Jurnal Teknik Komputer AMIK BSI
Published: LPPM Universitas Bina Sarana Informatika
website: http://ejournal.bsi.ac.id/ejurnal/index.php/jtk
Vol. 11, No. 1, Januari 2025, P-ISSN 2442-2436, E-ISSN: 2550-0120, Page. 70-79.

Application of Random Forest Algorithm To Classify Credit Status of KPR Customers at Bank BTN Based on Machine Learning

Maysade Fitri[1], Ahmad Sobri[2], Fido Rizki[3]

[1,2,3] Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Bina Insan, Lubuklinggau, Indonesia
Email : 2102020014@mhs.univbinainsan.ac.id, ahmadsobri506@gmail.com, fidorizki@univbinainsan.ac.id

ARTICLE INFORMATION

Artikel History::
Received: 8/1/, 2025
Revised: 15/1/, 2025
Accepted: 30/1/ , 2025

Keyword:
EDA
Random Forest
Credit Classification
KPR
Machine Learning

ABSTRACT

In the banking sector, this study is very suitable for determining and improving accuracy and determining credit status classification. This study aims to apply the Exploratory Data Analysis (EDA) method in supporting credit status classification at PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk. Exploratory Data Analysis (EDA) as data exploration and Machine Learning Algorithms such as Random Forest as modeling in determining classification. The results show that the Exploratory Data Analysis (EDA) method successfully determines data patterns, while Random Forest in modeling achieves accuracy, recall, Precision, F1-Score of 100% in predicting the credit status of KPR customers. This method is expected to be useful in making decisions on more accurate credit status by the bank.

Corresponding Author:

Maysade Fitri,
Fakultas Ilmu Komputer,
Universitas Bina Insan,
Jl. HM Soeharto No.Kel, Lubuk Kupang, Kec. Lubuk Linggau Sel. I, Kota Lubuklinggau, Sumatera Selatan 31626,
Email: 2102020014@mhs.univbinainsan.ac.id

INTRODUCTION

Banks are business entities that collect funds from the public in the form of savings and distribute them to the public in the form of credit and/or other forms in order to improve the standard of living of many people (Wulansari & Purwitasari, 2023). As the main provider of mortgage services, PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk offers a number of financial products and services, including Home Ownership Credit (Wikipedia, n.d.).

Credit is the provision of a loan to another party which requires the borrower to repay it along with interest over a certain period of time in accordance with a previously agreed agreement (Sidik, 2019). To maximize the potential of banks and ensure business success, it is essential to understand the factors that affect credit customer performance. In this context, machine learning and Exploratory Data Analysis (EDA) methods are increasingly needed to analyze credit data more comprehensively and accurately.

The EDA method can provide more useful information in understanding the patterns and characteristics of customer credit data. For example, research on Machine Learning for Enhanced Credit Risk Assessment shows that EDA is very helpful in identifying payment patterns that are difficult to see clearly and is very helpful in developing models that predict credit smoothness based on large amounts of data. This is very important for banks to reduce the risk of stock market fluctuations and ensure their financial stability (Risiko et al., 2020).

In the context of machine learning, algorithms such as Random Forest have proven to be more effective than traditional methods. Random Forest, also known as random decision forests, is a specific algorithm that uses a combination of decision trees based on a subset of a dataset (Andrew Murphy, 2019). The use of Exploratory Data Analysis (EDA) in customer credit analysis allows PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk to identify several features that have a major impact on credit evaluation, with relevant variables, such as several collectabilities, namely current, special attention, less current, doubtful and bad, the bank can improve risk prediction. Machine learning algorithms are then used to categorize data according to credit risk probability. It is expected that this random forest algorithm will provide faster, more accurate results, and work continuously without human assistance than traditional methods, increasing the operational efficiency of the bank (Ningsih et al., 2022).

So that the results of accuracy, recall, precision, and F1-score are 100% each. Then the results of the credit status of KPR customers who were declared smooth amounted to 25 customers and were declared non-smooth amounted to 63 customers from 292 datasets that had been preprocessed, namely splitting data.

This study aims to classify the form of mortgage customer data using the EDA method and the use of random forest as a machine learning algorithm to determine the smoothness of credit status, namely smooth and non-smooth customer credit, with model evaluation using metrics such as accuracy, recall, precision, F1-score and confusion matrix.

The results of this study are expected to provide insight into credit risk which shows that the Exploratory Data Analysis (EDA) method and the Random Forest Machine Learning algorithm are very important in the analysis. So that the credit status of mortgage customers who are declared non-performing is higher than the credit status declared current based on the customer's credit arrears variable. This illustrates that there are significant features that influence credit status, which are useful in making credit risk decisions. Regarding the pattern and distribution of mortgage customer credit data in an intuitive understanding for distribution analysis and visualization of prediction results.

RESEARCH METHOD

1. Research Stages

Figure 1 explains the research stages that will be used, starting from data collection to results and implementation.

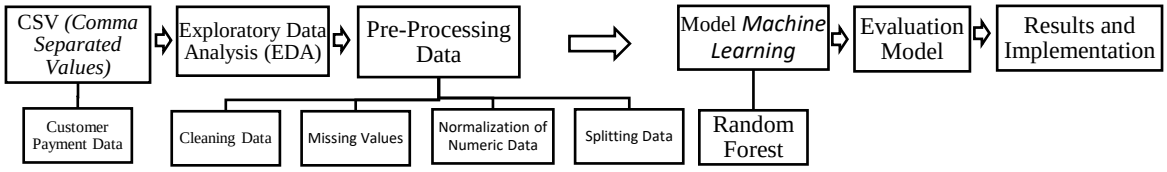

Figure 1. Research Stages

Figure 1 shows that the data used in this study is KPR customer payment data at PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk, with CSV (Comma Separated Values) data format. The data received will be processed using the Exploratory Data Analysis (EDA) method, at this stage data simplification is carried out through pre-processing with Cleaning Data, miising values, numeric data normalization, and data splitting.

The simplified data will be processed again in a machine learning model such as random forest to determine accuracy, recall, precision, F1-score and confusion matrix. Furthermore, the model evaluation stage is to assess how strong the model is to classify credit status. Each metric provides different model strengths and weaknesses. Cross-Validation as a validation process applies k-fold cross-validation to measure model ideas and stability in varying datasets.

The results and implementation of this study are expected to explain important patterns through a random forest algorithm based on machine learning on customer credit data, the resulting model will then be evaluated using accuracy, recall, precision, F1-score and confusion matrix. Implementation is useful for integrating the best models such as random forests in bank credit evaluations, can classify customer credit status smoothly or not smoothly.

This study has 292 KPR payment customer data that has been taken directly by researchers at PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk. This study uses two types of data, namely

primary and secondary. Primary data is data obtained through the results of an interview process with one of the bank staff who is in charge of managing customer credit data, while secondary data is a dataset that has been collected based on bank reports in the form of consumer payment data in the form of excel format which will be processed in CSV format to classify the Random Forest algorithm based on Machine Learning.

2. Random Forest

Random forest is an ensemble method used to solve classification problems (Budianti & Suliadi, 2022). Several studies have shown that the random forest method has a higher accuracy value than other methods. In addition, in large data sets, the random forest method is very suitable for use, because the data set formed on each tree is not the same data set, but rather the result of bootstrapping. However, if the resulting decision tree is increasing, then the creation of the model takes a long time. The random forest method is carried out by growing many trees so that a forest is formed, then the analysis is carried out on the collection of trees (Ulandari et al., 2024).

Random forest classification error estimation, research can be obtained in the following way (Liaw & Wiener, 2018) :

a. Make predictions for data not in the bootstrap sample (which Breiman calls "Out-Of-Bag," or OOB) using a tree created with the bootstrap sample.
b. The second step, combine the OOB predictions. On average, each data point will be out of the bag about 36% of the time.
c. The third step, calculate the error rate, and call it the OOB error estimate.

(Liaw & Wiener, 2018) The OOB error estimate is quite accurate, it is recommended that enough trees are planted so that the OOB estimate is not biased upwards.

3. Exploratory Data Analysis (EDA)

1) Application of Exploratory Data Analysis (EDA) and Machine Learning.

Exploratory Data Analysis (EDA) is an important step to understand the data thoroughly before applying machine learning models. EDA is essential to find patterns and gain insights from data. It often involves summarizing key data characteristics through visualization. EDA is an important step to understand the data thoroughly before applying Random Forest. While the Random Forest Algorithm can improve financial analysis by predicting banking financial metrics and trends. The Random Forest model is usually used for predictive analysis (Santosa et al., 2024).

EDA has a major benefit, one of which is its role in increasing the effectiveness of predictive models. By understanding the entire data, data scientists can make informed decisions about feature selection, transformation, and engineering that lead to more robust and accurate predictive models. EDA helps identify relevant variables, uncover potential confounding factors, and understand their interactions (*Exploratory Data Analysis (EDA) for Data Science and ML*, n.d.). Here are some graphical techniques that are often used:

1. Plotting raw data such as data traces, histograms, bihistograms, probability plots, log plots, and Youden plots.
2. Plotting simple statistics such as mean plots, standard deviation plots, box plots.

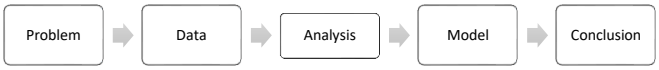

Figure 2. Exploratory Data Analysis (EDA) (Aryanti & Setiawan, 2019)

2) Review of Key Variables in Credit Data

There are three variables that have been selected by the author according to the characteristics and assessment of customer credit status, namely :

a. Collectibility

The collectability variable has five elements, namely col-1 which is interpreted as smooth, col-2 which is interpreted as under special attention with delays ranging from 31-90 days, col-3 which is less smooth with delays in payment of around 91-120 days, col-4 which is doubtful with delays of 121-180 days, and col-5 which is bad with delays <180 days.

b. Arrears

Arrears are the nominal installments that have not been paid by consumers to the bank.

c. Due Date

The due date is determined or calculated since the implementation of the credit agreement and the agreement on payment procedures between the consumer and the bank. The credit payment period for KPR is around 15 years and the due date has been calculated if the payment is late.

RESULTS AND DISCUSSION

1) Result

The dataset produced by the author through the data collection process amounted to 292 KPR customer data, with several variables including collectability, arrears, and due date. These variables will determine the credit status of customers who are declared smooth or not smooth KPR customer credit.

Statistik Deskriptif Dataset:

	BRANCH	BRANCH2	KANWIL	ACCTNO	STATUS	ORGAMT
count	292.000000	292.0	292.0	2.920000e+02	292.0	2.920000e+02
mean	44.054795	8.0	5.0	4.406505e+12	1.0	1.566107e+08
std	82.459761	0.0	0.0	8.245979e+12	0.0	2.142155e+07
min	8.000000	8.0	5.0	8.010214e+11	1.0	2.000000e+07
25%	8.000000	8.0	5.0	8.010214e+11	1.0	1.539000e+08
50%	8.000000	8.0	5.0	8.010214e+11	1.0	1.563800e+08
75%	8.000000	8.0	5.0	8.010214e+11	1.0	1.603400e+08
max	232.000000	8.0	5.0	2.320117e+13	1.0	3.000000e+08

	AMTREL	DLM_RIBU	BLN_AKAD	ORGDT6	...	\
count	2.920000e+02	292.000000	292.000000	292.000000	...	
mean	1.566107e+08	156610.712055	557.886986	190866.106164	...	
std	2.142155e+07	21421.550127	317.895211	83750.904766	...	
min	2.000000e+07	20000.000000	124.000000	11223.000000	...	
25%	1.539000e+08	153900.000000	224.000000	130624.000000	...	
50%	1.563800e+08	156380.000000	524.000000	210224.000000	...	
75%	1.603400e+08	160340.000000	824.000000	260349.000000	...	
max	3.000000e+08	300000.000000	1223.000000	310724.000000	...	

	BIKOLE	BISIFA	PREBAL	TERM	PDDAYS	\
count	291.000000	291.0	2.920000e+02	292.000000	77.000000	
mean	1.285223	9.0	1.566107e+08	186.746575	7.922078	
std	0.452299	0.0	2.142155e+07	38.240029	0.683763	
min	1.000000	9.0	2.000000e+07	60.000000	2.000000	
25%	1.000000	9.0	1.539000e+08	180.000000	8.000000	
50%	1.000000	9.0	1.563800e+08	180.000000	8.000000	
75%	2.000000	9.0	1.603400e+08	211.500000	8.000000	
max	2.000000	9.0	3.000000e+08	240.000000	8.000000	

	kolektabilitas	tunggakan	MEMAMT	REK_TAB	NO HP
count	292.000000	2.920000e+02	0.0	2.920000e+02	2.920000e+02
mean	3.092466	4.361700e+07	NaN	2.160314e+13	1.521461e+11
std	1.179002	5.220808e+07	NaN	6.462024e+12	4.096261e+11
min	1.000000	0.000000e+00	NaN	0.000000e+00	8.117872e+08
25%	2.000000	1.380001e+06	NaN	2.320150e+13	8.137890e+10
50%	3.000000	2.371350e+07	NaN	2.320150e+13	8.227978e+10
75%	4.000000	7.023012e+07	NaN	2.320150e+13	8.527995e+10
max	5.000000	4.210832e+08	NaN	6.010150e+13	6.281379e+12

[8 rows x 28 columns]

Figure 1. Statistik Deskriptif

Figure 1. Statistik Deskriptif shows descriptive statistics aimed at understanding the scale of the data, identifying outliers, and the distribution of values as a basis for statistical summary.
Continued with **Figure 2**. Data Exploration (EDA Visualization) This process aims to show the relationship between variables, such as collectability and customer maturity date.

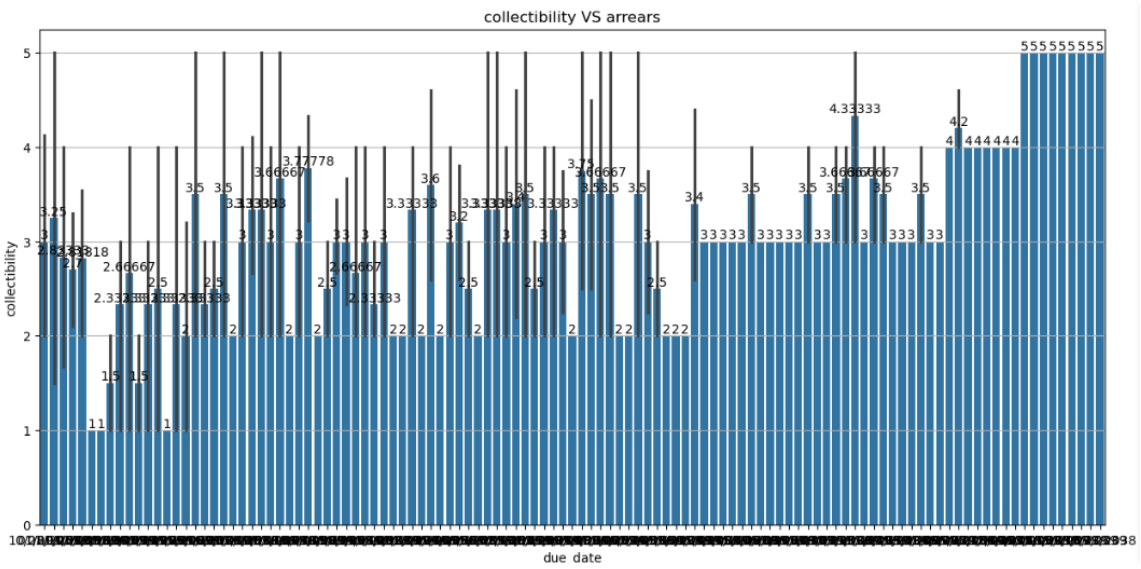

Figure 2. Data Exploration (EDA Visualization)

Then enter the process of visualizing the distribution of collectability and arrears data.

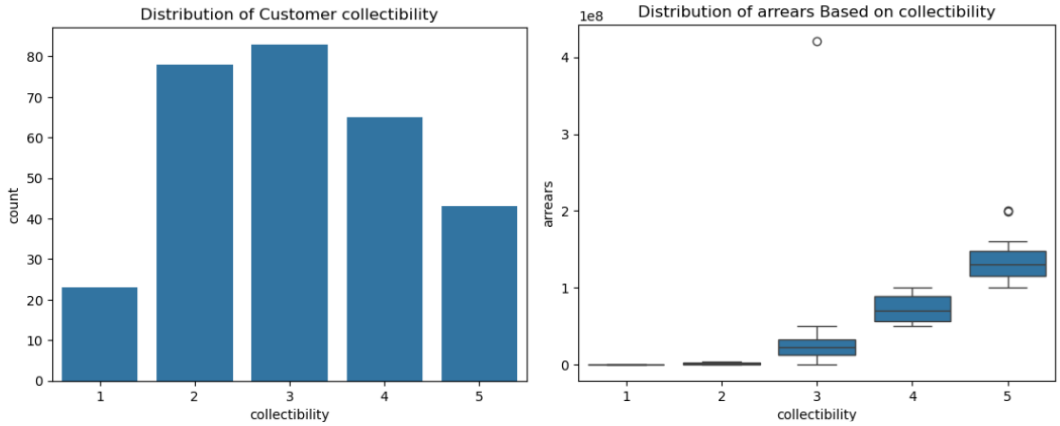

Figure 3. Distribution Collectibility and Arrears

The distribution of customer collectability displays the distribution of collectability categories, while the distribution of arrears based on collectability shows that the median value of arrears is according to the collectability category. So the following display makes it easier to understand in interpreting an analysis.

Next, pre-processing is data processing before entering the modeling process, namely missing values, data cleaning, numeric data normalization, and data splitting. In the missing values process, the results show that the three variables used have 0 missing values.

```
Column before dropna: Index(['BRANCH', 'BRANCH2', 'STA_CODE', 'CABANG', 'KANWIL', 'ACCTNO', 'CIFNO',
       'SNAME', 'STATUS', 'TYPE', 'DESC', 'KATEGORI', 'PRODUCT', 'KEL_PRODUCT',
       'ORGAMT', 'AMTREL', 'DLM_RIBU', 'BLN_AKAD', 'ORGDT6', 'RATE', 'AFRATE',
       'DEVPID', 'CCDVNM', 'LPDVCD', 'CCDVCD', 'LPDLOC', 'LPPINS', 'LPBUPT',
       'LPCAMT', 'ORGDT', 'TGL_AKAD', 'CBAL', 'DLM_JUTA', 'PMTAMT', 'BILINT',
       'MATDT6', 'CARCD', 'BIKOLE', 'BISIFA', 'BIREST', 'PREBAL', 'TERM',
       'PDDAYS', 'collectibility', 'arrears', 'due_date', 'DLRNO', 'MEMAMT',
       'PURBNK', 'REK_TAB', 'OUTLET', 'PROPER', 'NO HP', 'PERUSAHAAN'],
      dtype='object')
```

Figure 4. Cleaning Data

Figure 4. Cleaning data starting with ensuring the due date is not lost during cleaning, and ending with dropping the due date column if it is not needed. Continued by converting the collectability target into two classes, namely Current and Non-Current before the data splitting process.
Splitting Data in this study was divided into 2 elements, namely 70% training data and 30% testing data, the results obtained were (204, 2) (88, 2) (204,) (88,).

```
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1-Score: 1.0

Confusion Matrix:
 [[128   0]
 [  0  76]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       128
           1       1.00      1.00      1.00        76

    accuracy                           1.00       204
   macro avg       1.00      1.00      1.00       204
weighted avg       1.00      1.00      1.00       204
```

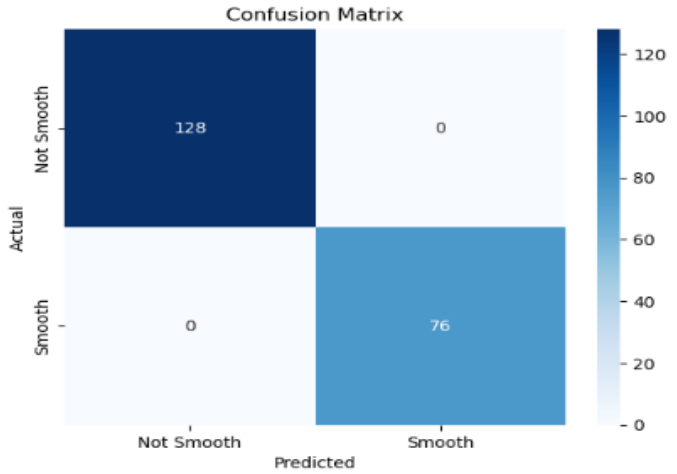

Figure 5. Evaluasi Model

The evaluation of this model shows that the value for accuracy is 1.0 (100%), precision 1.0 (100%), recall 1.0 (100%), and F1-Score 1.0 (100%). Then the Confusion Matrix as a prediction of actual test data and predictions shows:

1. 128, *true negatives* : for the number of non-smooth customer data.
2. 0, *false negatives* : for the number of smooth data.
3. 76, *true positives* : for the number of smooth data.
4. 0, *false positives* : for the number of non-smooth data.

After model evaluation, a unique value check is carried out to assist in further evaluation and ensure the quality of the model predictions. Continued by calculating the accuracy of each class with the training data prediction and calculating the total accuracy, the accuracy of each class is shown by an array of two elements, each of which is one in float format, then the result of the total accuracy is 1.0 (100%).

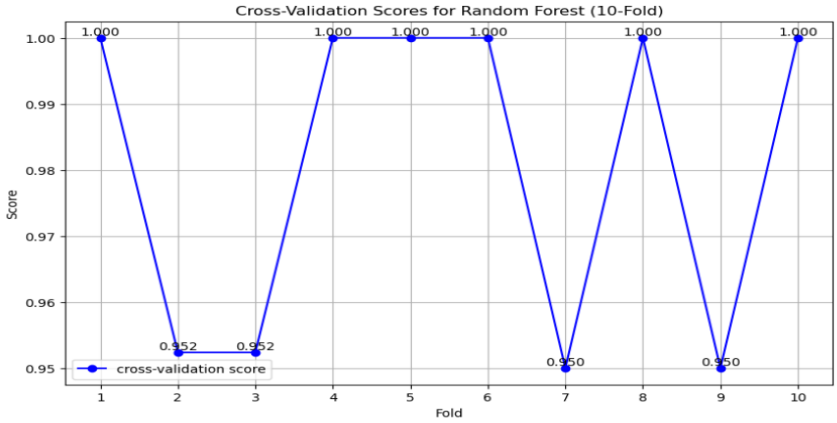

Figure 6. Cross-Validation

Shows that perfect testing on some folds reaches 1.0 (100%), and some folds show 0.95 (95%), which means there is inconsistent performance in the data subset. Continued with the accuracy of the test data of 100%.

```
Credit status prediction results:
Actual  Predicted
0       0           63
1       1           25
Name: count, dtype: int64
number of customers predicted to be smooth: 25
number of customers predicted to be not smooth: 63

Summary of arrears by predicted status:
               count      mean       std       min           25%       50%  \
credit_status
0              191.0  0.156949  0.123544  0.000056  5.700631e-02  0.128211
1              101.0  0.002663  0.002742  0.000000  5.937068e-07  0.002850

                    75%       max
credit_status
0              0.235147  1.000000
1              0.004513  0.009998

Customer data is not smooth:
      arrears  hari_sejak_jatuh_tempo
101  0.000056                0.302594
102  0.097867                0.481268
103  0.085505                0.403458
104  0.000694                0.763689
105  0.056758                0.466859
```

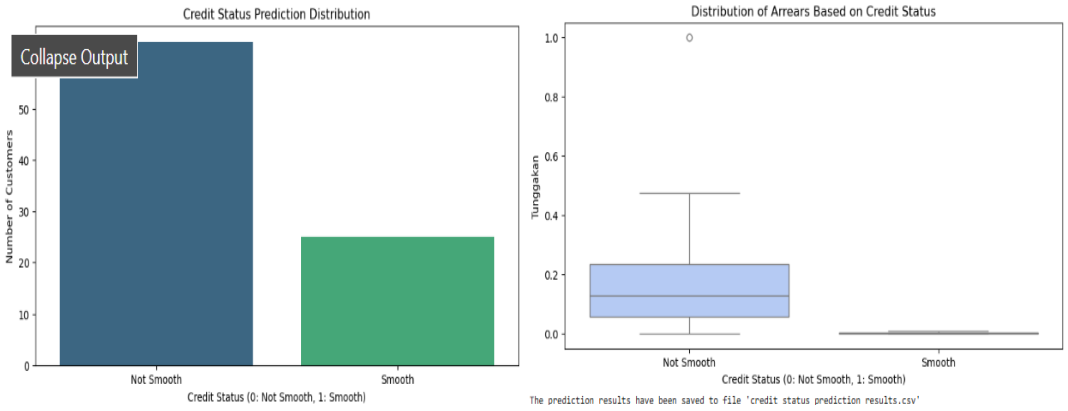

Figure 7. Result and Implementation

The credit status results show that there are 25 customers who are current and 63 who are not current. The results obtained have gone through a data splitting process, resulting in 25 and 63 customer data based on the current and non-current credit. The results have been displayed in the form of a bar chart, namely the distribution of credit status shows that non-current customer data is higher than current customer data. Then the boxplot diagram shows that the arrears value based on non-current is higher than current customers.

2) Discussion

This research was conducted with data taken in Microsoft Excel which has been tabulated in CSV (Comma Separated Values) format. Data processing uses KPR customer payment data of PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk totaling 292 data in 2024, from which data three variables were taken, namely Collectability, Arrears, and Due Date as determinants of KPR customer credit status. The following can be seen some examples of 292 KPR customer payment data.

Table 1. customer payment data table

collectibility	arrears	due_date
1	0	28/10/2039
1	0	28/10/2039

1	0	28/10/2039
2	1380000	21/05/2039
2	250.000	08/05/2039
2	750.000	29/05/2039
3	23,750.000	31/07/2039
3	41210001	30/05/2039
3	36004901	26/06/2039
4	50000009	22/01/2039
4	92000230	15/05/2039
4	56800000	08/03/2039
4	89000500	17/04/2039
4	59000785	03/04/2039
4	66003110	30/10/2039
4	98200000	07/05/2039
4	76901540	22/05/2039
5	130193200	22/04/2039
5	101230453	24/07/2039
5	149500000	06/06/2039
5	150500000	13/06/2039
5	100000500	14/06/2039
5	146380000	26/06/2039
5	160000000	10/07/2039
5	120000200	19/07/2039

The application of the Exploratory Data Analysis (EDA) method in this study aims to understand data patterns and data characteristics comprehensively, by displaying descriptive statistics and plots of three variables before entering preprocessing and algorithm models as determinants of customer credit status.

Model Evaluation, shows the results that Accuracy is 100% on training data, accompanied by recall, precision and F1-Score which are each 100%. This shows that the model can properly classify the credit status of KPR customers. The Confusion Matrix is used to determine the credit status with smooth and non-smooth classes, with correct predictions of true positive and true negative and incorrect predictions of false positive and false negative.

Cross-Validation Scores for RF (10 Fold), shows that perfect testing on some folds reaches 1.0 (100%), and some folds show 0.95 (95%) which means there is inconsistent performance in the data subset. This process indicates that the model performance works very well, but has some specific subsets that are less than optimal.

Implementation and Results, it can be seen that the results of the credit status prediction after Splitting Data are the number of customers predicted to be smooth is 25 customers, while the number of customers predicted to be not smooth is 63 customers. The display of a summary of arrears statistics based on the prediction status aims to make a comparison between the two categories. Customers who are said to be not smooth are much higher compared to current customers based on credit arrears. Then the results also observe arrears data and days since due date for customers whose credit status is not smooth.

The distribution of credit status using a bar chart shows the number of customers in the non-smooth category and the prediction of the smooth category has a small portion. Furthermore, the distribution of arrears based on credit status shows higher prediction results in the non-smooth category and less in the current category, the same as the distribution of credit status marked by the bar chart.

Based on the tests conducted by researchers, it was proven that the random forest algorithm successfully classified credit status into two classes, namely smooth and not smooth. Not only that, the algorithm also successfully showed an accuracy value of (100%), recall (100%), precision (100%), and F1-Score (100%). Then the test also showed the success of the EDA method in determining data patterns before entering the modeling process.

CONCLUSION

Based on the discussion and what the researcher explained previously, it can be concluded that the application of the Exploratory Data Analysis (EDA) method successfully processed data and analyzed credit status at PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk to form a classification model with machine learning compared to using traditional methods. Applying the Exploratory Data Analysis (EDA) technique to simplify data in analyzing in depth, explaining the relationship between variables, and performing preprocessing, namely cleaning data, handling missing values, normalizing numeric data, and splitting data very well.

The Random Forest algorithm is very accurate in classifying the credit status of KPR customers at PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk is capable of performing quite well, based on the results explained in the previous chapter that the evaluation uses accuracy, recall, precision, F1-Score, and confusion matrix with each being 1.0 (100%).

The results show that the Exploratory Data Analysis (EDA) method and the Random Forest Machine Learning algorithm are very important in the analysis. So that the credit status of mortgage customers who are declared non-performing is higher than the credit status declared current based on the customer's credit arrears variable. This illustrates that there are significant features that influence credit status, which are useful in making credit risk decisions. Regarding the pattern and distribution of mortgage customer credit data in an intuitive understanding for distribution analysis and visualization of prediction results.

Based on the results of the study conducted in this study, it can be suggested for further research to improve model reasoning, it is recommended to use a larger and more diverse dataset. Considering other Machine Learning algorithms such as GBoost (Gradient Boosting) as a consideration for comparison with Random Forest performance. It is recommended to expand the variables with analysis that affects credit status such as income variables, macroeconomics, and customer liabilities. he author hopes that PT. Bank Tabungan Negara KCP Lubuklinggau Persero Tbk can utilize the results of the customer credit status analysis as a consideration in formulating effective and proactive policies to deal with customers with high credit risk.

REFERENCES

Andrew Murphy. (2019). *No Title*. https://radiopaedia.org/articles/random-forest-machine-learning?lang=us

Aryanti, D., & Setiawan, J. (2019). Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia Periode 2014-2018. *Ultima InfoSys*, *9*(2), 86–91. https://doi.org/10.31937/si.v9i2.991

Budianti, L., & Suliadi. (2022). Metode Weighted Random Forest dalam Klasifikasi Prediksi Kelangsungan Hidup Pasien Gagal Jantung. *Bandung Conference Series: Statistics*, *2*(2), 103–110. https://doi.org/10.29313/bcss.v2i2.3318

Exploratory Data Analysis (EDA) for Data Science and ML. (n.d.). https://cognitiveclass.ai/courses/exploratory-data-analysis-eda-for-data-science-and-ml#about-course

Liaw, A., & Wiener, M. (2018). Klasifikasi dan Regresi oleh RandomForest. *R News*, *2*(3), 18–22.

Ningsih, P. T. S., Gusvarizon, M., & Hermawan, R. (2022). Analisis Sistem Pendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Machine Learning. *Jurnal Teknologi Informatika Dan Komputer*, *8*(2), 386–401. https://doi.org/10.37012/jtik.v8i2.1306

Risiko, M., Suhadolnik, N., Ueyama, J., & Silva, D. (2020). *Pembelajaran Mesin untuk Peningkatan Penilaian Risiko Kredit : Sebuah Pendekatan Empiris*.

Santosa, C. M. P., Sumirat, E., & Sudrajad, O. Y. (2024). An Exploratory Data Analysis (EDA) Approach for Analyzing Financial Statements in Pharmaceutical Companies Using Machine Learning. *International Journal of Current Science Research and Review*, *07*(07), 4676–4685. https://doi.org/10.47191/ijcsrr/v7-i7-12

Sidik, Z. (2019). *Klasifikasi Kelancaran Kredit Furniture Menggunakan Algoritma K-Nearest Neighbor Berbasis Forward Selection*.

Ulandari, K. P., Chamidah, N., & Kurniawan, A. (2024). *Prediksi risiko gagal bayar kredit kepemilikan rumah dengan pendekatan metode random forest*. *13*(2).

Wikipedia. (n.d.). *Bank Tabungan Negara (BTN)*. https://id.wikipedia.org/wiki/Bank_Tabungan_Negara

Wulansari, W., & Purwitasari, D. (2023). Algoritma Random Forest pada Prediksi Status Kredit Usaha Rakyat untuk Mengurangi Nonperforming Loan Rate. *Journal of Intelligent System and Computation*, *5*(2), 109–114. https://doi.org/10.52985/insyst.v5i2.358