

UTILIZING SCIENCE DATA TO INCREASING THE NUMBER MSME DEBTORS AT PT.BANK CENTRAL ASIA.TBK (CASE STUDY OF PT. BANK CENTRAL ASIA.TBK KCU TEBING TINGGI)

Effan Budiawan¹, Meilita Tryana Sembiring², Nazaruddin³

¹Master Management Study Program, Post Graduate School, Universitas Sumatera Utara

^{2,3}Lecturer at Master Management Study Program, Post Graduate School, Universitas Sumatera Utara

Corresponding Author: melita_tryana@yahoo.co.id

Abstract

This study aims to increase the number of MSME debtors at the BCA Tebing Tinggi Branch. Since the enactment of Bank Indonesia Regulation Number 23/13/PBI/2021 concerning the Macprudential Inclusive Financing Ratio (RPIM) for Conventional Commercial Banks, Sharia Commercial Banks, and Sharia Business Units. So Commercial Banks began to adjust the percentage of the use of funds that would be used to finance MSMEs and PBR. BCA Tebing Tinggi Branch is committed to meeting the increase in the percentage of RPIM. One way that can be used to explore Potential Funding is by Utilizing Data Science. Data science studies data, especially quantitative data, with the aim of finding hidden patterns in the data. Researchers will study profile information and transaction patterns in accounts to find MSME customers who are given the right financing. This study processes data using the Machine Learning method with the Random Forest algorithm.

Keywords: Data Science, Machine Learning, Random Forest, MSME Debtors, BCA.

INTRODUCTION

Since the enactment of Bank Indonesia Regulation Number 23/13/PBI/2021 concerning the Macprudential Inclusive Financing Ratio (RPIM) for Conventional Commercial Banks, Sharia Commercial Banks, and Sharia Business Units. So Commercial Banks began to adjust the percentage of the use of funds that would be used to finance Micro, Small and Medium Enterprises (MSMEs) and Low Income Individuals (PBR). This PBI was issued as one of Bank Indonesia's efforts to increase economic inclusion and open access to finance and strengthen the role of MSMEs in national economic recovery. BCA as the largest private commercial bank in Indonesia is committed to complying with these Bank Indonesia Regulations. As of June 2022, BCA has disbursed loans to MSMEs of IDR 94.2 trillion, or growing up to 19.6% on an annual basis or year on year (yoy). The portion only reached 13.94% of the company's total loans. This means that it is still far from the regulator's target. BCA's target in 2024 for lending to MSMEs is 30% in accordance with regulatory provisions. In line with this, BCA branches throughout Indonesia are expected to be able to more quickly and actively seek potential MSME financing.

Micro, Small and Medium Enterprises (MSMEs) thrive in a number of areas. This can be seen from the data reported by the Ministry of Cooperatives and Small and Medium Enterprises (Kemenkop UKM), the total number of MSMEs in Indonesia will exceed 8.71 million business units in 2022. Java dominates this sector. It was recorded that West Java became the MSME champion with a total of 1.49 million business units. Thin in second place is Central Java, which reached 1.45 million units. Third, there is East Java with 1.15 million units. Outside of the top three, the gap is quite far. DKI Jakarta, which won the fourth position, can carve almost 660 thousand units. Fifth, there is North Sumatra with 596 thousand units. While the least number of businesses are in three regions, namely West Papua with 4.6 thousand business units, North Maluku with 4.1 thousand units, and Papua with 3.9 thousand units. The following is the data and distribution of MSMEs in Indonesia in 2022.

As of April 2023 the number of BCA MSME debtors was 49,185 units and those managed by the BCA Tebing Tinggi branch were 132 units. When compared to the total number of MSMEs in North Sumatra, it can be seen that there is still considerable potential to grow the number of debtors. One way to get financing potential is to take advantage of the data available at BCA. As of December 2022, the number of customers in the KCU of the High Cliffs is 40,562 customers. From this data, 1140 K1 customers were selected to be used as potential financing. K1 customers are customers who are included in the list of Group 1 (K1) customers determined by the Network Management & Regional Planning Work Unit and the Individual Customer Business Development Division which,

based on data analysis, bring the biggest business and profit to BCA. Withdrawal of customer data is done with the help of RPA (Robotic Process Automation) and then the data is processed so that it is useful. To process such a large amount of data certainly requires Data Science because if it is done conventionally with tabulations and analysis it certainly takes a long time and is less efficient.

DataScience is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insight from structured and unstructured data (Dhar, 2013), and apply knowledge and actionable insights from data in various application domains. . Data science is related to data mining, machine learning, and big data. In simple terms, data science happens when we work with data to find answers to questions. The emphasis is more on the data itself and not on the science or knowledge. If we have data, then we have curiosity about the contents of useful data and then to answer this curiosity by studying the data, explore the data and do various things to analyze the data by utilizing certain science and technology to find answers. The ultimate goal of data science is to find insight from data. Data Science can be seen as a process of distilling or extracting or exploring insights from the data. The processed data can be very large. This insight can be likened to gold or diamonds, which, although only a few or small in size, are still valuable. This insight can be in the form of important information or models made from data that will be very useful in making a decision. Insights to be obtained from data need to start with a strong sense of curiosity from oneself or from an organization (in the form of a need because there is a problem to be solved by utilizing data). Furthermore, armed with this, later a data scientist can carry out various activities by utilizing appropriate knowledge and technology to get the desired insight (Sembiring & Hasibuan, 2021). Researchers through the BCA Tebing Tinggi Branch will explore this potential by utilizing Data Science to accelerate financing acquisitions. Data Science is a compilation of various techniques to process data to become valuable (Vijay Kotu, 2019). The technique that will be used by researchers is to look for certain patterns of customer account transactions that are likely to be potential MSME financing.

LITERATURE REVIEW

1. Bank Credit

Banking Law Number 10 of 1998 Article 1 concerning credit explains that credit is the provision of money or claims that can be equated with it, based on a loan agreement or agreement between banks and other parties that requires the borrower to repay the debt after a certain period of time with the provision of interest. . According to Hermansyah (2008) approval of a credit application is carried out by referring to the 5C Formula, namely Character, Capacity, Capital, Collateral, and Condition of Economy. The meaning of the word credit in the banking world is that creditors, both financial institutions, provide loans to debtors based on trust (Tjoekam, 1999). If associated with a business activity, credit means an activity that provides economic value to the debtor based on trust, and the same economic value will be returned by the creditor within a period determined by both parties.

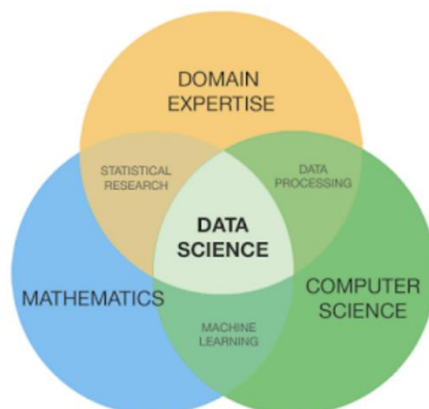
2. MSME Financing

Growth in access to credit by MSMEs can increase economic growth. Apart from being beneficial for the state in order to stabilize the economy, credit is also beneficial for MSMEs as a secure and sustainable source of funding. In Kakuru's (2008) study, almost all commercial banks include MSMEs in their credit schemes to develop access to formal credit. According to Lusimbo & Muturi (2012), access to formal credit is defined as the absence of obstacles related to administrative or procedural costs at formal credit provider institutions that are felt by MSMEs when applying for credit. MSMEs are inseparable from financing problems. According to Agus (2016), the problems faced by MSMEs are not only about financial management and human resources in it, but also related to financing. In order to increase ease of financing for MSMEs, the government is expanding MSME financing facilities in Indonesia with the People's Business Credit (KUR) program. KUR aims to facilitate capital for MSME actors, but it has not been fully realized. Until May 2016, KUR disbursement was only IDR 39.2 trillion of the total target of IDR 120 trillion this year (Noordiansyah, 2016). This has caused not all MSMEs to feel the KUR distribution.

3. Definition and concept of data science

Data Science is a concept to unify statistics, data analysis, informatics, and related methods to understand and analyze actual phenomena with data (Hayashi, 1998). It uses techniques and theories drawn from various fields in the context of mathematics, statistics, computer science, information science, and domain knowledge. However,

data science is different from computer science and information science. Turing Award winner Jim Gray envisions data science as the "fourth paradigm" of science (empirical, theoretical, computational and now data driven) and asserts that everything about science changes because of the impact of information technology and data deluge (Tony Hey, et al. 2009).



Source: Palmer, Shelly

Figure 2.1 Interdisciplinary Discipline of Data Science

The availability of very large data with inherent properties in big data affects the process of conducting data analysis. Data analysis includes the process of collecting, processing, and evaluating data to understand and draw conclusions from the information found in the data. The goal is to identify patterns, trends, and relationships between variables in data to solve problems or make informed decisions. Data analysis can involve various techniques such as statistics, mathematics, and computer engineering (Kudyba, S., 2014).

4. Data Mining

According to Davies (2004), simply data mining is mining or new information by looking for certain patterns or rules from a very large amount of data. Data mining is often also referred to as knowledge discovery in databases (KDD), which is an activity that includes collecting, using historical data to find regularities, patterns or relationships in large data sets (Santosa, 2007). Meanwhile, according to Turban (2005), data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases.

5. Machine Learning

Datamining is a more effective statistical learning approach and uses powerful features for data analysis. However, there are some problems with the complexity of the pre-processed data and the knowledge of experts. Recent advances in machine learning technology are helping to provide a new effective paradigm for deriving end-to-end machine learning models from complex data (Mastoli et al., 2019). Machine learning classification methods have long been used in the field of data mining and many other fields of computer science as a classification algorithm involving methods from statistics, artificial intelligence and database management, can also be categorized as a key element in data interpretation and data visualization (Berry & Linoff, 2004). Machine learning requires data to learn so it is also termed learn from data (Alpaydin, 2010).

RESEARCH METHODS

1. Types of Research

This type of research is a mixed research between descriptive and quantitative. The first method used in this study is descriptive analysis, where this method aims to provide an overview of the data to be studied. Furthermore, the second method is machine learning with the random forest algorithm where this method is used to classify credit usage on existing debtor data of BCA KCU Tebing Tinggi. Next, make predictions about the possibility of using credit for data on K1 customers who are not yet debtors.

2. Study Population

Population in this research is the customer data of BCA KCU Tebing Tinggi, namely as many as 1140 Group 1 customers. This data includes data on existing debtors, non-entrepreneur customers, MSME entrepreneurs and non-SMEs. For further research, several categories will be sorted before further processing.

3. Research Variables

There are five feature variables that will be trained in this study, namely average balance, credit transfer, credit and balance mutation ratio, age, and line of business. Then one target variable, namely the classification of usage which is a description of the average loan usage.

The following is a description of the variables above

Table 1 Variable Operational Definitions

No.	Variable	Definition	Roles
1.	C-Usage	Classification of the average use of the loan	<i>target</i>
2.	Average balance	Average customer deposit balance	<i>feature</i>
3.	Credit mutation	The number of customer account credit mutations	<i>feature</i>
4.	Ratio	The ratio between account credit movements compared to the average balance	<i>feature</i>
5.	Age	Customer age	<i>feature</i>
6.	Business fields	Type of customer's business	<i>feature</i>

4. Data Collection Techniques

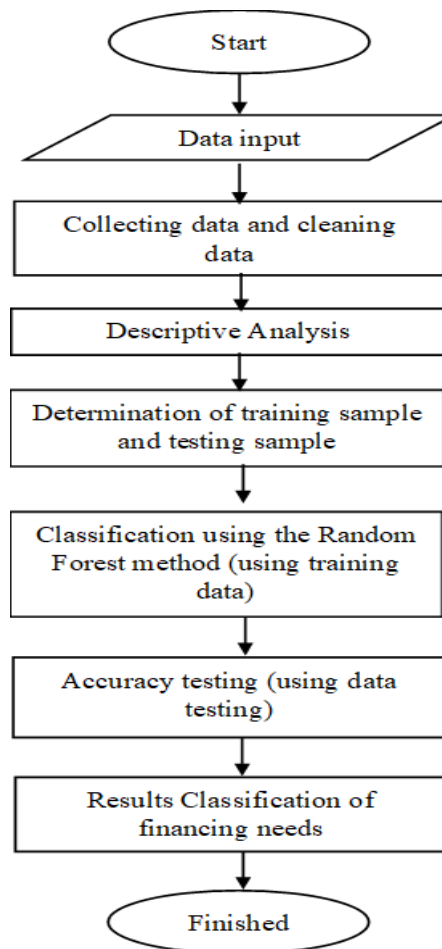
In this study, the data used were secondary data obtained from the BCA KCU Tebing Tinggi. The data is retrieved using the help of Robotic Process Automation (RPA) through data available on branch computers through a CRM application called Specta.

5. Tools and Ways of Organizing Data

The data processing tool used is the Ms. program. Excel and Orange Data Mining. In this stage, an assessment of the data that has been obtained is based on the theory, namely descriptive data analysis and using the random forest method, then implementing the following steps:

1. Perform a descriptive analysis of the variables used with the scatter plot
2. Determine the training data (training) that will be used to perform the random forest method.
3. Choose a value of n which represents the number of trees.
4. Make a plot to see the training data training process in order to get more optimal results.
5. Choose the sampling technique that will be used in the random forest method.
6. Classifying data using the random forest method using predetermined training data.
7. Conduct accuracy testing using test data (testing) to see how much accuracy the random forest method has for research data.

6. Research Flowchart



RESULTS AND DISCUSSION

1. Data mining

The research began with data collection from the SPECTA CRM BCA application which was carried out automatically with the help of Robotic Process Automation (RPA) technology using the Power Automate tool from Microsoft.

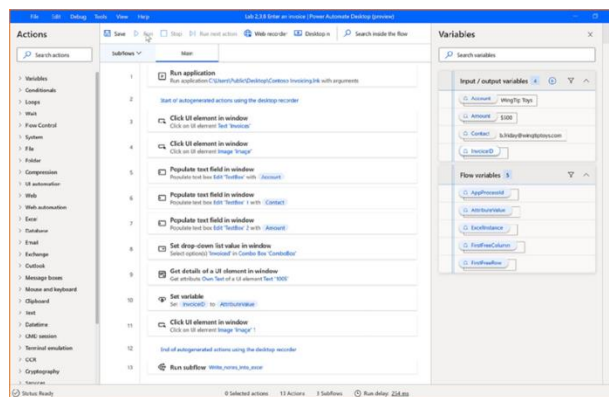


Figure 1 RPA Flow of automated data withdrawal from BCA

The resulting data will be cleaned and transformed using the Microsoft Excel application. This is necessary to correct some data errors, clean up confidential data and add variables. Obtained 102 existing debtor data that can be used as a source of train data to be trained by machine learning algorithms.

1	A	B	C	D	E	F	G	H	I	J
1	CIS	Nama	Saldo-rata2	Mutasi Kredit	ratio	Usia	Pekerjaan / Bidang t LOB		Usagae%	C-Usage
2			14.062.179,15	572.128,191	41	59	PENGUSAHA PABRII KILANG PADI DAN PLASTIK		99%	High
3			15.348.524,73	526.453.821,34	34	49	PENGUSAHA JASA	#N/A	87%	High
4			6.096.251.218,56	5.055.574.252,17	1	44	PEDAGANG PERTANIAN/PENJUALAN PUPIUK		21%	Low
5			65.147.314,30	1.243.261.240,93	19	38	PEDAGANG SPERPARIS MOTOR		32%	Moderate
6			23.563.596,26	1.522.620.917,80	65	42	PEDAGANG TOKO OBAT		96%	High
7			195.492.704,48	979.801.574,07	5	37	PEDAGANG KILANG PADI		88%	High
8			19.431.394,14	344.922.608,40	18	44	PEDAGANG JUAL BELI SEPEDA MOTOR		76%	High
9			76.681.946,85	320.282.109,46	4	38	PEDAGANG PONSEL		72%	High
10			12.709.100,44	219.792.945,17	17	60	PEDAGANG SPBU		89%	High
11			1.000.000,00	305.464.666,67	305	74	PEDAGANG PENGECEK MINYAK MAKAN & BERAS		46%	Moderate
12			236.094,43	294.972.982,50	1.249	35	PEDAGANG SPAREPART MOTOR & PERBAIKANNYA		86%	High
13			43.517.585,91	5.370.758.802,31	123	37	PEDAGANG SHOWROOM MOBIL		81%	High
14			154.194.361,72	130.487.419,97	1	53	PEDAGANG JUAL BELI MOTOR		0%	Low
15			784.555.481,67	192.349.105,56	0	46	PENGUSAHA JASA JASA LEVALANSIR		0%	Low
16			16.721.450,63	263.303.252,08	16	38	PEDAGANG TOKO BAHAN BANGUNAN		83%	High
17			15.085.785,22	159.508.164	11	49	PEDAGANG	#N/A	94%	High
18			50.976,45	673.315.474,87	13.208	39	PEDAGANG PENGEPEL BESI BAJA (BUTUT)		42%	Moderate
19			636.407.635,50	482.921.964,19	1	71	PEDAGANG BATU PERMATA/LOGAM MULIA		0%	Low
20			38.646.635,35	1.071.022.530,28	28	41	PEDAGANG DISTRIBUTOR BISKUIT & SNACK		53%	Moderate
21			1.000.000,00	261.291.666,67	261	49	PEDAGANG	#N/A	67%	Moderate
22			270.946.295,86	314.888.058,85	1	45	PENGUSAHA PABRII PERACIPAN KAYU		0%	Low
23			19.493.502,62	122.683.223,84	6	65	PEDAGANG TOKO JAM		98%	High
24			381.775.583,82	643.946.495,91	2	41	PEDAGANG PENGECEK CAT DINDING		100%	High
25			1.000.000,00	571.357.864,33	571	54	PEDAGANG TOKO PUPUK DAN PESTISIDA		73%	High
26			184.953.950,57	432.501.793,65	2	50	PEDAGANG TOKO PERABOT&INDUSTRI PERABOT		52%	Moderate

Figure 2 Data Cleaning and Transformation

Researchers use the Orange data mining application for data science processing. The Orange data mining application is an open source application that emphasizes visual programming in data mining.

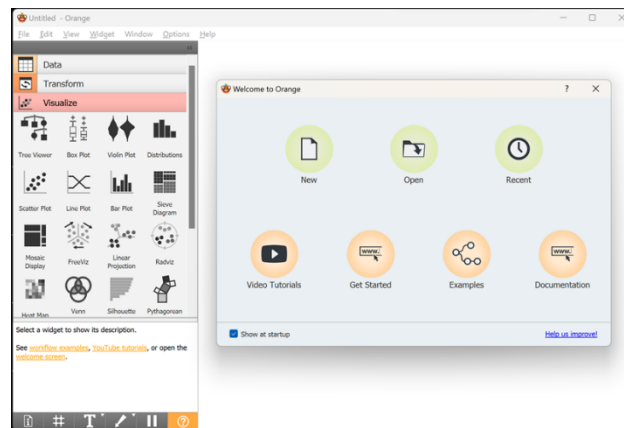


Figure 3 The initial appearance of the orange data mining application

Researchers will see an overview by making a descriptive analysis of each variable of customer data withdrawal results, researchers will review the data type of each variable used in the orange application. Average balances, credit transfers, ratios, ages, and line of business will be used as role features to be trained. Meanwhile, C-usage, which is the average credit usage, will be used as a target for machine learning training.

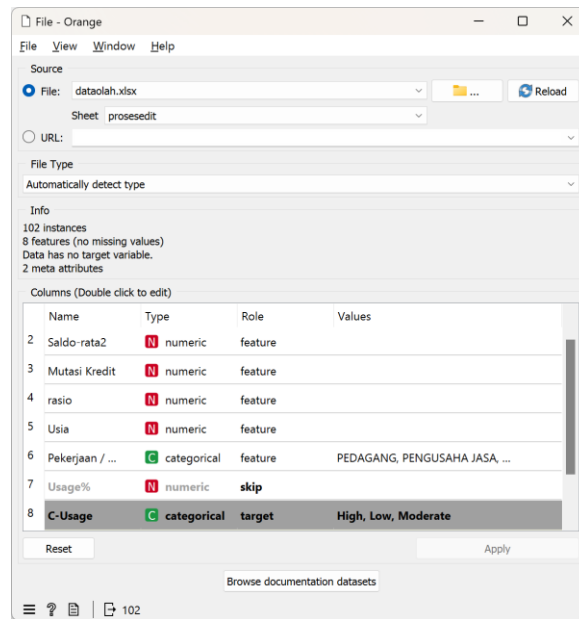


Figure 4 variable types and role determination

The results in the data table show that there are continuous variables such as average balance, credit transfer, ratio and age. There are also categorical variables, namely Business Field and C-Usage. C-Usage is categorized into 3, namely:

C-Usage	Use of Credit
high	>70%
Moderate	> 30% to 70%
Low	<=30%

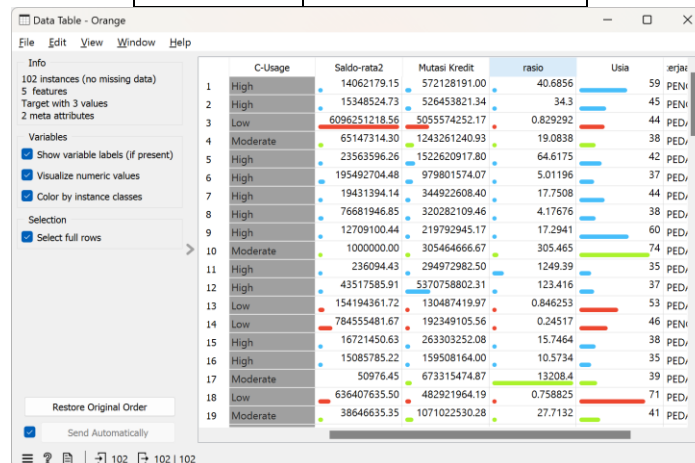


Figure 4.5 view data table

2.Descriptive Analysis

For descriptive analysis, the researcher used a scatter plot visualization and obtained an overview of variable data.

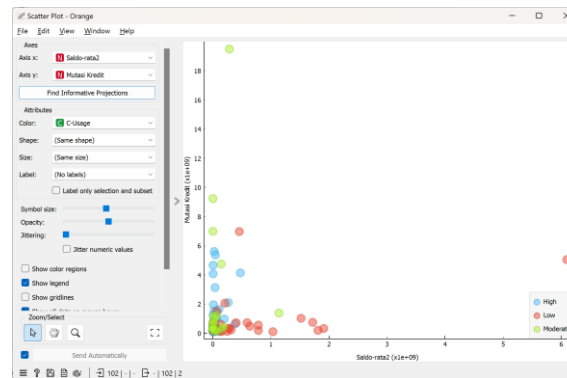


Figure 6 data outside the range

It can be seen that there is a relationship between the use of credit with the average balance and credit mutations. Collections of red dots or debtors with low usage predominate in the area of large average balances and small credit mutations. Meanwhile, a collection of blue dots or debtors with high usage dominates the area of small average balances and large credit mutations. From this descriptive analysis it can be concluded that the use of debtors' credit can be affected by the average balance and credit mutations. Then the results of the scatter plot also show that there are some data that are too far away from the data group. Some of this data needs to be excluded so that the machine learning algorithm can work optimally.

3. Random Forest Algorithm

With the help of the outlier function from Orange Software can help reduce data that is out of range. The results of the outliers obtained 10 amounts of data that are not valid for use.

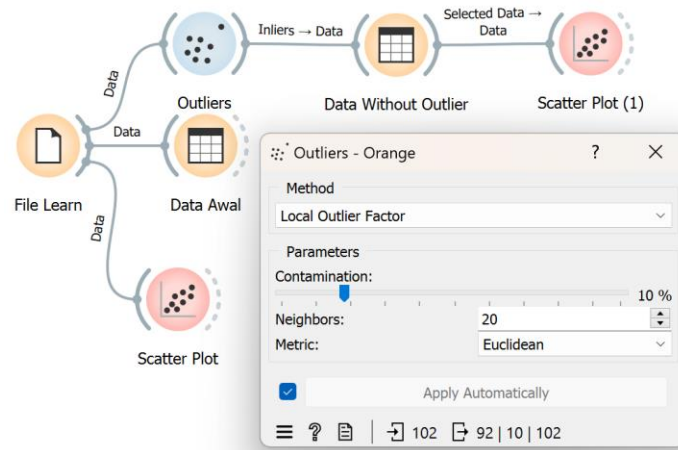


Figure 7 Outlier-Orange

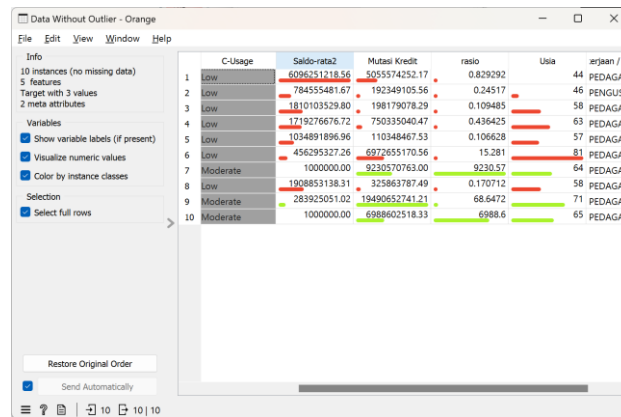


Figure 8 results of outlier data

After data reduction then it is ready to be used in machine learning algorithms. Here the researcher uses the random forest algorithm to train data obtained from existing debtors.

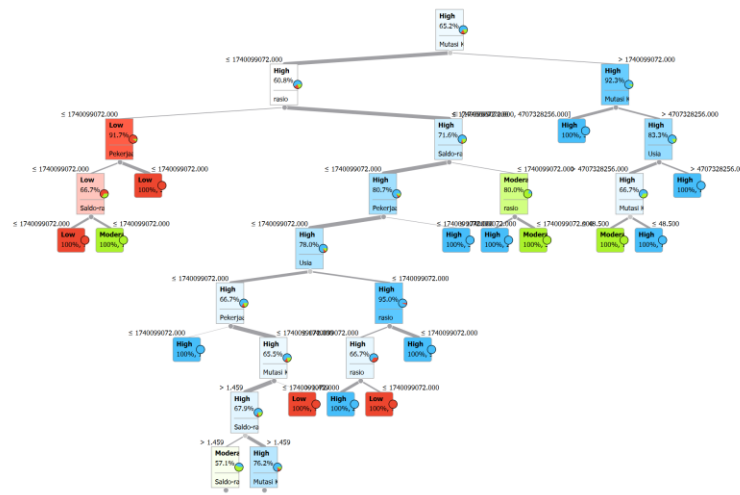


Figure 9 Decision Tree

The random forest algorithm is a collection of a group of decision trees that are used simultaneously using the Bagging or Bootstrap Aggregating method. The visualization results of the Pythagorean forest with a total of 10 trees show how the random forest algorithm works in determining predictions.

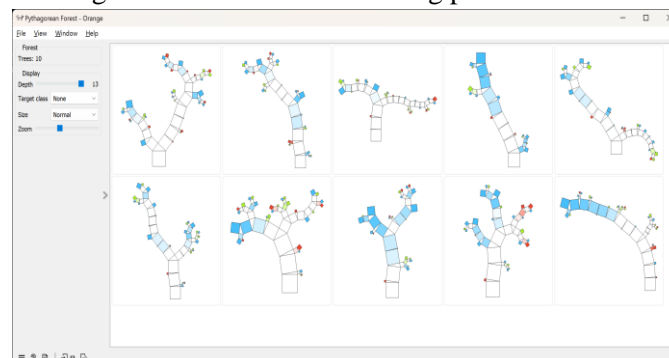


Figure 10 Pythagorean Random Forest

Performance predictions from the random forest algorithm can be measured using several models including the AUC area under the curve, CA Accuracy, F1 Score, Precision, Recall and MCC. The measurement results of several models show a fairly good predictive performance close to 1.

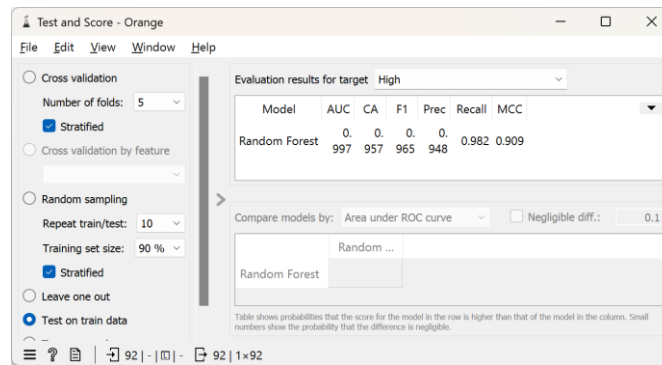


Figure 11 results of the random forest algorithm

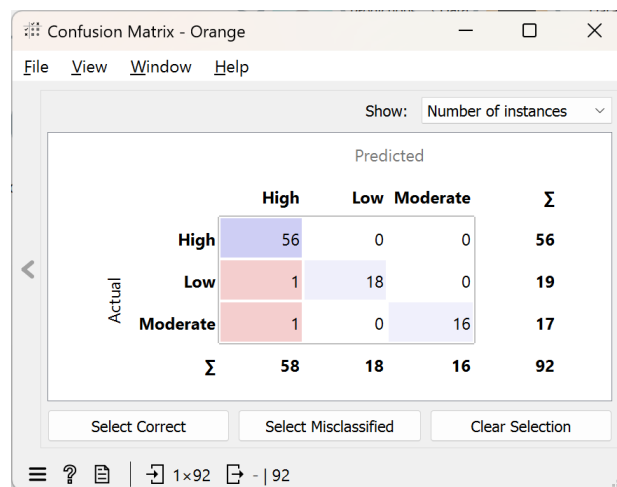


Figure 12 Confusion Matrix

By looking at the confusion matrix table, it can be concluded from the prediction results of the random forest algorithm that comes from a total of 92 data. For the use of High credit, the total prediction is entirely correct, namely from 56 data. Meanwhile, the use of Low credit has a total of 18 correct predictions out of a total of 19 data. And for the use of moderate credit, there are 16 correct predictions out of a total of 17 data.

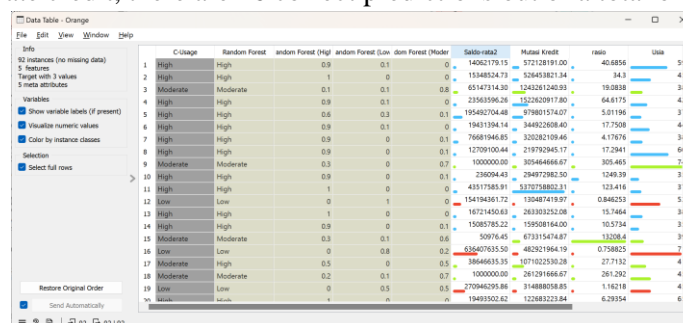


Figure 13 Learning results and predictions of the random forest algorithm with train data

The final stage of the research is the random forest algorithm which will be used as a prediction tool with the random forest algorithm which has learned from 92 train data of existing debtors. The algorithm is then used to predict K1 data and provide an overview of credit usage if the customer becomes an MSME debtor. Of the 1141 K1 data that were cleaned and transformed, there were 685 data that could be continued to the Prediction stage.

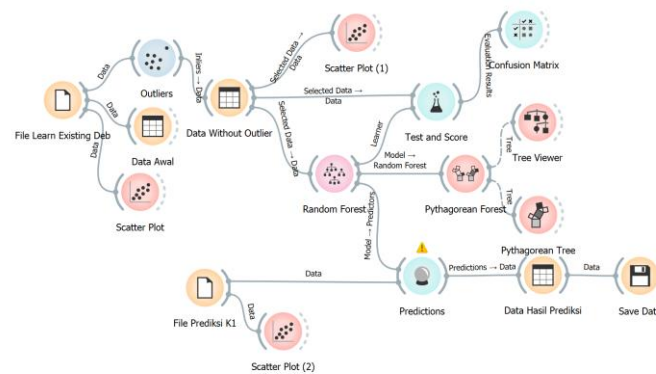
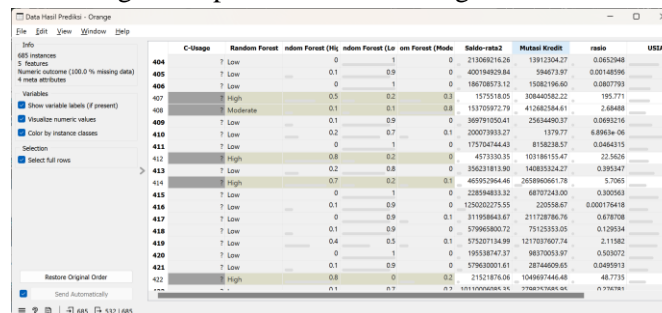


Figure 14 Random Forest Algorithm Final Flow and Orange Prediction

In summary, flow machine learning starts from the input data train, which has previously gone through a cleaning and transformation process from existing debtors, then the data is reduced using outliers for data outside the range. The data train then studied the random forest algorithm. Furthermore, the K1 data input that has gone through the cleaning and transformation process is as much as 685 data. The data is predicted with the random forest algorithm and produces the resulting data.

4. Machine Learning Prediction Results

The prediction results of 685 K1 customers concluded that if credit is given to these customers, it is likely that there will be usage according to the pattern of the existing debtor.



	C-usage	Random Forest	ndom Forest (H)	ndom Forest (L)	om Forest (M)	Seldo-rata2	Mutasi Kredit	ratio	USA
404	Low	0	1	0	21306216.26	1391204.27	0.0625948		
405	Low	0.1	0.9	0	40219829.84	394031.87	0.0148396		
406	Low	0	1	0	186708373.12	15082196.80	0.0807793		
407	High	0.5	0.2	0.3	1575518.05	30640582.22	195.771		
408	Moderate	0.1	0.1	0.8	151709372.79	41268204.61	2.68488		
409	Low	0.1	0.9	0	369791002.41	29244003.97	0.0823216		
410	Low	0.2	0.7	0.1	300073933.27	1379.77	6.8963e-06		
411	Low	0	1	0	175704744.43	8158238.57	0.0464315		
412	High	0.8	0.2	0	4573330.35	103189155.47	22.9426		
413	Low	0.2	0.8	0	352821913.90	14839324.27	0.395347		
414	High	0.7	0.2	0.1	465952964.46	265990661.78	5.7065		
415	Low	0	1	0	228584833.32	68707433.00	0.300563		
416	Low	0.1	0.9	0	1250202275.55	220558.67	0.00179418		
417	Low	0	0.9	0.1	311958462.67	211729186.76	0.678706		
418	Low	0.1	0.9	0	579968800.72	75125533.05	0.129534		
419	Low	0.4	0.5	0.1	575207134.99	1217037607.74	2.11582		
420	Low	0	1	0	19558747.37	9837003.97	0.509072		
421	Low	0.1	0.9	0	579630021.61	28744026.65	0.0499313		
422	High	0.8	0	0.2	21521876.06	104897446.48	48.7735		

Figure 15 results of predictions and recommendations for new MSME debtors

Prediction results obtained by 136 K1 customers when given credit then the usage is High, 17 K1 customers use Moderate and 532 K1 customers use Low. These results can provide Managerial Impact to determine strategies in prospecting prospective borrowers. Of course, if you offer credit to 136 customers with high usage, the probability of success is greater than if you offer credit to 532 customers with low usage. This strategy is expected to help the growth of MSME debtors at the BCA Tebing Tinggi branch.

CONCLUSION

Based on the analysis of research results that have been carried out with data on existing debtors and K1, it can be concluded:

1. The results of the descriptive analysis using scatter plots show that there is a relationship between variables such as average balance, account mutation, ratio, age and line of business with the use of credit from existing debtors. This can be the basis for making these variables a factor influencing the use of credit.
2. The random forest algorithm was applied to study the tendency to use credit that was trained using 92 data on existing debtors and the results have fairly good accuracy. Then the algorithm is used to predict 685 potential K1 customer data to become debtors. The prediction results show that there are 136 customers

with a high probability of using it, 17 customers with a moderate probability and 532 customers with a low probability.

REFERENCES

- Alpaydin, E. 2010. Introduction to Machine Learning. 2nd Edition, The MIT Press, Cambridge.
- Amir Mahmud Husein dkk. 2021. Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning. Data Science Indonesia 2021 : vol 1, number 1
- Atieno, R. 2001. Formal and informal institutions lending policies and access to credit by small-scale enterprises in Kenya. AERC Research Paper 111 : 1-53.
- Aznovri Kurniawan et al. 2022. Pemilihan Metode Predictive Analytics dengan Machine Learning untuk Analisis dan Strategi Peningkatan Kualitas Kredit Perbankan. Indonesian Journal of Applied Statistics : 5(1), 1-11, 2022
- Baas, T., dan Schrooten, M. 2006, Realtionship Banking and SMEs: A Theoretical Analysis . Small Business Economics, 27.
- Berry, M., and Linoff, G. 2004. Data Mining Techniques for Marketing, Sales and Customer Support (2nd ed.). Wiley, New York.
- Bonggo Bawono dan Rochdi Wasono. 2019. Perbandingan Metode Random forest dan Naïve Bayes untuk klasifikasi debitur berdasarkan kualitas kredit. ISBN : 2685 – 5852
- Han, Jaiwei., Kamber, Micheline. Pei, Jian. 2011. Data Mining: Concepts and Techniques. 3rd ed. Morgan Kaufmann. ISBN 978-0-12-381479-1
- Hayashi, Chikio. 1998. What is Data Science? Fundamental Concepts and a Heuristic Example. In Hayashi, Chikio., Yajima, Keiji., Bock, Hans-Hermann., Ohsumi, Noboru., Tanaka, yutaka., Baba, Yasumasa. Data Science, Classification, and Related Methods. Studies in Classificataion, Data Analysis, and Knowledge Organization. Springer Japan. pp. 40-51.
- Hermansyah. 2008, Hukum Perbankan Nasional Indonesia. Prenada, Yogyakarta
- Hey, Tony., Stewart Tansley., Kristin Michele Tolle. 2009. The Fourth Paradigm : Data-Intensive Scientific Discovery. Microsoft Research. ISBN 978-0-9825442-0-4
- Kakuru, J. 2008, The Supply-Demand Factors Interface and Credit Flow to Small and Micro Enterprises (SMEs). Stirling, University of Stirling.
- Kotu, Vijay and Bala Deshpande. 2019, Data Science Concepts and Practice Second Edition. Elsevier Inc, Cambridge
- Kumar Arun et al. 2016. Loan Approval Prediction based on Machine Learning Approach. IOSR Journal of Computer Engineering : 2278-0661
- L. Sathish kumar, V. Pandimurugan, D. Usha et al. 2021. Random forest tree classification algorithm for predicating loan. Proceedings. 57 : 2216–2222
- Larose, D.T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley , New York
- Lin Zhu et al. 2020. A study on predicting loan default based on the random forest Algorithm. Procedia Computer Science 162 : 503–513
- Lusimbo, N. E., and Muturi, W. 2015, Financial Literacy and The Growth of Small Enterprises in Kenya: A Case of Kakamega Central Sub-Country, Kenya. International Journal of Economics, Commerce and Management :828-845.
- MacLennan, J., Tang, Z., dan Crivat, B. (2009). Data mining with Microsoft SQL server 2008. Indianapolis, IN: Wiley Pub.
- Manahan, S.E. 2005, Environmental Chemistry. CRC Press, Boca Raton
- Mehul Madaan et al. 2020. Loan default prediction using decision trees and random forest: A comparative study.

- IOP Conf. Series: Materials Science and Engineering : 1022-012042
Nkundabanyanga, K. S., Kasozi, D., and Nalukenge, I. 2014. Lending terms, financial literacy, and formal credit accessibility.
International Journal of Social Economics : 41(5), 342-361.
Moh. Tjoekam. 1999, Perkreditan Bisnis Inti Bank Komersial. Gramedia, Jakarta
Olson, D. L. 2007. Data Mining in Business Services. Service Business, 181-193
Paweł Ziemba et al. 2021. Credit Decision Support Based on Real Set of Cash Loans Using Integrated Machine Learning Algorithms.
Electronics : 10-2099
Rudiantoro, R., dan Siregar, S. V. 2012. Kualitas Laporan Keuangan UMKM serta Prospek Implementasi SAK ETAP.
Jurnal Akuntansi dan Keuangan Indonesia : 9(1), 1-12.
Santosa, Budi. 2007, Data Mining : Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Graha Ilmu, Yogyakarta
Sembiring, Meilita Tryana dan Chalis Fajri Hasibuan. 2021, Data Science Strategi UMKM Dalam Pengambilan Keputusan. USU Press, Medan
Stiglitz, J., dan Weiss, A. 1981. Credit Rationing in Markets with Imperfect Information.
The American Economic Review : 71(3), 393-410.
Subhiksha R et al. 2022. Bank Loan Approval Prediction Using Data Science Technique (ML).
SJ Impact Factor : 7.538 Volume 10 Issue V May 2022
Tambunan, Tulus T.H. 2021, UMKM di Indonesia. Prenada, Jakarta
Turban, E., dkk. 2005, Decision Support Systems and Intelligent Systems. Andi Offset, Yogyakarta
Vasant Dhar. 2013, Data Science and Prediction , Article in Communications of the ACM, Vol. 56 No. 12.
X. Ye et al. 2018. Loan evaluation in P2P lending based on Random Forest optimized by genetic algorithm with profit score.
Electronic Commerce Research and Applications 32 : 23–36