Eduvest – Journal of Universal Studies
Volume 5 Number 2, February, 2025
p- ISSN 2775-3735- e-ISSN 2775-3727

Data Prediction Of Receivables In 2021-2023 At Bank Syariah Indonesia Tbk With Regression And Clustering Methods

Putra Pribowo
Universitas Lampung, Indonesia
Email: putrapribowo@indotechpren.org

ABSTRACT

This study aims to predict receivables data at Bank Syariah Indonesia Tbk for the 2021-2023 period using regression and clustering methods. Data analysis methods such as regression and clustering have been used to predict credit risk and receivables payment behavior. A linear regression model is applied to predict the future value of different types of receivables (Murabahah, Istishna, Multijasa, Qardh, Serent), while K-Means clustering is used to group data based on five main variables. The results of the analysis show that the linear regression model is able to predict future values with quite good accuracy, shown by the compatibility between the actual value and the predicted value. K-Means clustering produces three fairly good clusters, with a silhouette score of 0.51, which indicates adequate cluster quality. Visualization of the results of the analysis shows the distribution and patterns in the data, providing insight into the relationships between different types of receivables. This research provides a deeper understanding of the structure of receivables data and aids in decision-making based on future predictions and data grouping.

KEYWORDS Regression, Clustering, Receivables, BSI

INTRODUCTION

Islamic banking has shown significant growth in recent decades, especially in countries with large Muslim populations such as Indonesia. Bank Syariah Indonesia Tbk (BSI) is one of the financial institutions that plays an important role in providing sharia-based banking services. One of the critical aspects of bank operations is receivables management, which includes monitoring and predicting receivables to ensure liquidity and minimize credit risk. Basically, borrowing, called bank receivables, is one of the alternatives that can be done by several people in the process of improving the economy (Nguyen et al., 2021). The lending process is carried out with various objectives, one of which is to improve the economy individually in the form of entrepreneurial activities (Mahdaoui & SBAI, 2021). Institutions that can provide lending services to customers and the general public are banking and non-banking institutions (Abedini et al., 2016). The performance results are obtained based on a review analysis conducted on data on the financing that has been provided (Mustaidah & Fauzan, 2021).

Machine learning data analysis methods, such as regression and clustering, have been widely used in various studies to predict credit risk and payment behavior of receivables. According to (Jiang et al., 2019), machine learning methods can

How to cite: Putra Pribowo. (2025). Data Prediction Of Receivables In 2021-2023 At Bank Syariah Indonesia Tbk With Regression And Clustering Methods. *Journal Eduvest*. *5*(2): 1830-1845
E-ISSN: 2775-3727

provide high accuracy in predicting bank loan defaults using various historical variables. This research shows that techniques such as logistic regression and decision trees can be effective in identifying default patterns.

According to (Chaudhuri & Ghosh, 2020) conducted a comparative study of various machine learning techniques for loan default prediction. The results of their study show that regression methods, such as linear regression and logistic regression, as well as clustering methods can provide significant results in classifying accounts receivable data based on default risk. Their research emphasizes the importance of choosing the right model based on the characteristics of the data and the purpose of the analysis.

(Kim & Min, 2021) in their research on loan repayment prediction using various machine learning algorithms, found that a combination of regression and clustering methods can improve prediction accuracy. They also highlighted the importance of model validation to ensure that the predictions made are reliable and can be used in management decision-making.

(Luo et al., 2018) examined credit risk assessment for small and medium enterprises (SMEs) using clustering analysis and logistic regression. The results of their research show that the clustering approach can help in better segmentation of data, making it easier to identify groups with high and low risk.

(Wang & Liu, 2022) in their research on loan default prediction using clustering and classification algorithms, showed that a combination of these two methods can improve predictive performance. They emphasized the importance of clustering analysis to find hidden patterns in the data, which can then be used to build more accurate prediction models.

Based on these studies, this study aims to predict accounts receivable data at Bank Syariah Indonesia Tbk for the 2021-2023 period using regression and clustering methods. By utilizing historical data and appropriate analysis techniques, it is hoped that significant patterns can be found that help in managing accounts receivable risk, improving the bank's financial performance, and seeing whether the increase from the 2021-2023 period has increased or not.

RESEARCH METHOD

This research adopts study research with quantitative methods that aim to produce a research result based on mathematical calculations. The calculation process is based on the application of *Machine Learning* with the Regression and Clustering approach used. The activities of this research are described in a research framework that can be seen in Figure 1.

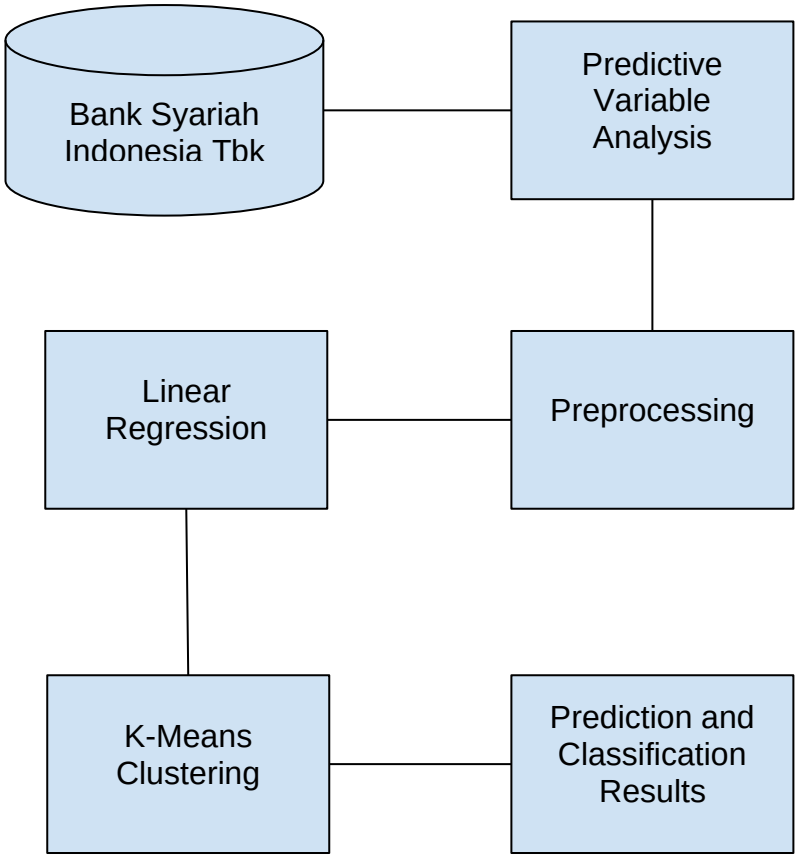

Figure 1. Research Framework

Figure 1 explains that the stages of the research were carried out with 2 methods, namely Linear Regression and Clustering with K-Means which resulted in a Prediction and Classification of Receivables at Bank Syariah Indonesia Tbk. Before the preprocessing stage is carried out, the process begins by analyzing the variables used in building determining the results of the prediction of existing receivables at Bank Syariah Indonesia Tbk. The dataset is sourced from the official website of Bank Syariah Indonesia Tbk which was published in the 2021-2023 period which is used as a research dataset. Datasets taken as variables of Murabahah Receivables, Istishna 'Receivables, Multipurpose Receivables, Qardh Receivables, Lease Receivables can be presented in Table 1, Table 2, Table 3, Table 4, and Table 5.

Table 1. Dataset of Murabahah Receivables.

No	Month	Year	Murabahah Receivables
1	February	2021	IDR 90,762,814.00
2	April	2021	IDR 93,186,337.00
3	May	2021	IDR 93,468,639.00
4	July	2021	IDR 94,251,987.00
5	August	2021	IDR 95,191,126.00
6	October	2021	IDR 97,930,228.00
7	November	2021	IDR 99,481,405.00
8	January	2022	IDR 102,529,531.00
9	February	2022	IDR 103,889,549.00
10	April	2022	IDR 108,869,564.00
11	May	2022	IDR 109,657,642.00
12	July	2022	IDR 114,400,066.00
13	August	2022	IDR 116,572,568.00
14	October	2022	IDR 120,777,445.00
15	November	2022	IDR 122,750,609.00
16	February	2023	IDR 125,055,048.00
17	April	2023	IDR 127,655,075.00
18	May	2023	IDR 127,735,328.00
19	July	2023	IDR 130,512,345.00
20	August	2023	IDR 132,088,589.00
21	October	2023	IDR 133,941,781.00
22	November	2023	IDR 134,427,748.00

Based on Table 1 is one of the Murabahah Receivables datasets whose data is used to be studied with Regression and Clustering in producing a prediction.

Table 2. Istishna' Receivables Dataset

No.	Month	Year	Istishna' receivable
1	February	2021	IDR 610.00
2	April	2021	IDR 581.00
3	May	2021	IDR 568.00
4	July	2021	IDR 430.00
5	August	2021	IDR 415.00
6	October	2021	IDR 390.00
7	November	2021	IDR 372.00
8	January	2022	IDR 348.00
9	February	2022	IDR 336.00
10	April	2022	IDR 309.00
11	May	2022	IDR 296.00
12	July	2022	IDR 249.00
13	August	2022	IDR 196.00
14	October	2022	IDR 149.00
15	November	2022	IDR 140.00
16	February	2023	IDR 114.00
17	April	2023	IDR 98.00
18	May	2023	IDR 86.00
19	July	2023	IDR 54.00
20	August	2023	IDR 46.00
21	October	2023	IDR 33.00
22	November	2023	IDR 31.00

Based on Table 2 is the Istishna' Receivables dataset whose data is used to be studied with Regression and Clustering in producing a prediction.

Table 3. Multijasa Receivables Dataset

No.	Month	Year	Multijasa Receivables
1	February	2021	Rp 0
2	April	2021	Rp 0
3	May	2021	Rp 0
4	July	2021	Rp 0
5	August	2021	Rp 0
6	October	2021	Rp 0
7	November	2021	Rp 0
8	January	2022	Rp 0
9	February	2022	Rp 0
10	April	2022	Rp 0
11	May	2022	Rp 0
12	July	2022	Rp 0
13	August	2022	Rp 0
14	October	2022	Rp 0
15	November	2022	Rp 0
16	February	2023	IDR 234,423.00
17	April	2023	IDR 225,588.00
18	May	2023	IDR 222,098.00
19	July	2023	IDR 217,756.00
20	August	2023	IDR 215,377.00
21	October	2023	IDR 213,840.00
22	November	2023	IDR 208,864.00

Based on Table 3 is a Multijasa Receivables dataset whose data is used to be studied with Regression and Clustering in producing a prediction.

Table 4. Qardh Receivables Dataset

No.	Month	Year	Qardh Receivable
1	February	2021	IDR 8,939,049.00
2	April	2021	IDR 9,318,957.00
3	May	2021	IDR 9,568,412.00
4	July	2021	IDR 9,601,610.00
5	August	2021	IDR 8,989,363.00
6	October	2021	IDR 9,534,000.00
7	November	2021	IDR 9,319,900.00
8	January	2022	IDR 8,961,565.00
9	February	2022	IDR 8,954,247.00
10	April	2022	IDR 9,325,164.00
11	May	2022	IDR 9,239,654.00
12	July	2022	IDR 9,040,587.00
13	August	2022	IDR 9,211,831.00
14	October	2022	IDR 9,582,902.00
15	November	2022	IDR 9,575,131.00
16	February	2023	IDR 9,807,220.00
17	April	2023	IDR 9,800,109.00
18	May	2023	IDR 9,838,862.00
19	July	2023	IDR 10,404,096.00
20	August	2023	IDR 10,487,682.00
21	October	2023	IDR 10,981,754.00
22	November	2023	IDR 11,065,204.00

Based on Table 4 is a dataset of Qardh Receivables whose data is used to be studied with Regression and Clustering in producing a prediction.

Table 5. Rent Receivable Dataset

No.	Month	Year	Rent Receivable
1	February	2021	IDR 42,908.00
2	April	2021	IDR 47,546.00
3	May	2021	IDR 57,475.00
4	July	2021	IDR 77,312.00
5	August	2021	IDR 72,640.00
6	October	2021	IDR 8,588.00
7	November	2021	IDR 92,492.00
8	January	2022	IDR 110,002.00
9	February	2022	IDR 122,574.00
10	April	2022	IDR 139,465.00
11	May	2022	IDR 148,816.00
12	July	2022	IDR 163,259.00
13	August	2022	IDR 169,872.00
14	October	2022	IDR 13,312.00
15	November	2022	IDR 13,510.00
16	February	2023	IDR 12,467.00
17	April	2023	IDR 11,653.00
18	May	2023	IDR 10,482.00
19	July	2023	IDR 11,151.00
20	August	2023	IDR 11,155.00
21	October	2023	IDR 11,127.00
22	November	2023	IDR 11,214.00

Based on Table 5 is the Rent Receivable dataset whose data is used to be studied with Regression and Clustering in producing a prediction. Based on Table 1, Table 2, Table 3, Table 4, and Table 5 are variables selected from the research dataset in the prediction and classification process. The data will then be processed in the preprocessing analysis stage to convert data type strings into float data type before going to the Linear Regression process and the K-Means Clustering process to perform clustering on each data used. Conceptually, K-Means is one of the unsupervised learning algorithms used in Machine Learning (Saadat Far et al., 2020), by clustering the same and similar data (Yuan, 2020). The algorithm in K-means is carried out by determining the number of clusters first, or initialized with variable K. Randomly select K initial data as the initial centroid. Calculate the distance of the object to the centroid using Euclidean Distance (Beltran et al., 2020). Calculation of centroid distance from each data by maximizing the Euclidean Distance function. After preprocessing is done, the process will produce predictions and classifications, in the prediction analysis process using the Linear Regression and K-Means Clustering approaches. This method is a method for classifying objects based on data that has the closest distance to the object whose cluster value has been determined (Nalic et al., 2020).

RESULT AND DISCUSSION

Results

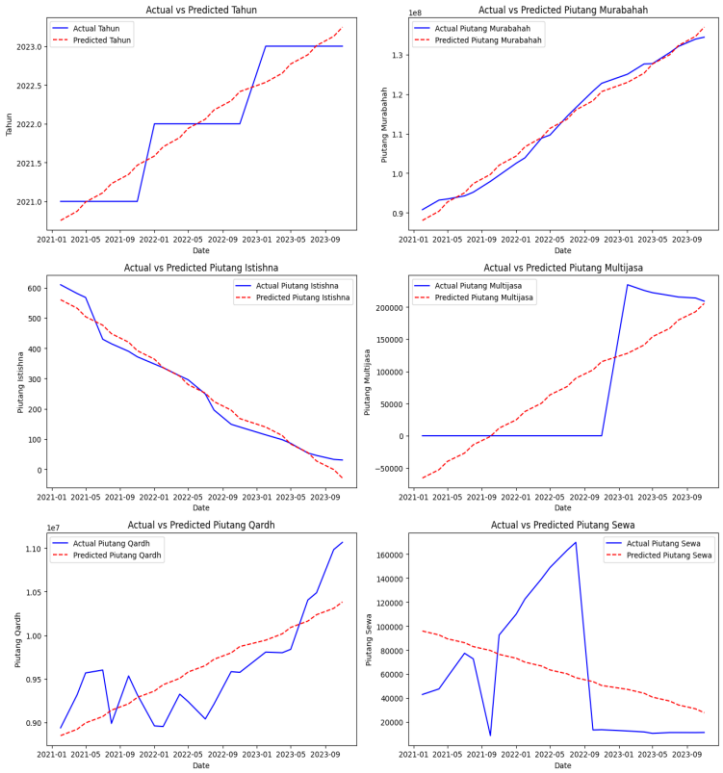

Figure 2. Actual and Predicted Graphs on Receivables (Murabahah, Istishna', Multijasa, Qardh, and Rent)

The conclusion in Figure 2 *Actual VS Predicted* Year graph shows a fairly good match between the actual value and the predicted value and the upward trend is consistently in accordance with historical data, indicating that the linear regression and clustering model is able to predict the year quite well and produce information that there is an increase in a particular month that exceeds the prediction for two increases. The *Actual VS Predicted* Murabahah graph produces a predicted value (red dotted line) almost parallel to the actual value (blue line), indicating the linear regression model and clustering can capture a stable pattern of increase and in accordance with the predicted value. The *Actual VS Predicted* Istishna' graph produces a consistent downward trend value that is quite good at both actual and predicted values, which provides information that it has decreased. The *Actual VS Predicted* Multijasa graph of model prediction shows a consistent increase, while the actual data shows a sudden spike in a certain period. The *Actual VS Predicted* Qardh graph shows a smoother upward trend compared to the more pronounced fluctuations in the actual data. The *Actual VS Predicted* Rent graph shows a downward trend, while the actual values show larger fluctuations. The conclusion that can be drawn from the overall graph is that there are some noticeable differences between the predictions and actual data, especially for Multijasa receivables and Rental receivables, which show high variability in the actual data.

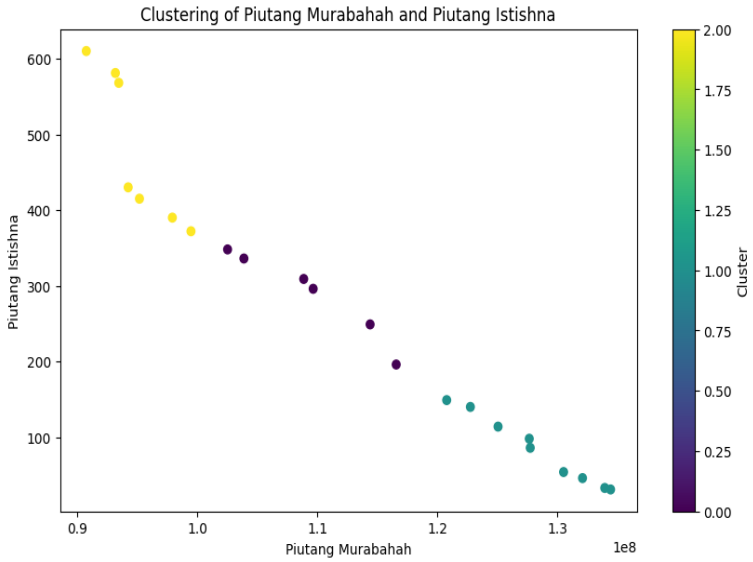

Figure 3. Clustering of Murabahah Receivables and Istishna' Receivables

The conclusion of Figure 3 is:

a. Cluster Identification:

1. Cluster 0 (marked in purple): This cluster has Istishna Receivables values ranging from 200 to 400, with Murabahah Receivables values ranging from 1.0e8 to 1.2e8.

2. Cluster 1 (marked in yellow): This cluster has higher Istishna Receivables values, ranging from 400 to 600, with Murabahah Receivables values ranging from 0.9e8 to 1.1e8.
3. Cluster 2 (marked in light green): This cluster has lower Istishna Receivables values, ranging from 100 to 200, with Murabahah Receivables values ranging from 1.1e8 to 1.3e8.

b. Cluster Separation:

1. Clusters 0 and 1: These two clusters show a decreasing distribution of Istishna Receivables value with increasing Murabahah Receivables value. Cluster 1 has a higher Istishna Receivables value than Cluster 0 at a similar Murabahah Receivables value.
2. Cluster 2: This cluster is clearly separated from the other clusters as it has a lower value of Istishna Receivables, but a higher value of Murabahah Receivables. This indicates a group of receivables with higher values and different characteristics from the other two clusters.

c. Business Insights:

1. Risk Analysis: Cluster 2, with a low value of Istishna Receivables and a high value of Murabahah Receivables, may require more attention in risk management and credit monitoring due to the high value of Murabahah Receivables.

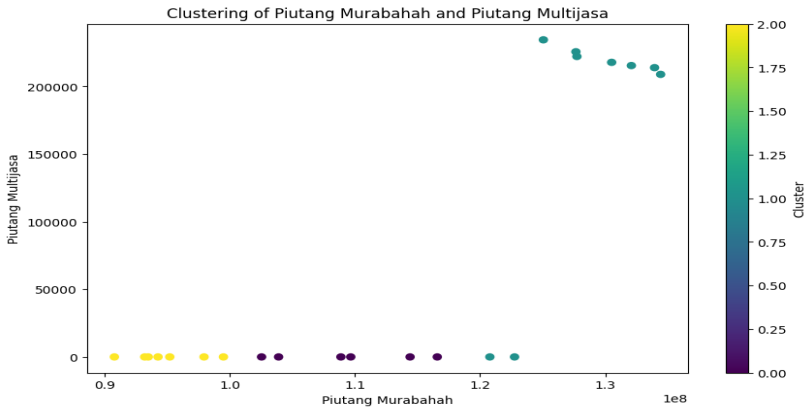

Figure 4. Clustering of Murabahah Receivables and Multijasa Receivables

The conclusion of Figure 4 is:

a. Cluster Identification:

1. Cluster 0 (marked in purple): This cluster has a Multijasa Receivables value close to zero and includes a wide range of Murabahah Receivables values, from the lowest to around 1.2e8.
2. Cluster 1 (marked in yellow): This cluster also has a value of Multijasa Receivables close to zero, but its Murabahah Receivables value is slightly higher than Cluster 0, ranging from 1.0e8 to 1.1e8.
3. Cluster 2 (marked in light green): This cluster has significantly higher values of Multijasa Receivables, from around 1.0e5 to over 2.0e5, with varying values of Murabahah Receivables but mostly over 1.2e8.

b. Cluster Separation:

1. Clusters 0 and 1: These two clusters are characterized by a near-zero value of Multijasa Receivables, but are differentiated by a range of Murabahah Receivables values. This indicates that there are two clusters of receivables with low Multijasa values but differ in terms of Murabahah values.
2. Cluster 2: This cluster is clearly separated from Clusters 0 and 1 as it has a high value of Multiservice Receivables. This shows that this group has unique characteristics with a much higher value of Multijasa than the other two clusters.

c. Business Insights:

1. Risk Analysis: Cluster 2 with a high value of Multijasa Receivables may require more attention in terms of risk management and credit monitoring, as high values may indicate greater exposure.

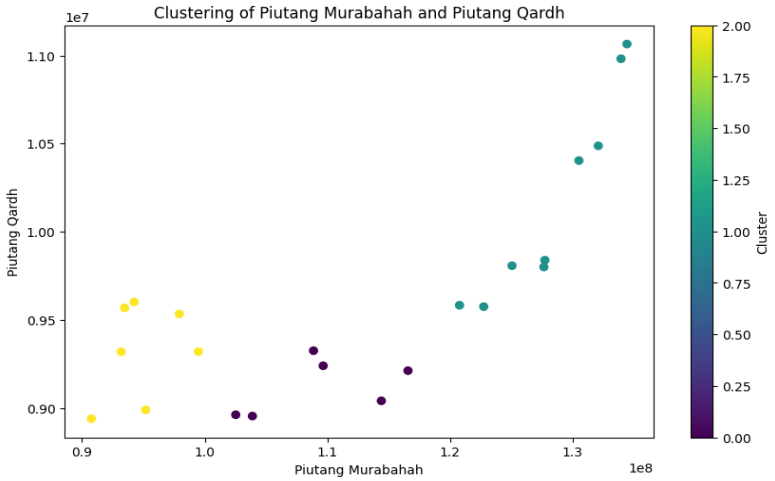

Figure 5. Clustering of Murabahah Receivables and Qardh Receivables

The conclusion of Figure 5 is:

a. Cluster Identification:

1. Cluster 0 (marked in purple): This cluster has low Qardh Receivables values, ranging from 0.9e7 to 0.92e7, with Murabahah Receivables values ranging from 1.0e8 to 1.1e8.
2. Cluster 1 (marked in yellow): This cluster has slightly higher Qardh Receivables values, ranging from 0.92e7 to 0.97e7, with Murabahah Receivables values ranging from 0.9e8 to 1.0e8.
3. Cluster 2 (marked in light green): This cluster has the highest value of Qardh Receivables, ranging from 1.0e7 to 1.1e7, with the value of Murabahah Receivables varying from 1.1e8 to 1.3e8.

b. Cluster Separation:

1. Clusters 0 and 1: These two clusters show a distribution of Qardh Receivables values that increases gradually with an increase in Murabahah

Receivables values. This indicates a relationship between the two types of receivables in the lower value range.

2. Cluster 2: This cluster is clearly separated from the other clusters as it has higher values of Qardh Receivables and Murabahah Receivables. This indicates a group of receivables with higher values and different characteristics from the other two clusters.

c. Business Insights:

1. Risk Analysis: Cluster 2 with high values of Qardh Receivables and Murabahah Receivables may require more attention in risk management and credit monitoring as the high value of receivables may indicate greater exposure.

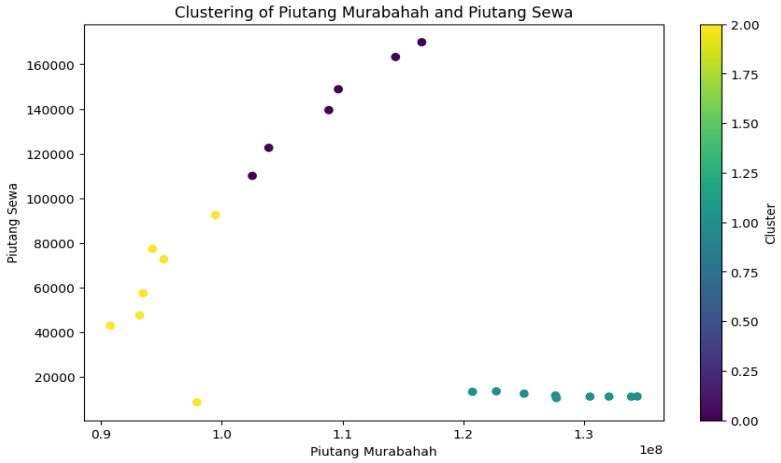

Figure 6. Clustering of Murabahah Receivables and Lease Receivables

The conclusion of Figure 6 is:

a. Cluster Identification:

1. Cluster 0 (marked in purple): This cluster has the highest Lease Receivables value, ranging from 120,000 to 160,000, with Murabaha Receivables value varying from 1.0e8 to 1.2e8.
2. Cluster 1 (marked in yellow): This cluster includes moderate Rental Receivables values, ranging from 20,000 to 100,000, with Murabaha Receivables values ranging from 0.9e8 to 1.1e8.
3. Cluster 2 (marked in light green): This cluster has low Lease Receivables values, close to zero, but Murabaha Receivables values vary between 1.2e8 to 1.3e8.

b. Cluster Separation:

1. Clusters 0 and 1: These clusters show an increasing distribution of Lease Receivables values as Murabaha Receivables values increase. This indicates a positive relationship between the two types of receivables within these clusters.
2. Cluster 2: This cluster is clearly separated from the other clusters as it has a low value of Lease Receivables. However, the value of Murabaha

Receivables in this cluster is higher compared to the other clusters, showing the unique characteristics of this cluster.

c. Business Insights:

1. Risk Analysis: Cluster 0, with high Rental Receivables values, may require more attention in risk management and credit monitoring, as the high values may indicate greater exposure.

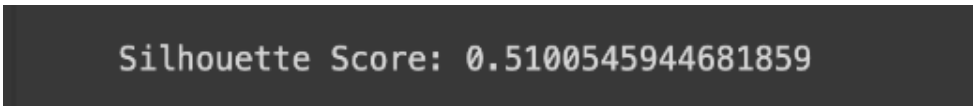

Figure 7. Prediction Silhouette Score Results

Figure 7 provides a conclusion that the prediction results on Receivables at Bank Syariah Indonesia Tbk using linear regression and K-Means Clustering methods are quite good because they produce a silhouette score value of 0.51 which is close to one and away from negative numbers which means it is quite good.

Discussion

Alternative lending is carried out by several people with the aim of restoring the economy and is called receivables in transactions that occur at Bank Syariah Indonesia Tbk, in research can see the results of receivables for the 2021-2023 period are good or not good by displaying a prediction result data from various types of receivables variables and based on the measurements taken, the prediction output provides good results to find out information that the 2021-2023 period is good enough. After the prediction and classification analysis process is carried out using the Linear Regression and K-Means Clustering methods combined in the implementation on the accounts receivable variable at Bank Syariah Indonesia Tbk. The analysis results produced a Silhouette Score value with a value of 0.51. Based on the analysis conducted, Linear Regression and K-Means Clustering are able to provide fairly good prediction output results in the accounts receivable variable. Prediction with Machine Learning using the Linear Regression and K-Means Clustering approaches has been able to provide fairly good prediction results on the accounts receivable variable.

CONCLUSION

The prediction results show that the linear regression and clustering model is able to predict future values for various types of "Receivables" (Murabahah, Istishna, Multijasa, Qardh, Sewa) reasonably well, based on the available historical data. The resulting plots show a good fit between the actual and predicted values, indicating that the model can capture trends with fairly good accuracy. Meanwhile, clustering using K-Means with five main variables produced three good clusters, with a silhouette score of 0.51. This score indicates that the clusters formed are of fairly good quality but can still be improved. The scatter plot visualization shows the distribution and patterns in the data, providing insight into the relationships between different types of receivables. Overall, these predictions provide a deeper understanding of the structure of receivables data and aid in decision-making based on future projections and data clustering.

REFERENCES

Abedini, M., Ahmadzadeh, F., & Noorossana, R. (2016). Customer credit scoring using a hybrid data mining approach. *Kybernetes*.

Beltran, B., Vilariño, D., Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A., & Pinto, D. (2020). K-means based method for overlapping document clustering. Journal of Intelligent and Fuzzy Systems, 39(2), 2127–2135. https://doi.org/10.3233/JIFS-179878.

Chaudhuri, A., & Ghosh, D. (2020). Prediction of Loan Default: A Comparative Study of Machine Learning Techniques. Information Sciences, 508, 66-89.

Jiang, D., Zhang, J., Wang, Z., Feng, C., Jiao, K., & Xu, R. (2020). A Prediction Model of Blast Furnace Slag Viscosity Based on Principal Component Analysis and K-Nearest Neighbor Regression. JOM, 72(11), 3908–3916. https://doi.org/10.1007/s11837-020-04360-9.

Kim, M., & Min, J. (2021). Prediction of Loan Repayment using Various Machine Learning Algorithms. Journal of Forecasting, 40(3), 535-551.

Luo, L., Liu, Q., & Deng, S. (2018). Credit Risk Assessment for SMEs: A Cluster Analysis and Logistic Regression Approach. Economic Modelling, 70, 350-358

Mahdaoui, A., & SBAI, E. H. (2021). 3D point cloud simplification based on the clustering algorithm and introducing the Shannon's entropy. Thirteenth International Conference on Machine Vision, 11605, 87. https://doi.org/10.1117/12.2588384.

Mustaidah, E., & Fauzan, A. (2021). Analisis Pembiayaan Bermasalah Pada Unit Pengelola Kegiatan (UPK) Bangkit Mandiri Kecamatan Lebakwangi Kuningan. BanKu: Jurnal Perbankan Dan Keuangan, 2(2), 68–77. https://doi.org/10.37058/banku.v2i2.3572.

Nalic, J., Martinovic, G., & zagar, D. (2020). New Hybrid Data Mining Model For Credit Scoring Based On Feature Selection Algorithm And Ensemble Classifiers. Advanced Engineering Informatics, 45, 101130. https://doi.org/10.1016/j.aei.2020.101130.

Nguyen, X. V. C., Duong, L. M., & Nguyen, H. S. (2021). Appliance Classification On Low-Cost Smart Power Outlets Based On Frequency Feature Of Electric Current. Proceeding - 2021 26th IEEE Asia-Pacific Conference on Communications, APCC 2021, 128–133. https://doi.org/10.1109/APCC49754.2021.9609819.

Saadat far, H., Khosravi, S., Joloudari, J. H., Mousavi, A., & Shamshir Band, S. (2020). A New K-Nearest Neighbors Classifier For Big Data Based On Efficient Data Pruning. Mathematics, 8(2), 286. https://doi.org/10.3390/math8020286.

Wang, S., & Liu, C. (2022). Loan Default Prediction Using Clustering and Classification Algorithms. IEEE Access, 10, 58319-58328.

Yuan, C. Z. (2020). Determining Optimal Lag Time Selection Function with Novel Machine Learning Strategies for Better Agricultural Commodity Prices Forecasting in Malaysia. In ACM International Conference Proceeding Series (pp.37–42). https://doi.org/10.1145/3417473.3417480.