursement Category Class

The evaluation of the first predictive model with the Disbursement Category class will be carried out in two stages, namely the evaluation of each individual model of the three algorithms, and the evaluation of the final model resulting from the ensemble method.

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **22**

Tabel 2. Model accuracy - numerous disbursement ratio category target

Model	Accuracy
kNN-Classifier	71,70%
Random Forest	69,54%
Logistic Regression	71,94%
Ensemble Model	73,14%

The first evaluation is on the individual model. For the kNN-Classifier model, the accuracy of the built model is 71.70%. The Random Forest Classifier model has an accuracy value of 69.54%. While the Logistic Regression model, the accuracy of the resulting model is 71.94%. The confusion matrix for each individual model with the target disbursement ratio category can be seen in Graph 14.

Next is the evaluation of the final model of the results of the ensemble method. The confusion matrix for the ensemble model with the target disbursement ratio category can be seen in Graph 15. It is known that the model resulting from the voting classifier algorithm has an accuracy of 73.14%. The accuracy of the ensemble model is higher than the accuracy of the three models that make up the individual models.

4.4.2 Predictive Model Evaluation - Numerous Drawing Limit Amendment Class

As with the evaluation of the first predictive model, the evaluation of the second predictive model will also be carried out in two stages, namely the evaluation of each individual model of the three algorithms, as well as the evaluation of the final model resulting from the ensemble method.

Tabel 3. Model accuracy - numerous drawing limt amendment target

Model	Accuracy
kNN-Classifier	86,33%
Random Forest	88,01%
Logistic Regression	86,81%
Ensemble Model	89,69%

The first evaluation is on the individual model. The confusion matrix of individual predictive models with Numerous Drawing Limit Amendments can be seen in Figure 16. For the kNN-Classifier model, the accuracy of the model built is 86.33%. The Random Forest Classifier model has an accuracy value of 88.01%. While the Logistic Regression model, the accuracy of the resulting model is 86.81%.

Next is the evaluation of the final model of the results of the ensemble method. The confusion matrix for the predictive ensemble model with Class Numerous Drawing Limit Amendments can be seen in Graph 17. It is known that the model resulting from the voting classifier algorithm has an accuracy of 89.69%. The accuracy of the ensemble model is higher than the accuracy of the three models that make up the individual models.

There are additional evaluation procedures for models targeting numerous drawing limit amendments. After knowing that the predictive model generated through ensemble learning has the highest accuracy, the next step is to evaluate the model with metrics precision and recall and visualize it using the Receiver Operating Characteristic (ROC) Curve and Precision-Recall Curve (PRC).

ROC Curve is related to the true positive rate (TPR) and false positive rate (FPR). FPR is the ratio of False Positives to the number of True Negatives. In the context of this predictive model, the ratio describes a measure of how many cases the model predicts that a loan has numerous drawing limit amendments than all loans that do not actually have numerous drawing limit amendments. In the data used, the FPR value is 0.18.

On the other hand, TNR or Specificity is the ratio of True Negatives and the actual number of negatives. In this model, the TNR is a measure of how many cases the model correctly predicts that a loan has no numerous drawing limit amendments than all loans that do not actually have numerous drawing limit amendments. TNR for data above 0.82. From these 2 definitions, we can also conclude that Specificity or TNR = 1 – FPR.

After knowing about TNR and TPR, then we can plot the Receiver Operating Characteristic (ROC) Curve. ROC Curve is a plot between TPR and FPR. Since the built predictive model classifies loans as having numerous drawing limit amendments or not based on the probabilities generated for each class, a probability threshold can be decided. For example, if you want to set the threshold value at 0.4, this means that the model will classify loans as having numerous drawing limit amendments if the probability for loans having numerous drawing limit amendments is greater than 0.4. This of course will provide a high recall value and reduce the number of False Positives. Similarly, it can be visualized how this predictive model performs for different threshold values using the ROC curve.

The ROC curve for the ensemble model with numerous drawing limit amendment targets can be seen in Graph 18. The lowest point in the ROC curve, namely the coordinates (0,0) illustrates that the threshold value is set at 1.0, which means the model classifies all loans will have numerous drawing limit amendments. Meanwhile, the top point on the ROC curve, namely coordinates (1,1) illustrates that the threshold is set at 0.0, which means that the model classifying all loans will not have numerous drawing limit amendments. The remainder of the curve is the FPR and TPR values for threshold values between 0 and 1. At some threshold values, it can be observed

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **23**

that for FPR close to 0, the model is made to reach TPR close to 1. This is when the developed model predicts borrowing has numerous drawing limit amendments almost perfectly.

The area with the curve and the FPR axis as the boundary is called the Area Under Curve (AUC). This area is considered a good model metric. With this metric ranging from 0 to 1, predictive model development should target high AUC values. Models with high AUC are referred to as models with good skills. It is known that the area under the ROC Curve is 90.04. This means that the predictive model has the capability to distinguish loans that will undergo numerous drawing limit amendments and those that will not as much as 90.04% of observations.

Next up is the Precision-Recall Curve (PRC). As the name implies, this curve is a direct representation of precision and recall. The development of this predictive model certainly does not prioritize cases that are True Negative or loans that do not have numerous drawing limit amendments. This is very important especially for the preparation of predictive models with unbalanced datasets as in the case of developing this predictive model, where the number of negative datasets (does not have numerous drawing limit amendments) is more than positive ones. In this case, the main concern of model development is to correctly detect loans that have numerous drawing limit amendments.

The PRC curve for the ensemble model with numerous drawing limit amendment targets can be seen in Figure 19. In the PRC curve, the lowest point (0.0) is when the threshold is set at 1.0 or in other words the resulting predictive model cannot distinguish loans with numerous drawing limit amendments and those that are not. Meanwhile, the highest point (1.0) is when the precision and recall of the predictive model produced is very high so that the predictive model can very well distinguish loans that have numerous drawing limit amendments and those that do not. As with ROC, the area under the PRC curve shows how good the model is and seeks to maximize the AUC value. Because the high AUC for the PRC curve indicates that the predictive model produces high precision predictions and can detect the majority (majority) of the datasets that have a positive target or in this context will have numerous drawing limit amendments. It is known that the AUC of the resulting PRC curve is 80.54.

4.5 Model Deployment

To be able to deploy, it is necessary to first store the predictive model that has been developed and meets the expected evaluation targets. To be able to store predictive models, you can use a library called pickle.

After that, deployment of the predictive model that has been created is carried out through dashboard development. The dashboard was developed using the dash library from plotly under the name Republic of Indonesia Sovereign Debt Performance Predictive Analytics Dashboard.

The dashboard allows users to predict the performance of loan disbursement by the executing agency, by setting features related to the characteristics of the lender, the characteristics of the executing agency, and the characteristics of the loan. After setting these features, the dashboard will generate two kinds of predictions for the performance of the loan. The first is about the rate of disbursement of the loan and the second is about the possibility of an amendment to the loan disbursement deadline.

The source code of the dashboard developed can be obtained in the following link: https://s.id/loan-prediction-working-papers.

5. CONCLUSION

Of the two predictive models developed, the multiclass classification model on the disbursement rate of government loans has an accuracy rate of 73.14%. This means that the predictive model developed can accurately predict 73.14% of observations on the level of government loan disbursement. On the other hand, for the predictive model of the drawing limit amendment, the accuracy is quite high at 89.69%. This means that this predictive model is able to accurately predict government loan amendments as much as 89.69% of observations.

Although the model was developed with an unbalanced composition of target features, the next test through Precision and Recall Curve showed that the area under the curve (AUC) was 80.54%. This high AUC value represents that the resulting model is not only capable of producing accurate predictions (precision), but can also detect the majority (80.54%) of loans that will undergo repeated amendments (numerous drawing limit amendment).

6. IMPLICATION AND LIMITATION

The strategic contribution of the predictive model developed lies in its ability to predict the performance of loan plans that have not been executed or signed. With insight in the form of predictions on the loan plan, the Minister of Finance can give special emphasis to the Ministries/Institutions/SOEs/Local Governments who are the executing agencies for the loan plan, for example at the trilateral meeting in the preparatory stage of loan procurement.

The predictive model developed in this study has limitations. The first is related to the limitations of the data obtained to be used as predictors (features). The data source used only comes from the internal

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **24**

Ministry of Finance which is based on records in the DMFAS application belonging to the DJPPR. This predictive model built will certainly be more perfect if it also uses relevant external data.

In addition, the criteria for determining loan performance proxies as predictive targets are formulated based on the team's consideration and experience in assessing risk, not referring to a specific official document.

And lastly, this predictive model is prepared with unbalanced datasets, for example, the number of negative datasets (does not have numerous drawing limit amendments) is much higher than positive ones.

ACKNOWLEDGEMENT

REFERENCES

Albon, Chris. 2018. *Machine Learning with Python Cookbook*. O'Reilly Media Inc.
   In-text reference: (Albon, 2018)
Geron, Aurelien. 2019. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media Inc.
   In-text reference: (Geron, 2019)
Muller, A.C., Guido S. 2017. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media Inc.
   In-text reference: (Muller & Guido, 2017)
Raschka, S., Mirjalili, V. 2017. *Python Machine Learning 2nd Edition: Manchine Learning and Deep Learning with Python, scikit-learn, and Tensor Flow*. Packt Publishing.
   In-text reference: (Raschka & Mirjalili, 2017)
Schreiber-Gregory, D. et.al. 2018. *Logistic and Linear Regression Assumptions: Violation Recognition and Control*. SESUG Paper 247 -2018
   In-text reference: (Schreiber-Gregory et.al., 2018)
Shwartz, S.S, David, S.B. 2014. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.
   In-text reference: (Shwartz & David, 2014)
Stoltzfus, J. C. 2011. *Logistic Regression: A Brief Primer*. Academic Emergency Medicine, 18(10), 1099–1104. doi:10.1111/j.1553-2712.2011.01185.x
   In-text reference: (Stoltzfus, 2011)
Zhang, Z. 2016. *Introduction to machine learning: k-nearest neighbors*. Annals of Translational Medicine Vol 4, No 11 (June, 2016). doi: 10.21037/atm.2016.03.37
   In-text reference: (Zhang, 2016)

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30
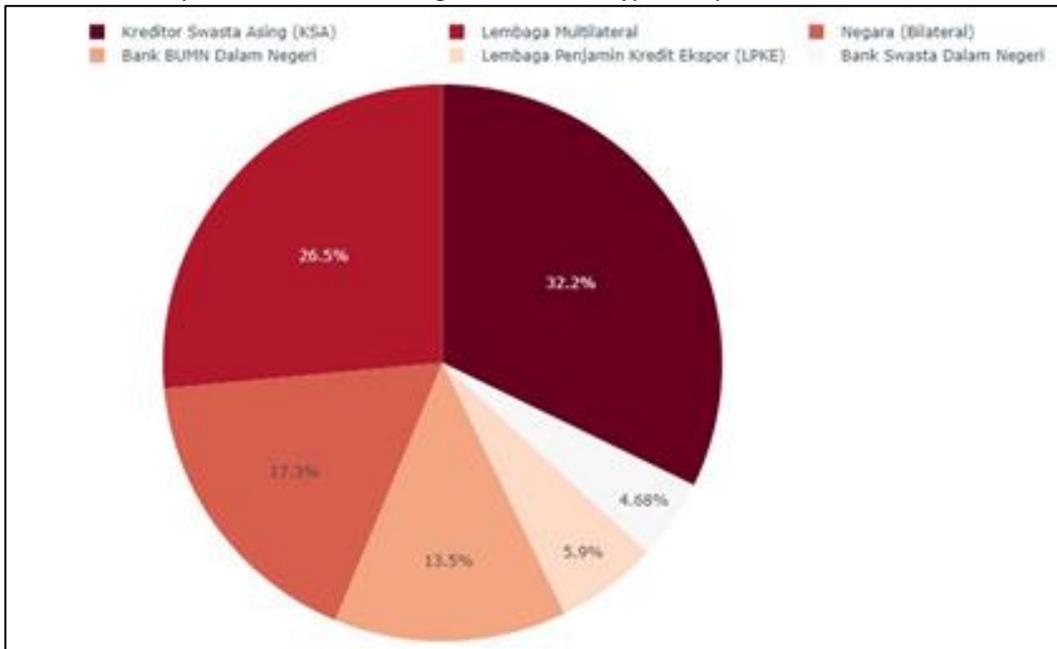
Halaman **25**

GRAPH ILLUSTRATIONS

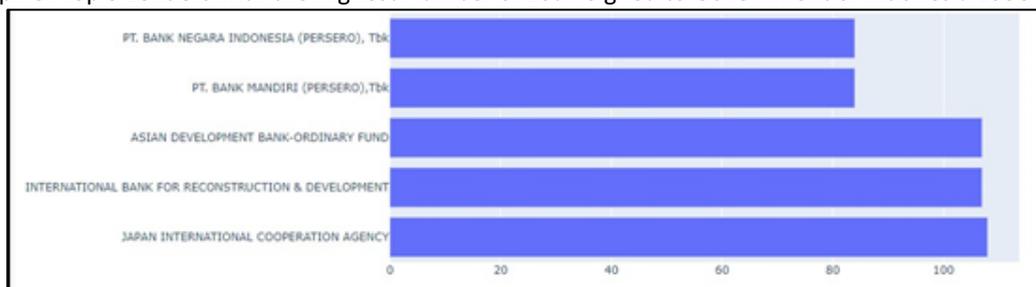Graph 1. 5 Executing Agency with Top Amounts of Total Loan 1998-2020
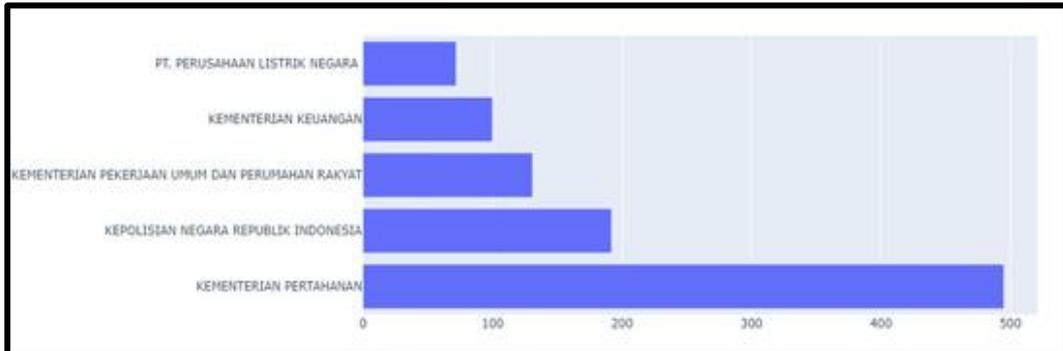
Source: DJPPR DMFAS.

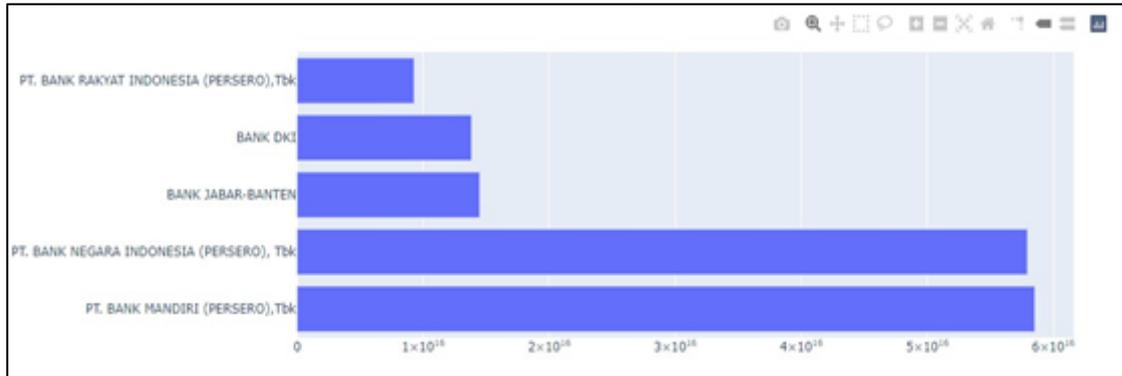Graph 2. Indonesia Sovereign Loan Lender Type Composition 1998-2020

Source: DJPPR DMFAS.

Graph 3. Top 5 Lenders with the Highest Number of Loan Signed to Government of Indonesia 1998-2020

Source: DJPPR DMFAS.

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **26**

Graph 4. Top 5 Lenders with the Most Amount of Loan Signed to Government of Indonesia 1998-2020

Source: DJPPR DMFAS.

Graph 6. Top 5 Executing Agency with the highest Number of Loans having Numerous Drawing Limit Amendment
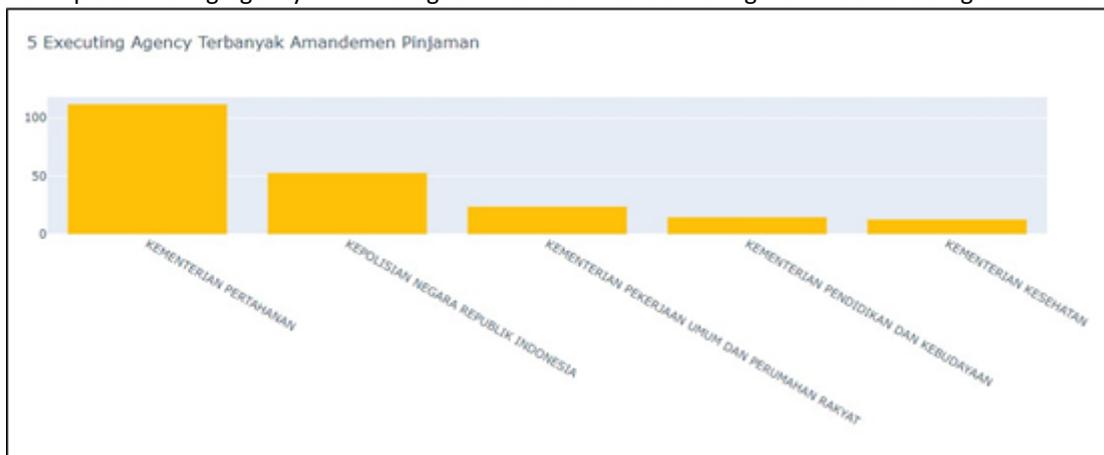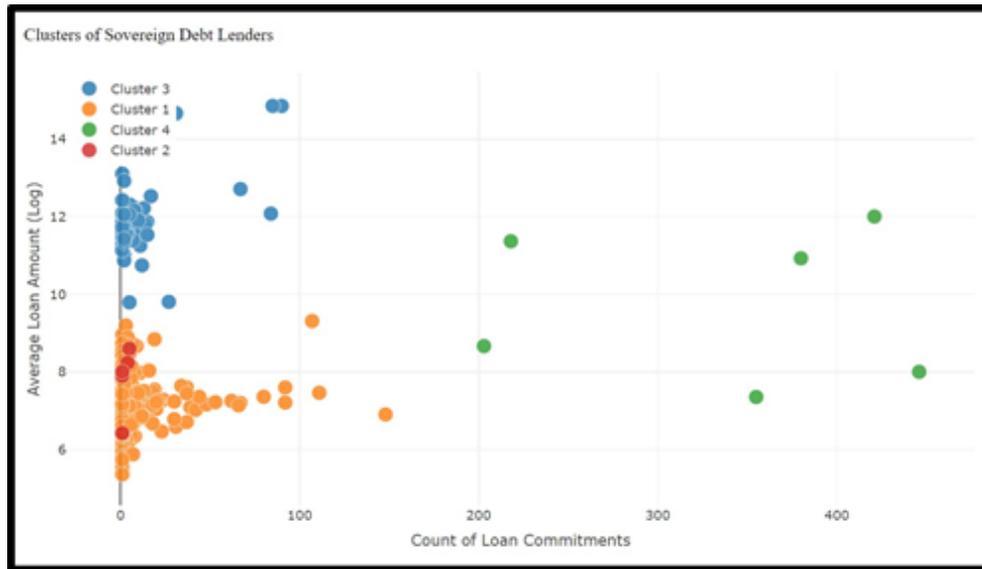

Source: DJPPR DMFAS

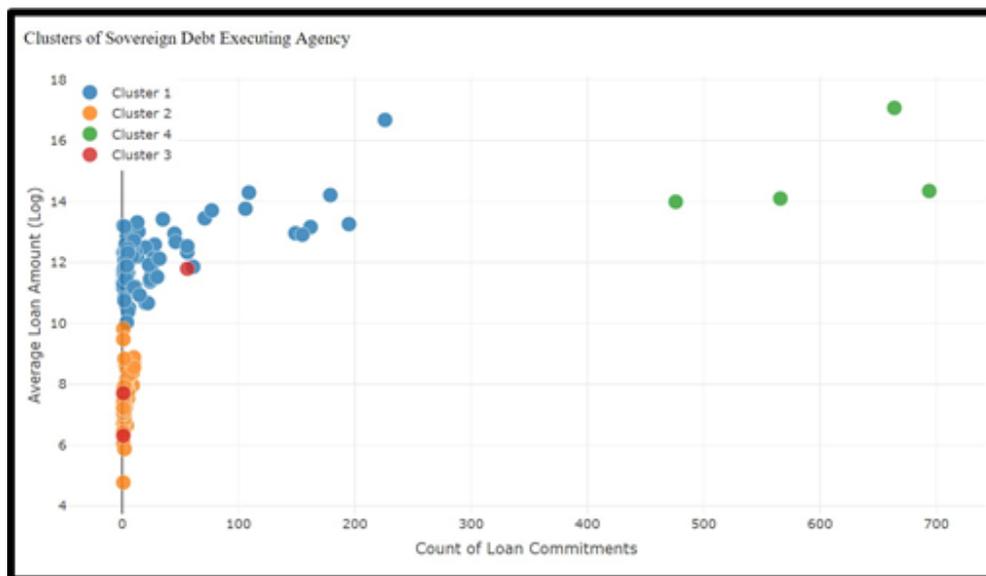### Graph 8. Top 5 Executing Agency with the highest Number of Loans Having Low Disbursement Ratio
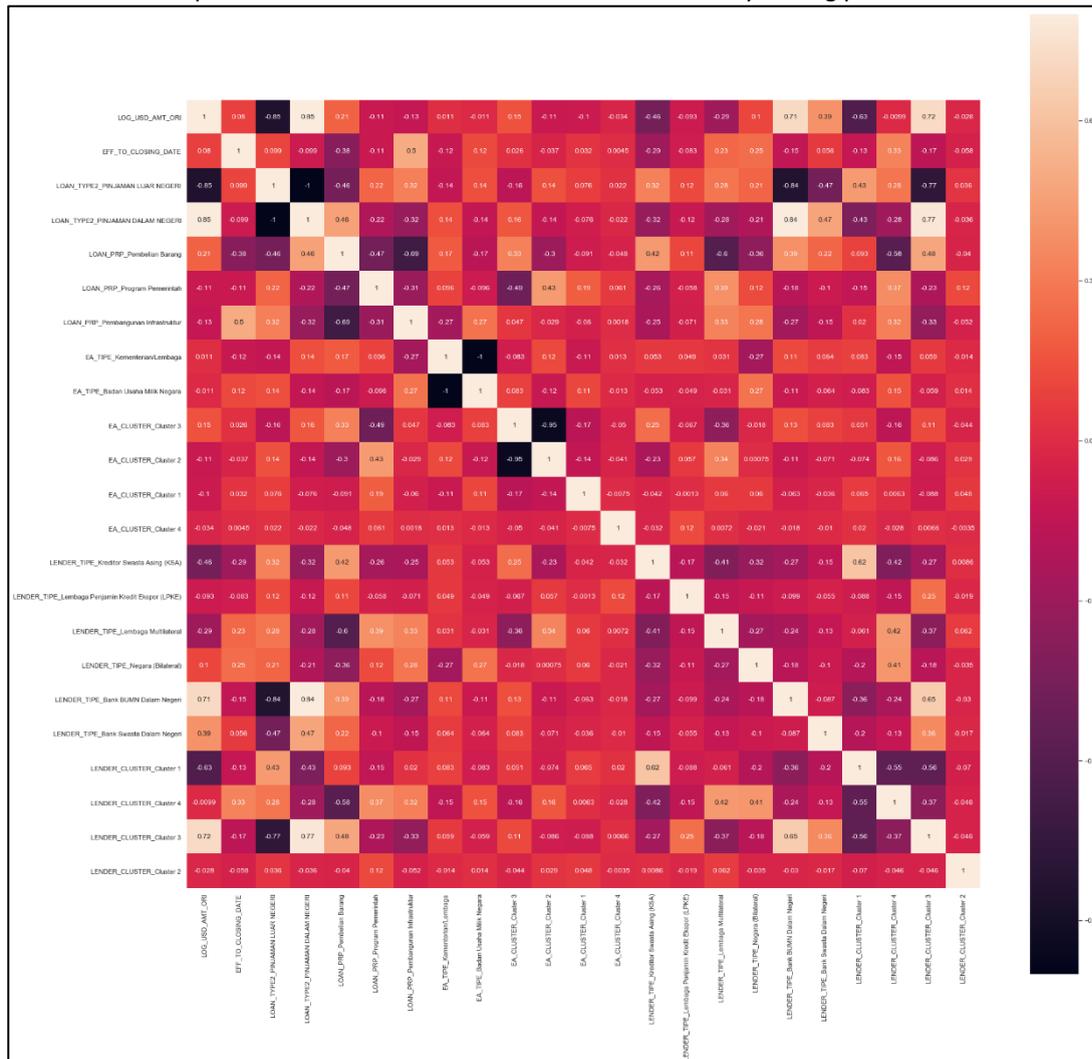

Source: DJPPR DMFAS

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **27**
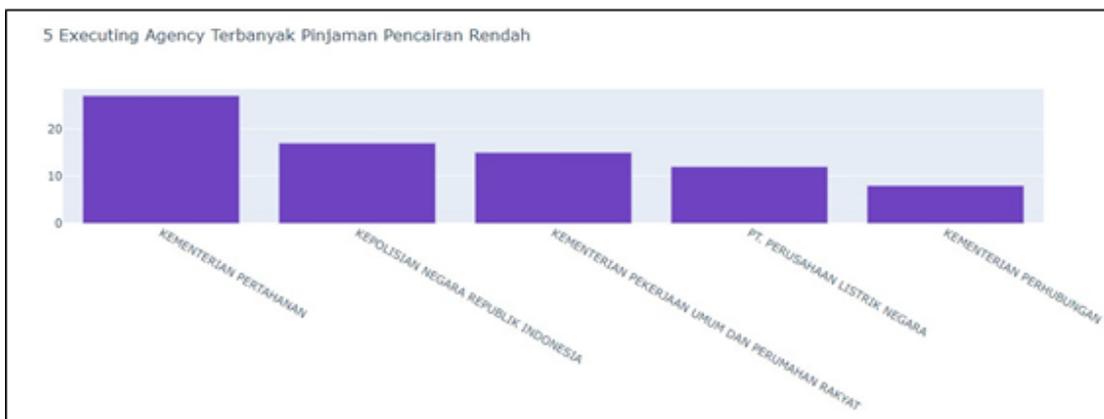
Graph 10. Government of Indonesia Lender Cluster 1998-2020

Graph 12. Government of Indonesia Loan Executing Agency Cluster 1998-2020

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

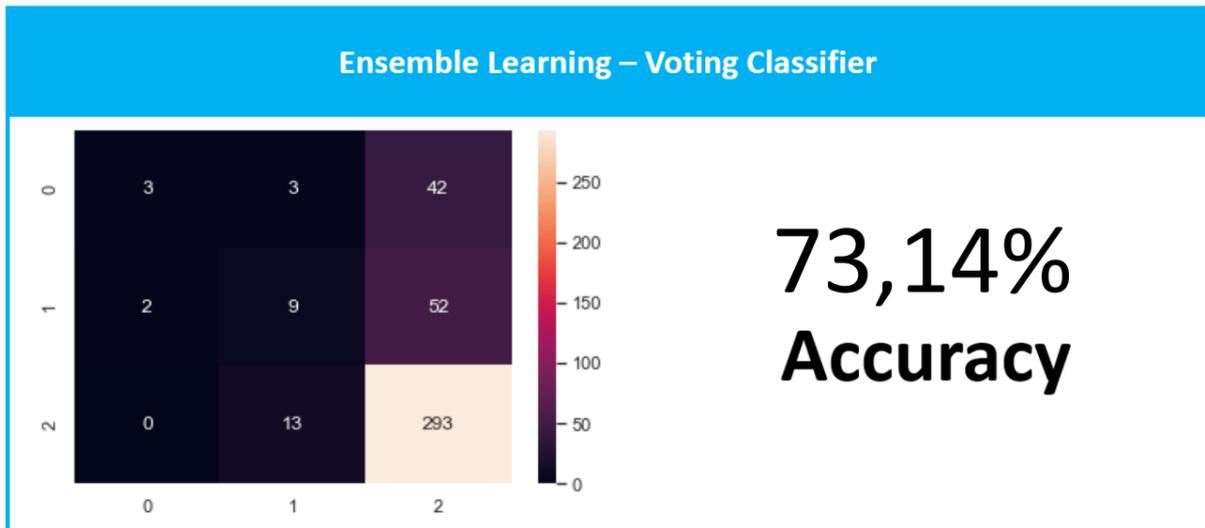Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **28**

Graph 13. Pearson-Correlation test for multicollinearity among predictors

Graph 14. Confusion matrix for individual model with disbursement ratio category as target

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

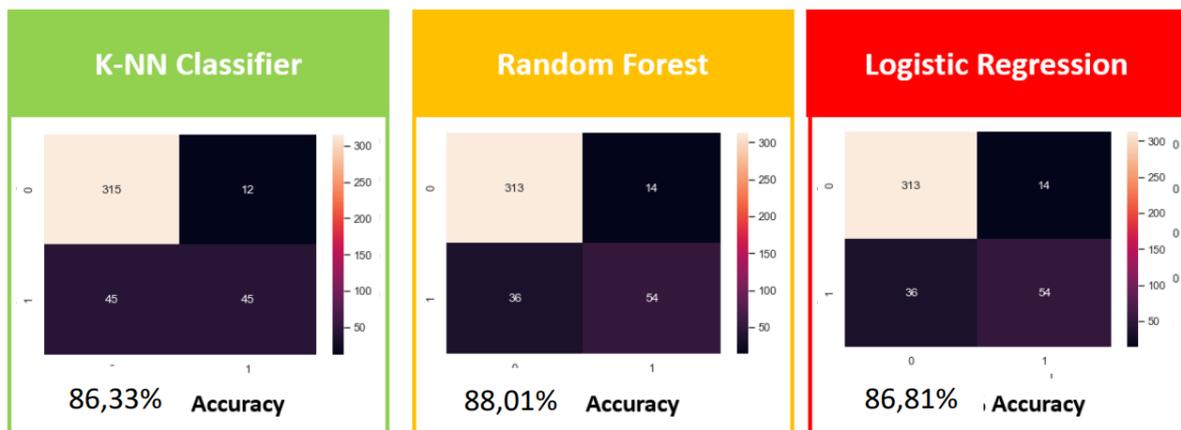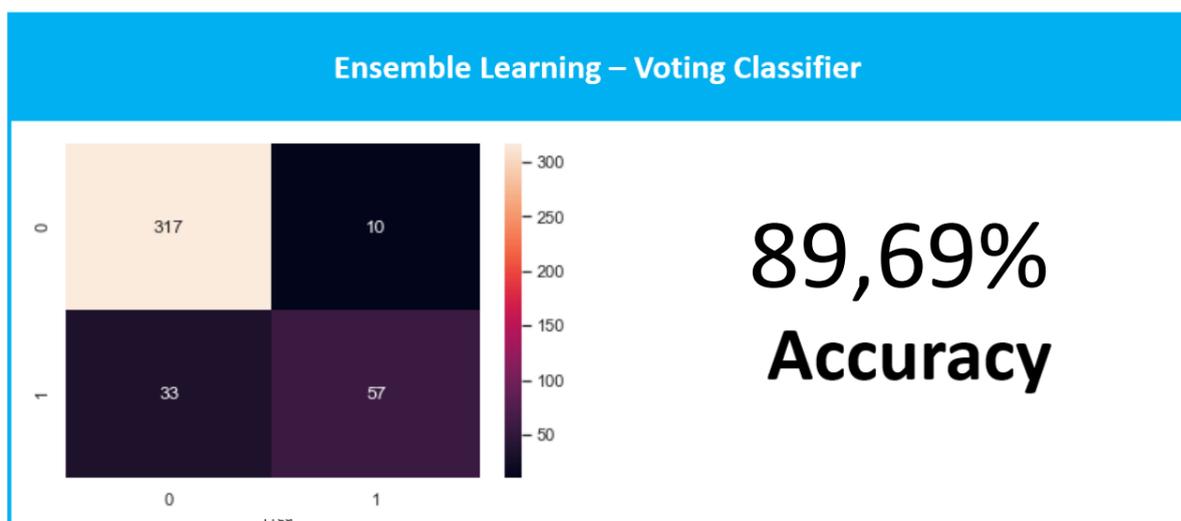Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **29**
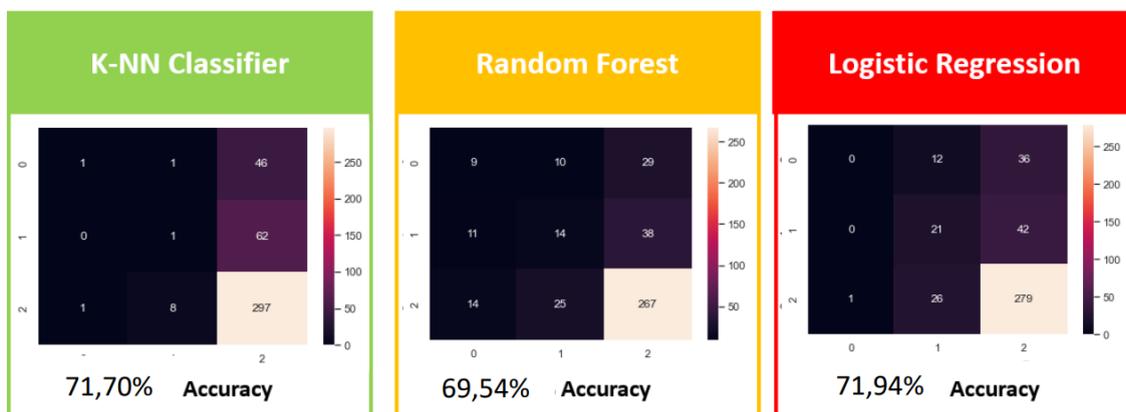
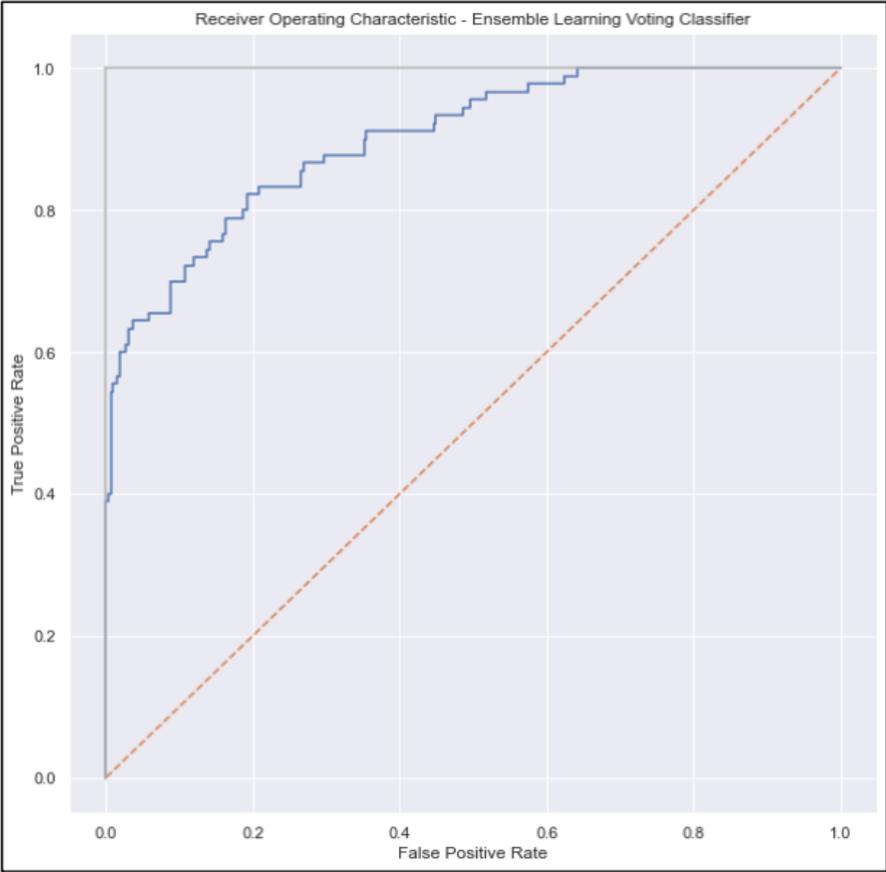Graph 15. Confusion matrix for ensemble model with disbursement ratio category as target

Graph 16. Confusion matrix for individual model with numerous drawing limit amendment as target.

Graph 17. Confusion matrix for ensemble model with numerous drawing limit amendment as target.

PREDICTING INDONESIA SOVEREIGN DEBT MANAGEMENT'S
RISK OCCURRENCE USING MACHINE LEARNING ALGORITHMS
Darmawan, Putra, Fridayanto

Jurnal Info Artha Vol.6, No.1, (2022), Hal.13-30

Halaman **30**

Graph 18. ROC Curve for ensemble model with numerous drawing limit amendment as target.

Graph 19. PRC Curve for ensemble model with numerous drawing limit amendment as target