PAPER – OPEN ACCESS

Utilizing Data Science for Credit Decision-Making to Minimize Non-Performing Loan (Case Study PT Bank Central Asia, Tbk Kisaran Branch)

Author : Jumawal Halim, et al
DOI : 10.32734/lwsa.v9i2.2858
Electronic ISSN : 2654-7066
Print ISSN : 2654-7058

Volume 9 Issue 2 – 2026 TALENTA Conference Series: Local Wisdom, Social, and Arts (LWSA)

LWSA Conference Series 09 (2026)
TALENTA Conference Series
Available online at https://talentaconfseries.usu.ac.id

Utilizing Data Science for Credit Decision-Making to Minimize Non-Performing Loan (Case Study PT Bank Central Asia, Tbk Kisaran Branch)

Jumawal Halim, Nazaruddin, Rulianda Purnomo

Master of Management Study Program, Postgraduate School, Universitas Sumatera Utara, Medan, 20155, Indonesia

jumawalhalim@gmail.com, nazaruddin@usu.ac.id, rulianda_wibowo@usu.ac.id

Abstrak

Salah satu kontributor utama terhadap pendapatan bank adalah kredit; oleh karena itu, risiko kredit merupakan salah satu risiko utama yang harus dimitigasi dengan tepat. Tingkat risiko kredit diukur melalui rasio *Non-Performing Loan* (NPL), yang menjadi indikator penting dalam menilai kesehatan keuangan bank. Setiap penyaluran kredit harus dilakukan secara hati-hati dan penuh kehati-hatian. Salah satu upaya strategis yang dilakukan oleh BCA Kisaran untuk mendukung pertumbuhan laba adalah dengan meningkatkan portofolio kredit, khususnya pada Kredit Usaha Kecil dan Menengah (*Small Medium Enterprise Loans*). Oleh karena itu, setiap keputusan kredit harus dibuat secara prudent atau berhati-hati. Peneliti menggunakan *Data Science*, yang berfokus pada analisis data yang tersedia, terutama data kuantitatif, untuk menemukan pola-pola tersembunyi yang dapat mendukung pengambilan keputusan strategis. Analisis data dalam penelitian ini akan menggunakan metode *Machine Learning* dengan algoritma *Random Forest* untuk mempercepat identifikasi calon nasabah potensial serta meningkatkan efisiensi proses seleksi. Dengan demikian, setiap keputusan kredit dapat dibuat secara lebih hati-hati guna meminimalkan *Non-Performing Loan*, khususnya pada Kredit Usaha Kecil dan Menengah.

Kata Kunci*: Data Science; Random Forest; Non-Performing Loan; Machine Learning.*

Abstract

One of the main contributors to a bank's revenue is credit; therefore, credit risk is one of the key risks that must be appropriately mitigated. The level of credit risk is measured by the Non-Performing Loan ratio, which is a critical indicator of a bank's financial health. Every credit disbursement must be carried out with prudence. One of the strategic efforts undertaken by BCA Kisaran to support profit growth is to increase its credit portfolio, particularly in the Small Medium Enterprise Loans and credit decision must be made prudently. Researcher uses Data Science, which focuses on analyzing existing data primarily quantitative data to uncover hidden patterns that can inform strategic decisions. The data analysis in this research will employ Machine Learning Methods with algorithm Random Forest to accelerate the identification of potential customer and the efficiency of the selection process so that each credit decision is made prudently in order to minimize Non-Performing Loan Especially Small Medium Enterprise Loans.

Keywords: Data Science; Random Forest; Non-Performing Loan; Machine Learning.

1. Introduction

Banks are generally institutions that collect funds from society and redistribute them to the society in loans in order to improve the socety welfare [14]. Every bank must release credit to customers in need because credit is the most significant component in providing or contributing profit to the bank. In this thesis, the researcher focuses more on examining the SME (Small Medium Enterprise) Credit segment. Every credit given is inseparable from various risks that can threaten the health of a bank. The measure of credit risk that is a parameter for the level of health of a bank is Non-Performing Loan (NPL). Banks will have potential endanger the continuity of their business if the ratio of non-performing loans (NPL) is more than 5% of total credit (PBI No.15/2/PBI/2013). The higher the NPL level, the more unprofessional the bank is in managing its credit, while also indicating that the level of risk for providing credit to the bank is relatively high in line with the high NPL. The determining factors for a bank's profitability can be seen from the bank's internal factors, namely capital adequacy, operational efficiency, liquidity, and asset size. Internal bank information, especially related to financial conditions, is reflected in the bank's financial statements, from which a number of financial ratios can be calculated, which are one of the tools used by decision makers for both internal and external parties in determining subsequent policies.

Selection and peer-review under responsibility of The 2nd International Joint Conference on Digitalisation of Sustainable Social Sciences (IJCDSS) 2025
p-ISSN: Jumawal Halim, e-ISSN: et al, DOI: 719

Number of Non-Performing loan is used as an indicator for bank's health and a measure of bank asset quality. Research reveals that poor credit quality, higher operational inefficiency costs, and the size of the banking industry can significantly increase NPL, while larger banks with higher profitability reduce NPL. Therefore, in order to minimize risk and optimize bank performance, when releasing credit, banks need to analyze the debtor's ability to repay their obligations, which will ultimately affect the bank's profitability level. Credit that is distributed must be managed properly, this is to avoid the risk of credit losses because credit that is not managed properly will cause non-performing loans which can have a negative impact on profitability. Credit that is not managed properly will cause non-performing loans (Non-Performing Loans) to continue to increase, resulting in a decrease in bank interest income and a decrease in principal credit, which in turn will cause the bank to suffer losses, and it is not impossible that it will eventually go bankrupt.

As of February 2025, the number of BCA SME Debtors at the Kisaran Branch was 366 debtors with an NPL (SME Credit) of 0.28% and has the potential to increase to 0.42%. Compared to 2024, there was an increase in NPL of 0.28% which was previously 0%, so BCA Kisaran branch have to decrease the number of Non-Performing loan. One way to reduce the Non-Performing Loan figure other than through credit settlement is to increase the credit portfolio with credit decisions given must be prudent and appropriate so that the Non-Performing Loan figure can be minimized properly. One way to get potential financing is by utilizing customer data at BCA. Researchers will explore the potential in the data by utilizing Data Science, which aims to accelerate the identification process to increase the effectiveness of the credit decision-making process to minimize the occurrence of problematic credit (Non-Performing Loan) for the credit given. The customers studied are individual customers or individuals who have the potential to be given productive credit financing. The technique that will be used by the researcher is to find a specific or particular pattern of customer transactions that are likely to be potential for SME credit financing by providing prudent credit decisions in order to minimize NPL. The purpose of the above research is to find an effective, fast and accurate method or way to accelerate data identification and increase the efficiency of the prudent credit decision-making process in order to minimize Non-Performing Loans, especially SME credit at BCA Kisaran Branch which supports the increase in the branch's credit portfolio. The expected benefit of this research is to provide the company with strategic considerations that can be used by BCA Kisaran Branch to minimize the occurrence of problematic loans (Non-Performing Loans) in relation to the credit decisions made.

Table 1. Comparison of Average Realization vs. Target of Productive AO (2022–2024)

Year	Num AO	Average Realization	Average Target	Achievement Percentage
2022	65	Rp 8,61 M	Rp 14,64 M	58,81%
2023	64	Rp 7,37 M	Rp 14,54 M	50,69%
2024	41	Rp 8,15 M	Rp 16,92 M	48,16%

Source: Data AO PT. Bank Sumut (2025)

2. Literature Review

2.1. Definition and Functions of Banks

Banks are generally institutions that collect capital in the form of funds from the public in the form of savings and redistribute them to the public in the form of loans and/or other forms in order to improve the welfare of the community [14]. Banks are one of the sources of capital needed by the community to start or develop a business [8]. Sufficient capital is important to start a businessand banks can provide capital loans as an alternative. Banks also have an important role in development by providing credit to the community, which is very much needed to encourage economic growth in a country [25].

2.2. Definition of Credit

According to Law No.10 of 1998 definiton of credit is provision of money based on a loan agreement between a bank and a borrower which requires the borrower to repay the debt by paying interest during a certain agreed period. Credit provision is an activity carried out by banks or other financial institutions in providing loans to customers or other parties who need funds for certain purposes [28]. According to Hermansyah in [12], approval of a credit application is carried out based on the 5C Formula, namely Character, Capacity, Capital, Collateraland Condition of Economy. The assessment carried out by the bank is in accordance with the principle of prudence in order to reduce the risk that will arise in the future. The manifestation of the implementation of the principle of prudence in the context of providing credit is reflected in the criteria called "The Five C's Principle of Credit Analysis" [22].

2.3. Definition Of Small and Medium Enterprise

Law No. 20/2008 defines Small Medium Enterprise based on net assets and annual sales results. Small Business, have annual sales about three hundred million rupiah and up to a maximum two billion five hundred million rupiah. Medium Enterprises, have annual sales about two billion five hundred million rupiah and up to a maximum fifty billion rupiah.

2.4. Non-Performing Loan

Choudhury et al in [7] stated that non-performing loans are not a "uniclass" concept but a "multiclass" concept, meaning that NPLs can be classified into various types usually based on the "duration" of the loan. NPLs are seen as a common by-product of the financial crisis: NPLs are not the main product of the lending function but rather an unintentional occurrence of the lending process, which has excellent potential to exacerbate the severity and duration of the financial crisis and complicate macroeconomic management [31].

Based on BI Circular Letter Number 17/11/PBI/2015, the criteria for assessing the health level of the Non-Performing Loan (NPL) ratio are are explained in Table 1.

Table 1. NPL Ratio Health Criteria

Ratio	Predicate
NPL < 5%	Healthy
NPL > 5%	Unhealthy

Source: https://www.bi.go.id

If the NPL level exceeds 5%, the bank's health is disrupted. That's why the bank should maintain it NPL number as best as possible.

2.5. Data Science

Data Science is a concept to unify statistics, data analysis, informaticsand related methods to understand and analyze actual phenomena with data [11]. Data analysis involves the process of collecting, processingand evaluating data to understand and draw conclusions from the information found in the data. The goal is to identify patterns, trendsand relationships between variables in the data to solve problems or make informed decisions. Data analysis can involve a variety of techniques such as statistics, mathematicsand computer engineering.

2.6. Data Mining

According to Davies (2004), data mining is mining or new information by searching for certain patterns or rules from a very large amount of data. Data mining is also often referred to as knowledge discovery in database (KDD), which is an activity that includes collecting, using data, history to find regularities, patterns or relationships in large data sets [22].

2.7. Random Forest

Random forest is one of the most popular machine learning algorithms. Random forest consists of many decision treesand the resulting class output is the mode of the class output through the individual trees. Many of these resulting trees create a good forest for classification problems like this and other tasks like functional regression as described above by creating multiple trees when traininging and dropping classes or predicting the average of a particular tree. Many deep decision trees are traininged on separate batches of the same dataset and averaged with a target of variance reduction invalid source specified. The tuning parameters for Random Forest are criterion, maximum depth, maximum features, and number of estimators. Criterion is the value used to process the model. Maximum features are the best split for the random forest model. Numerical estimators are used to create leaf nodes from parent nodesand maximum depth is the number of instance boundaries at each node.

3. Methods

The type of research used is Mixed Method Research, which is a mixed research which is a combination of qualitative and quantitative. The first method is qualitative analysis which aims to provide an overview of the data to be studied. Furthermore, the second method is machine learning with the Random forest algorithm, which aims to classify the Potential for Non-Performing Loans (Based on color classification) in the existing debtor data of BCA KCU Kisaran branch. Furthermore, a prediction of the possibility of a credit decision that can be given prudently in order to minimize NPL for the provision of credit obtained from K1 customer data, who are not yet debtors. The population in this study is All Individual Customers focused on K1 BCA Kisaran Branch who have the potential to be given productive credit financing (SME). Fixed variables that will be used in this study consist of Credit Mutation, Average Credit Balance, Age, Business Field, and Credit Mutation Ratio to Average Credit Balance. At the same time, the target variable is the NPL Risk Colormap. Based on the existing data, the number of customers is 983 customers so that all customers are used as research samples so that this research is conducted using the Census/Saturated Sampling method. The data used in the study is secondary data available at the BCA Kisaran Branch. A summary scheme of the research stages can be seen in Figure 1 below.

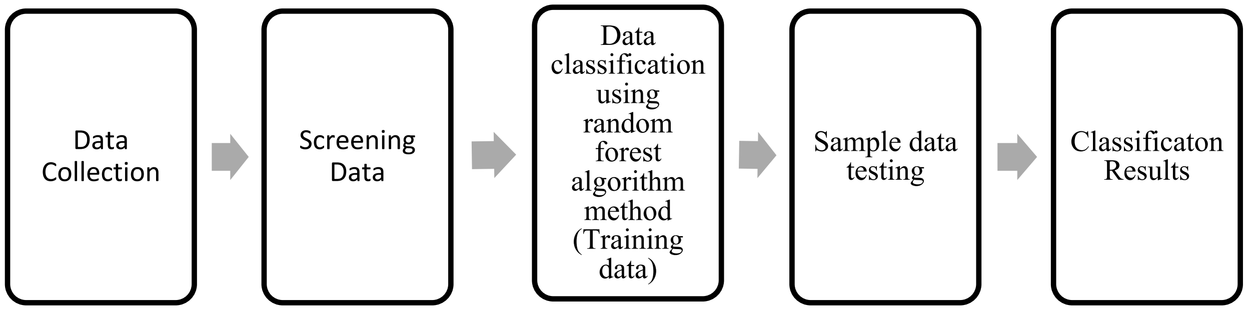

Figure 1. Research Stages

4. Results And Discussion

The research began with data collection through the specta and DIAN BCA applications in the form of K1 customer data and BCA debtor data in the North Sumatra area covering Medan City - Tebing Tinggi - Kisaran - Tanjung Balai - Rantau Prapat. The data collected will be cleaned and screened using the Microsoft Excel application. This is necessary to correct some data errors, clean confidential data, and add variables needed in the study. From the cleaning results, 686 debtor data were obtained which were used as training data to be used by the random forest algorithm. After the training data and testing data have been cleaned, the next step is to process data science using the Orange Data Mining application with the aim of studying the patterns hidden in the data. Researchers make a descriptive analysis of each predetermined customer data variable from the previous withdrawal data that has gone through the cleaning and screening process. Then the researcher will review the data for each variable used in the orange application. Average balance, credit mutation, mutation ratio to average balance, age, and type of business are used as role features in the traininging data. While the NPL risk Colormap which reflects the potential Non-Performing Loan from the traininging data is used as a target in machine learning. The NPL risk colormap is categorized into three colors as seen in Table 2.

Table 2. Non-Performing Loan Risk Colormap

Colormap	Criteria
Green	Usage of Non-installment Credit facilities up to 50% and Installment Facilities with collectibility 1 (Fluent or Good)
Yellow	Usage of Non-installment Credit facilities 51 to 95% and Installment Facilities with collectibility 1 (Fluent or Good)
Red	Usage of Non-installment Credit facilities above 95% and Installment Facilities with collectibility other than 1

Descriptive analysis is done using the Scatter plot tool. Figure 2 is a visualization of the Scatter Plot results from the traininging data.

From the Inlier data visualization above, it can be seen that there is a relationship between the Non-Performing Loan risk colormap and the average credit balance and credit mutations. The collection of blue dots or debtors with non-installment facility credit usage of up to 50% and installment facilities with a collectibility of 1 (fluent or good) dominates the area of significant average balances and large credit mutations. While the collection of green dots or debtors with non-installment facility credit usage above 95% and installment facilities with a collectibility other than fluent or good dominate the area of small average balances and small credit mutations. From this descriptive analysis, it can be concluded that the credit risk colormap can be influenced by the average balance and credit mutations. The scatter plot results also show that there are some data dominate are too far from the data group. Some data that is far from the group needs to be excluded so that the machine learning algorithm used can work optimally. Descriptive analysis with scatter plots shows a relationship between variables such as average balance, account mutations, ratio of mutations to average balance, age, and business sector with the Non-Performing Loan risk colormap created from existing debtor data. This can be used as a basis for making this variable a factor that influences Non-Performing Loan risk. After the outlier data is removed, the inlier data can then be used in the machine learning algorithm to training the existing debtor data that has been determined previously. 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 3 is the result of the Pythagorean Forest.

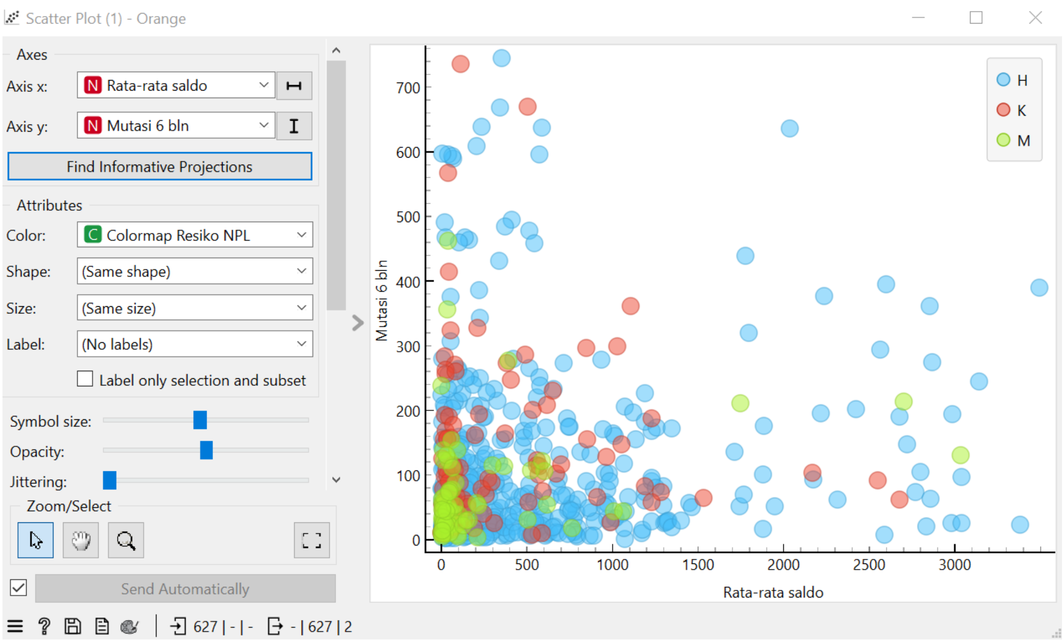

Figure 2. Inlier Scatter Plot Data Visualization

Figure 3. Phytagorean Forest Result

The performance Random Forest's prediction is measured using several models including AUC area under curve, CA Accuracy, F1 Score, Precision, Recall and MCC. Figure 4 is the result of the Random Forest accuracy test. The measurement results of several models show relatively good prediction performance approaching approximately value = 1.

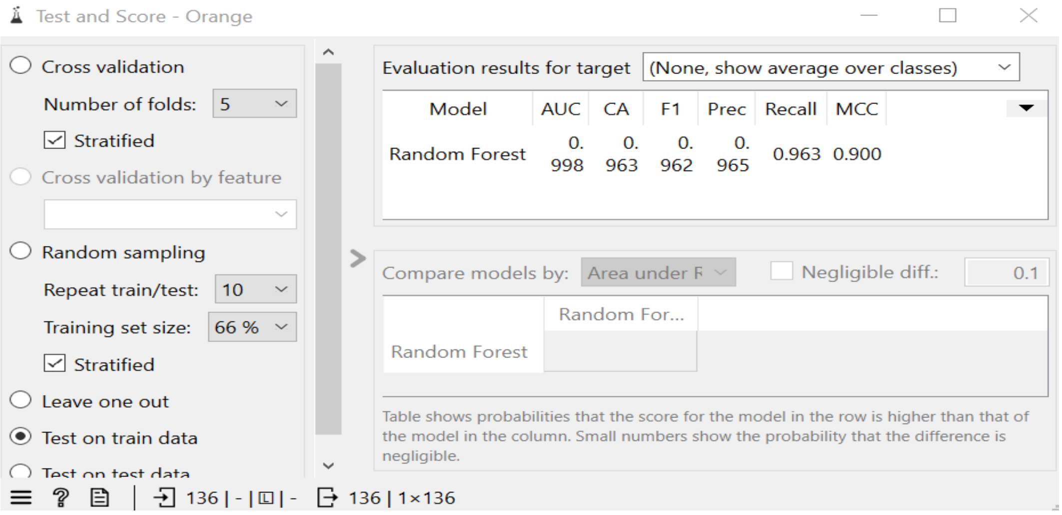

Figure 4. Random Forest Accuracy Test Results

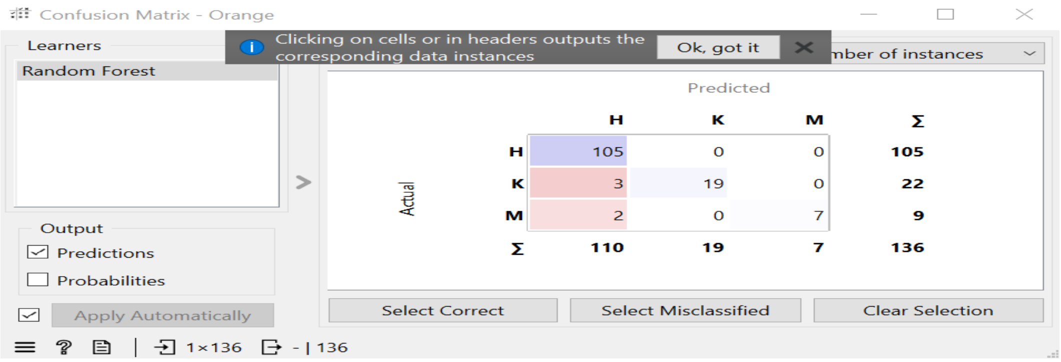

Figure 5. Confusion Matrix

From the confusion matrix data shown in Figure 5 above, the conclusion of the random forest algorithm prediction results is obtained from a total of 627 training data. The Green NPL risk colormap (H) has a total of 460 correct predictions. The Yellow colormap (K) has a total of 87 correct predictions from a total of 104 data and the Red colormap (M) has a total of 50 correct predictions from a total of 63 data.

The random forest algorithm that has been tested and studied 627 training data will be used as a prediction tool to predict K1 non-debtor data that has been screened and cleaned. Of the 747 K1 data that were cleaned and screened, 625 data can be continued to the Prediction stage. Figure 6 is the flow of the Random Forest algorithm carried out in this study.

Flow machine learning starts from inputting training data (Logo “File” in the flow) which has previously gone through a screening and cleaning process from existing debtors, then the data is reduced using outliers to remove data outside the range. The training data is tested by the Random Forest algorithm. Then the K1 data is input (Logo ‘File1’ in the flow), which is 625 data. The K1 data is predicted by the random forest algorithm and produces the result data.

Prediction data from 625 K1 customers, if credit is given to all customers, the following conclusions are obtained:

- 612 customers in the Green colormap category (H), which means they have a low risk Non-Performing Loans
- 10 customers in the Yellow colormap category (K), which means they have a Medium/Moderate Non-Performing Loan risk
- 3 customers in the Red colormap category (M), which means they have a High risk Non-Performing Loans

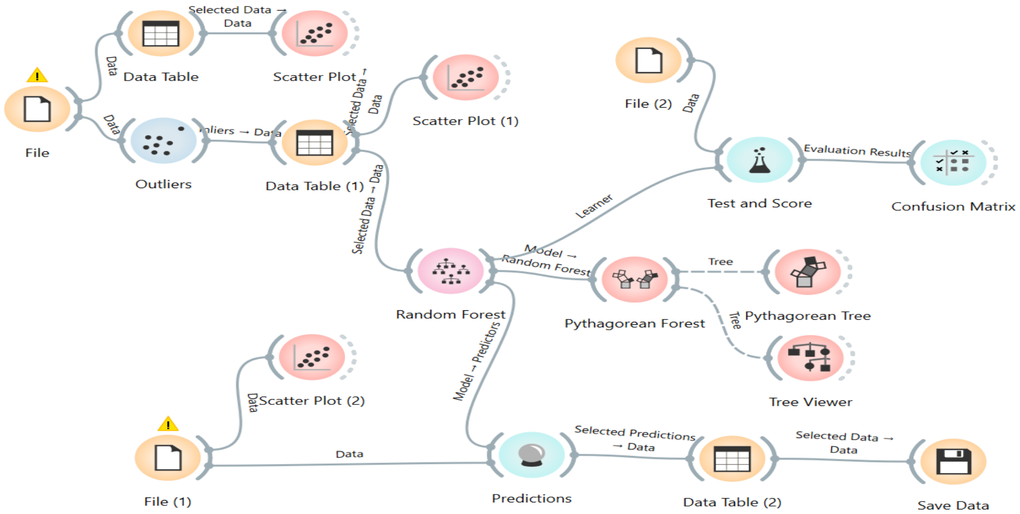

Figure 6. Random Forest Algorithm Flow

From the results of the research above with the research data used is limited to customers domiciled in North Sumatra which includes Medan - Tebing Tinggi - Kisaran - Rantau Prapat with the Random Forest algorithm Data Science method can be used to predict BCA Kisaran Branch focus customers who have the potential to be given credit with low Non-Performing Loan risk potential so that the identification and prospecting process can be carried out efficiently and quickly by providing the right credit decisions to increase the branch credit portfolio with well-controlled Non-Performing Loan (low). Data Science with the Random Forest algorithm can be used for comprehensive data prediction because it accommodates the use of various variables, so that more accurate results are obtained. Random Forest has only been used for data classification, but it turns out that it can be used to predict the results of data with reasonably good accuracy. For this method to be implemented more effectively and accurately, proper and comprehensive database management with continuous and periodic data updates is necessary. The results of this research implementation may differ for customers in other regions, cities, or provinces, depending on the demographics of the population and businesses in those regions, as different demographics will produce different prediction results. Suggestions for further research include the addition of other variables, such as the 5C analysis: character, capital, 7P, and 3R.

5. Conclusion

Based on the results of the research that has been conducted, it was concluded that the use of data science can be used to accelerate the selection process and identification of data on many customers who have the potential to be given credit effectively and efficiently by making prudent credit decisions to minimize Non-Performing Loans (NPL) which is also a strategy to increase the credit portfolio

References

[1] Aliero, I. H., & Yusuf, M. M. (2015). Analysis of impact of credit on the performance of SMEs in Sokoto metropolis of Sokoto State of Nigeria. International Journal of Small Business and Entrepreneurship Research, 3(4), 22-31.
[2] Alpaydin, E. 2010. *Introduction to Machine Learning*. 2nd Edition, The MIT Press, Cambridge.
[3] Andrianto. (2020). Manajemen Kredit. In Q. Media (Ed.), CV. Penerbit Qiara Media. CV. Penerbit Qiara Media.
[4] Atieno, R. (2001). *Formal and informal institutions' lending policies and access to credit by small-scale enterprises in Kenya: An empirical assessment*. AERC.
[5] Avendano, R., Boehm, N., & Calza, E. (2013). Why scarce small and medium enterprise financing hinders growth in Latin America. *Journal of the role for public policies*, *2*, 112-114.
[6] Berry, M. and Linoff, G. 2004. *Data Mining Techniques for Marking, Sales and Customer Support (2nd ed.).* Wiley, New York.
[7] Choudhury, T. A., & Adhikary, B. K. (2002, January). Loan classification, provisioning requirement and recovery strategies: A comparative study on Bangladesh and India. In *Seminar Paper, Bangladesh Institute of Bank Management, January* (pp. 21-54).
[8] Effendi, S., & Harahap, B. (2020). Pengaruh Sistem Akuntansi Dan Pengendalian Internal Terhadap Pemberian Kredit Oleh Pt Bank Ocbc Nisp Tbk Kota Batam. Jurnal Akuntansi Barelang, 5(1), 37–46. https://doi.org/10.33884/jab.v5i1.2648
[9] Greuning, H. Bratanovic, S. B. 2003. Analyzing and Managing Banking Risk. A Framework for Assessing Corporate Governance and Financial Risks. 2nd edition. The World Bank, Washington, D.C.
[10] Han, J., Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan kaufmann*, *340*(1), 94104-103205.
[11] 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, Classificationand Related Methods. Studies in Classification, Data Analysis and Knowledge Organization*. Springer Japan. pp. 40-51.
[12] Hermansyah. 2008, Hukum Perbankan Nasional Indonesia. Prenada, Yogyakarta
[13] Kakuru, J. 2008, *The Supply-Demand Factors Interface and Credit Flow to Small and Micro Enterprises (SMEs).* Stirling, University of Stirling. http://hdl.handle.net/1893/493
[14] Kashmir. 2017. Bank dan Lembaga Keuangan Lainnya. *Cetakan Kedelapan Belas*. Jakarta: Rajawali.

[15] Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining.* John Wiley, New York
[16] Liedholm, C., McPherson, M., & Chuta, E. (1994). Small enterprise employment growth in rural Africa. *American Journal of Agricultural Economics*, *76*(5), 1177-1182. https://doi.org/10.2307/1243413
[17] 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.
[18] MacLennan, J., Tang, Z., dan Crivat, B. (2009). *Data mining with Microsoft SQL Server 2008*. Indianapolis, IN: Wiley Pub.
[19] Mahmoeddin, 2010, Melacak Kredit Bermasalah. Cetakan Pertama. Jakarta: Pustaka Sinar Harapan
[20] Moh. Tjoekam. 1999, Perkreditan Bisnis Inti Bank Komersial. Gramedia, Jakarta
[21] Olson, D. L. 2007. *Data Mining in Business Services*. Service Business, 181-193
[22] Santosa, Budi. 2007, *Data Mining* : Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Graha Ilmu, Yogyakarta
[23] Sembiring, M. T., & Hasibuan, C. F. (2021). Data Science: Strategi UMKM dalam Pengambilan Keputusan.
[24] Sembiring, S. (2007). Arti Penting Jaminan dalam Pemberian Kredit dalam Transaksi Bisnis Perbankan. *Gloria Juris*, *7*(1).
[25] Sumartik. (2018). Buku Ajar Manajemen Perbankan. In S. B. Sartika & M. T. Multazam (Eds.), UMSIDA Press.
[26] Tushabomwe-Kazooba, C. (2006). Causes of small business failure in Uganda: A case study from Bushenyi and Mbarara towns. *African studies quarterly*, *8*(4), 27-35. https://journals.flvc.org/ASQ/article/view/136329/140886
[27] Ulwan, N. (2016). Pattern Recognition Pada Unstructured Data Teks Menggunakan Support Vector Machine Dan Association. Yogyakarta: Universitas Islam Indonesia.
[28] Ulfah, L., & Massora, A. (2021). Pengaruh Sistem Pengendalian Internal dan Sistem Informasi Akuntansi terhadap Pemberian Kredit di BPR Harapan Saudara. Jurnal Akuntansi FE-UB, 15(2),93–109. https://ejournal.borobudur.ac.id/index.php/akuntansi/article/viewFile/918/829
[29] Van Stel, A. J. & Storey, D.J. (2004): The link between firm births and job creation: Is there an up as tree effect? Regional Studies, 38(8).
[30] Walpole, R. (1992). Pengantar Statistika Edisi ke-3. Gramedia Pustaka.
[31] Woo & David. 2000, "Two Approaches to Resolving Non-Performing Assets during Financial Crisis." IMF working paper 00/33, March: 2-5