ISSN 1533-9211

DOI: 10.5281/zenodo.10049842

MACHINE LEARNING PREDICTIVITY APPLIED TO INDONESIAN SMARTPHONE USERS' CREDITWORTHINESS AFTER COVID-19 PANDEMIC

RADEN RARA KARTIKA KUSUMA WINAHYU*

Lecturer, Department of Informatics Management, Politeknik Astra, Jl. Gaharu Blok F3 Delta Silicon II, Cibatu, Cikarang Selatan, Kab. Bekasi, Jawa Barat, Indonesia. ORCHID ID: https://orcid.org/0009-0000-3640-0004, *Correspondence Author Email: raden.rara@polytechnic.astra.ac.id

LATHIFAH ALFAT

Lecturer, Department of Informatics, Universitas Pembangunan Jaya, Jl. Cendrawasih Raya Blok B7/P Bintaro Jaya, Ciputat, Tangerang Selatan, Banten, Indonesia. ORCHID ID: https://orcid.org/0000-0002-9873-9510, Email: lathifah.alfat@upj.ac.id

Abstract

In this research work, we used machine learning techniques to predict the creditworthiness of smartphone users in Indonesia after the COVID-19 pandemic. Principal Component Analysis (PCA) and K-means algorithms were used for dimensional reduction and clustering using a dataset of 803 respondents consisting of twelve questions to smartphone users in Indonesia after the COVID-19 pandemic. To classify the creditworthiness of smartphone users in Indonesia, the four different classification algorithms (Logistic Regression, Support Vector Machine, Decision Tree, and Naïve Bayes) were tested. The tests carried out included testing for accuracy, precision, recall, F1-score, and Area Under Curve Receiver Operating Characteristics (AUCROC) assessment. When compared to other models, the Logistic Regression algorithm outperforms them. The findings of this study also provide new information about the most influential and non-influential variables based on the twelve questions posed to smartphone users in Indonesia, which can assist financial institutions, particularly banks, in assessing the creditworthiness of prospective customers after COVID-19 pandemic.

Keywords: Creditworthiness, Smartphone, Machine Learning, COVID-19 Pandemic

INTRODUCTION

Most people are familiar with banks these days. Credit is one of the banking products that helps individuals meet their day-to-day demands, but there are still some people who are unable to receive credit owing to the old technique that is still employed by financial institutions, particularly banks. According to projections, the number of smartphone users in Indonesia could reach 239 million by 2026 (Kartika Winahyu et al., 2022). Technology-based financial innovations have grown significantly in most countries around the world, including Indonesia, in recent years in order to facilitate the provision of financial services and improve the quality of financial services.(Santoso et al., 2020). Clients' smartphone data usage can be utilized to analyze credit ratings, eliminate paperwork, shorten loan delivery periods, and boost the selection of suitable clients (Shema, 2019). Because of the vast number of possible borrowers, it is vital to utilize models and algorithms that avoid human faults in credit application analysis in consumer lending (Aniceto et al., 2020).

The Seybold REPORT
ISSN 1533-9211
DOI: 10.5281/zenodo.10049842

Creditworthiness is influenced by elements such as ability to pay and willingness to pay. The financial factor is tied to the ability to pay, but the personal factor is related to the willingness to pay (Kousayri, n.d.).

However, there has been a change in financial condition experienced by Indonesian people who were also smartphone users during the COVID-19 pandemic since according to (Nugroho, n.d.), Indonesia experienced an economic crisis in which several people lost their jobs and more than 80% of companies saw a significant drop in profits. Production fell and distribution/logistics lines were disrupted as a result of mobility restrictions. The condition of economic crisis, combined with mobility restrictions during the implementation of the Community Activity Restrictions (PPKM), resulted in a decrease in people's income and spending interest which led to a decrease in credit expansion and indicated sufficient liquidity in the banking sector (Nugroho, n.d.). Furthermore, banks and other financial institutions in Indonesia are also seeing increased loan default rates as a result of the economic crisis during pandemic times (Rizwan et al., 2020).

To mitigate the detrimental impact of the Covid-19 outbreak on the Indonesian economy, the Indonesian government has implemented fiscal policies such as acceleration of government spending, relaxation of income taxes, and revival of the national economy through executing state financial policies through relaxation of the State Budget (APBN) (Yenni Ratna Pratiwi, 2022). According to a press release of Bank Indonesia as the central bank of the Republic of Indonesia, Indonesia's economy grew strongly in second quarter (Q2) of 2023. This economic growth was characterized by an increase in domestic demand which supported rapid economic expansion. Household consumption increased by 5.23% year on year, reflecting increased mobility, increased income expectations, controlled inflation, and the positive impact of National Religious Holidays, and the payment of the 13th salary to the State Civil Servant. (Erwin Haryono, 2023).

With this research it is hoped that it will be able to clarify the results of previous studies that are different and even contradictory, this research also aims to expand this literature by adding attributes that have a significant impact and attributes that do not have a significant impact on smartphone users in Indonesia after the COVID-19 pandemic based on the twelve attributes (Alfat et al., 2019b). Then, grouping and classifying smartphone users based on the twelve attributes using machine learning algorithms which can later help financial institutions in Indonesia, especially banks in assessing the creditworthiness of prospective customers after pandemic times.

LITERATURE REVIEW

1. Credit Scoring using Mobile Phone Usage

According to The Five Characteristics of credit model, five variables must be considered when determining a borrower's creditworthiness. These requirements include Capacity, Capital, Character, Conditions, and Collateral, which is an alternative source of payment that is not required for loan approval (Alfat et al., 2019b). Commercial banks frequently rely on highly

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.10049842

subjective risk control personnel to assess the credit risk of loan applications (Wang et al., 2020). However, the traditional method of credit scoring are problematic for a variety of reasons because they do not appropriately depict the economic characteristics of thin-file borrowers. Second, they are not prepared regularly enough to incorporate the most recent developments in borrowers' lives. Third, they are not easily and promptly accessible at an inexpensive cost (Lainez, 2021). Daniel Björkegren and Darrell Grissen was successful in finding solutions to problems faced by residents of developing countries in South America by using alternative data sources, such as cell phone usage history that can indicate creditworthiness in previously unbanked persons due to a lack of formal financial history (Björkegren & Grissen, 2019). On the other hand, MobiScore proved that mobile phone data could be used to construct credit profiles that are as reliable as those given by traditional credit agencies by merging demographic data and call detail records (CDR) with credit card data from a financial institution in the same nation. (Pedro et al., 2015).

2. Principal Component Analysis and K-means

PCA is a commonly used method for lowering the dimension of huge datasets. PCA is another approach that can be used to characterize the subspace. On PCA, orthonormal transformation yields variables that are linearly independent of the data set and have the potential to be correlated. PCA is also a statistical method that substitutes a large number of original elements with a few key ones (Migenda et al., n.d.).

The K-means method takes as inputs the number of clusters and the starting point of the cluster center. The steps of the K-means algorithm are illustrated below (Hossain et al., 2019).

Step 1: Select the first cluster centers from the dataset of n data objects by selecting the number of clusters k.

Step 2: Determine the distance between each data object in the dataset with i=0,1,2,...k-1, and all k cluster centers c with j, where j=0,1,2,...k-1. Allocate the distance between each data object in dataset i and the cluster nearest to it.

Step 3: Reassess the worth of each cluster center j.

Step 4: If the cluster's centers remain unchanged, repeat Steps 2 and 3.

3. Logistic Regression

The subordinate components in Logistic Regression are normally subjective factors, and the binary logistic regression is consistent with the circumstance in this study. Logistic regression is constructed by considering the components that influence a credit score (Ruyu et al., 2019).

4. Support Vector Machine (SVM)

The appropriate balance between the model's complexity and learning capacity is sought in order to ensure the best expansion capacity, and SVM is capable of interpreting information based on limited test data. When compared to other classification methods, the SVM can identify the objective function's global least squares (Ruyu et al., 2019) .

The Seybold REPORT

ISSN 1533-9211

DOI: 10.5281/zenodo.10049842

5. Naïve Bayes (NB)

NB discusses selecting the best class marker based on alternatives and misunderstanding losses (Ruyu et al., 2019). Given the class variable, the NB classifier assumes that the presence or absence of a single characteristic has no bearing on the presence or absence of any other feature. In practice, the maximum likelihood technique is used to estimate NB parameters. In practice, Naive Bayes classifiers have excelled in a wide range of challenging real-world situations (Omidiora et al., 2013) .

6. Decision Tree

The data is partitioned into a series of rectangular subdivisions using the Decision Tree approach. The critical nodes for splitting are shown by the position of each node at the top of the tree. The data is then linked to additional nodes in the given order via branches, resulting in a tree-like classification. When the "impurity" is calculated, the nodes are divided (Ruyu et al., 2019).

Related Work

Research on the accuracy of phone data in anticipating client creditworthiness has revealed that the precise patterns of real-time financial behavior represented in smartphone data can aid in enhancing customers' creditworthiness knowledge (Shema, 2019). Mobiscore, a Caribbean microcredit organization successfully reduced default by nearly half of total credit while keeping three-quarters of its debtors by combining demographic and call detail records (CDR) with financial information from the same country's financial institution (Pedro et al., 2015). While research that employed CDR and data received from users' mobile handsets in conjunction with standard credit scoring system data to surpass the selection of loan applications screening procedures, improving acceptance rates and decreasing late rates (Neftali et al., 2017) . On the other hand, social network factors between individual networks such as age, gender, ethnicity, language, economic factors, geography, urbanization, and epidemics can be studied through credit scoring assessment using CDR data (Óskarsdóttir et al., 2019). Furthermore, research on microfinance credit scoring models based on social media network information taken from Facebook accounts conducted by De Cnudde et al. found that an explicit network of interacting people was more predictive than an explicit network of friends who did not, but that a network of people exhibiting similar behavior outperformed both explicit networks (Óskarsdóttir et al., 2019).

METHOD

Questionnaire

This research uses a survey method using a questionnaire. This study used a questionnaire-based survey approach. The questionnaire was meant to provide the researchers with important information. The closed questions method was employed to allow the researchers to express their perspective by selecting what they believe. The Likert scale was developed to identify five levels of response. The arithmetic mean range (1.5 to 2.5) indicates low approval, the field

ISSN 1533-9211

DOI: 10.5281/zenodo.10049842

(2.5 to 3.5) shows medium acceptance, and the range (3.5 to 5) suggests strong approval. The twelve questions asked in the questionnaire were classified into three different categories: background (2 questions), smartphone usage parameter (8 questions), trustworthiness (2 questions) (Alfat et al., 2019a) as shown in Table 1. The implementation of the twelve variables manifested in the form of questionnaire questions is shown in Table 2.

Table 1

Category	Variables
Background	Age
	Occupation
Trustworthiness	Spend the rest of the money
	Maintain assets
Smartphone usage	Phone manufacturer
	Game data usage
	Social media data usage
	Expense money for smartphone
	Internet costs
	Switching phones on a regular basis
	The reason for switching phones
	Pay for yourself

Table 2

Variables	Questionnaire Questions
Age	How old are you?
Occupation	What is your occupation?
Spend the rest of the money	If I have money left in my wallet then I will spend it instead of saving it.
Maintain asset	I would rather save my money than spend it on buying things I don't need.
Phone manufacturer	What is your smartphone's manufacturer?
Game data usage	How much data do you use for games per month?
Social media data usage	How much data do you use for social media per month?
Expense money for smartphone	How much money do you spend on your smartphone per month?
Internet cost	How much money do you spend on the internet per month?
Switching phones on a regular basis	How often do you change your smartphone?
The reason for switching phones	What is the reason you change your smartphone last time?
Pay for yourself	I always pay for my personal needs including my smartphone expenses.

Respondents

The survey was conducted in July 2023 after Presiden Joko Widodo officially withdrew the COVID-19 pandemic national emergency designation on 21st June, 2023. The survey attended by 803 smartphones users in Indonesia from various age and economic backgrounds as shown in Fig. 1 and Fig. 2. The number of respondents in this survey was calculated by taking the total population of Indonesia in 2021, which was 272 million people (Kartika Winahyu et al., 2022). The minimum number of respondents is 663 when using a 99% confidence level and a 5% margin of error (Hazra, 2017). The respondents were given a link to the online

ISSN 1533-9211
DOI: 10.5281/zenodo.10049842

questionnaire. Then, based on their smartphone usage, they voluntarily responded to the questionnaire.

The age range of respondents was between 10 years to more than 60 years. The age distribution of respondents consisted of 22% was 10 – 20 years, 30% was 21 – 30 years, 29% was 31 – 40 years, 10% was 41 – 50 years, and 6% was 51 – 60 years. The occupation distribution of respondents consisted of 18% was private employees, 10% was civil servants, 5% was state-owned company employees, 6% was teachers/lecturers, 4% was entrepreneurs, 31% was students, 2% was content creator, and 1% was farmers. The rest, about 27% worked as freelancer and other professions.

Monthly social media data distribution of respondents as shown in Fig. 3 consisted of >10 Gigabyte (GB) (26%), 5,1 – 10 GB (30%), 1 – 5 GB (36%), and < 1 GB (8%). Monthly internet cost distribution of respondents as shown in Fig. 4 consisted of < Rp 100.000 (33%), Rp 100.000 – Rp 200.000 (24%), Rp 200.001 – Rp 300.000 (17%), Rp 300.001 – Rp 400.000 (19%), Rp 400.001 – Rp 500.000 (5%), and >Rp 500.000 (2%). Monthly smartphone expenses distribution of respondents as shown in Fig. 5 consisted of < Rp 100.000 (29%), Rp 100.000 – Rp 200.000 (26%), Rp 200.001 – Rp 300.000 (21%), Rp 300.001 – Rp 400.000 (16%), Rp 400.001 – Rp 500.000 (4%), and the rest > Rp 500.000 (4%).

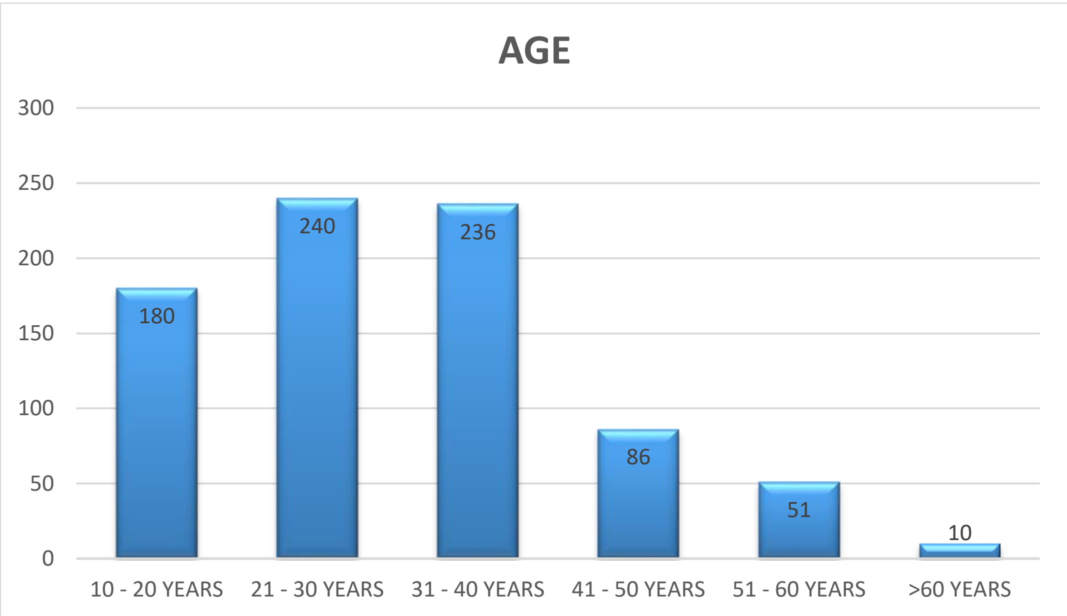

Fig 1: The age distribution of respondents

ISSN 1533-9211
DOI: 10.5281/zenodo.10049842

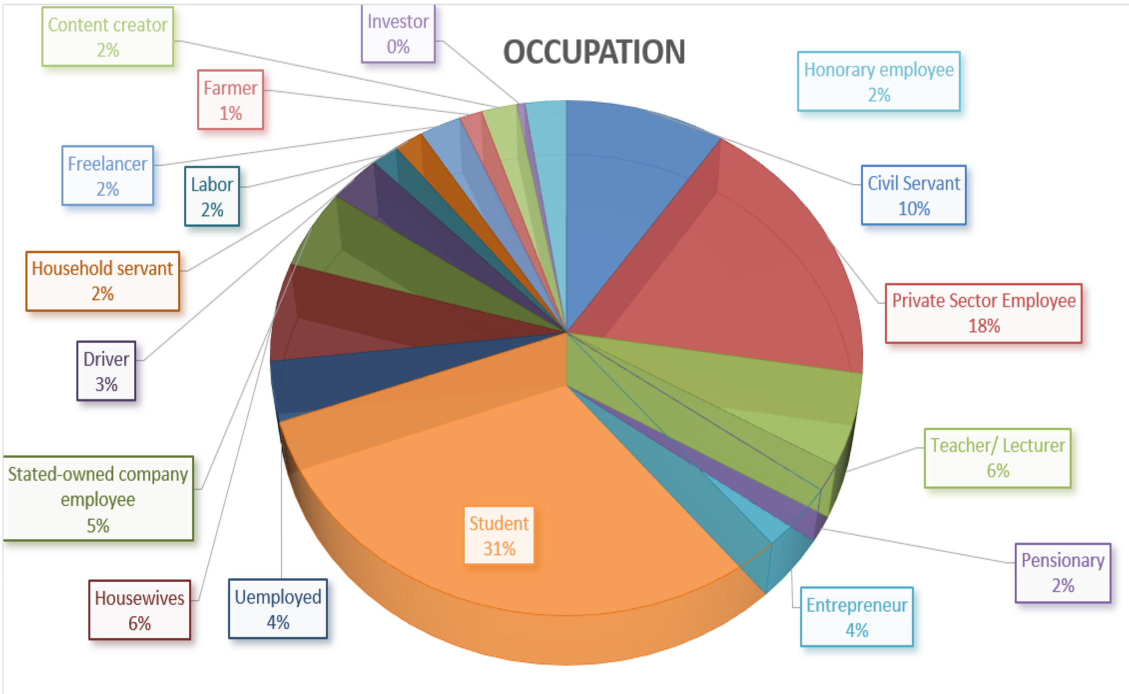

Fig 2: The occupation distribution of respondents

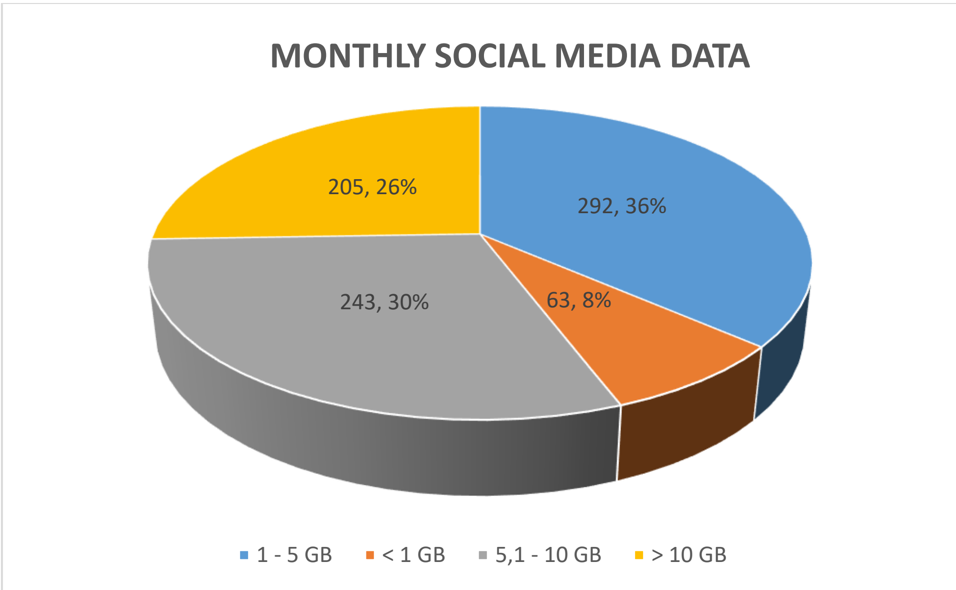

Fig 3: Monthly social media data distribution of respondents

ISSN 1533-9211

DOI: 10.5281/zenodo.10049842

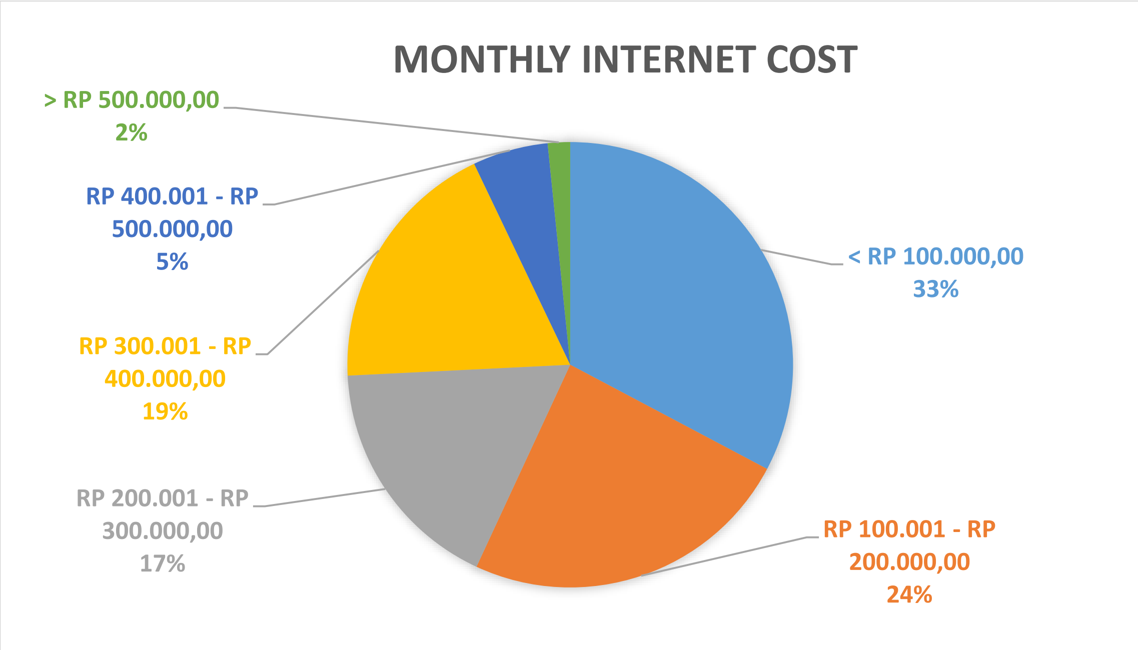

Fig 4: Monthly internet cost distribution of respondents

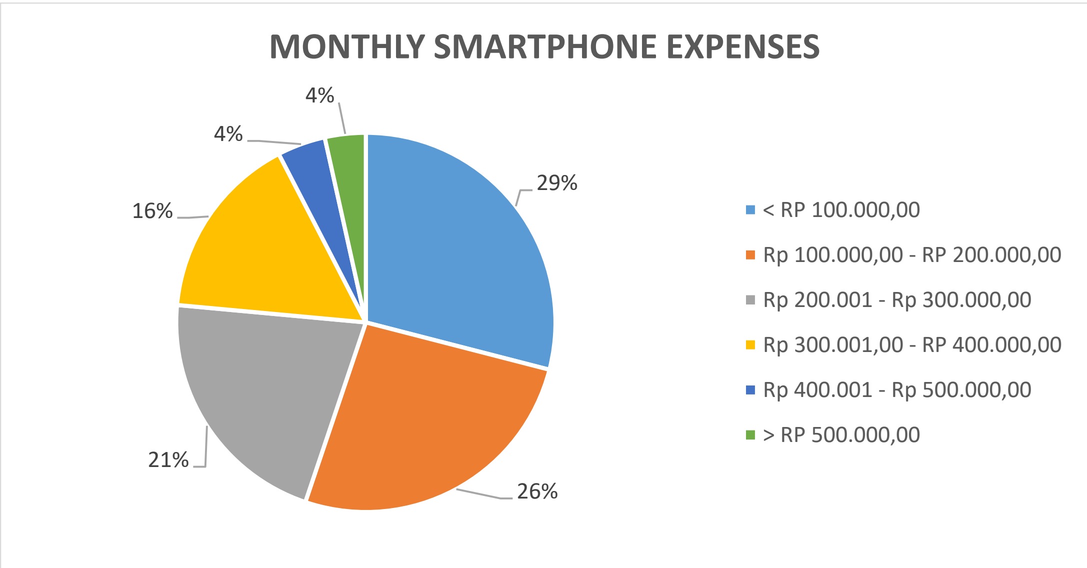

Fig 5: Monthly smartphone expenses distribution of respondents

PCA and K-means Clustering Implementation

Outlier reduction is performed prior to the Principal Component Analysis (PCA) process because outliers have a high impact on PCA (Jollife & Cadima, 2016). From the heatmap correlation between attributes shown in Fig. 6, we can conclude that only 10 of 12 attributes had a correlation between attributes. The two attributes that did not have a significant impact

ISSN 1533-9211

DOI: 10.5281/zenodo.10049842

on the condition of smartphone users in Indonesia after pandemic times were the reason for switching smartphone and pay for yourself. These results further convince us that the financial condition of smartphone users in Indonesia has changed after the economic crisis of COVID-19 pandemic. From the heatmap shown in Fig. 6, we can also conclude that attribute monthly smartphone expenses had a high correlation with the attribute monthly internet cost by 0.72 and attribute spend the rest of the money had a high correlation with attribute saving by 0.59. The next process was applying PCA calculation to the new dataset to reduce the dimension of the dataset. From the PCA process, the attribute frequency of changing smartphones and monthly smartphone expenses were the two principal components that account for 54% of the data variation as shown in Fig. 7. Following the PCA process, the K-means algorithm was applied to visualize the cluster based on the results of previous calculations.

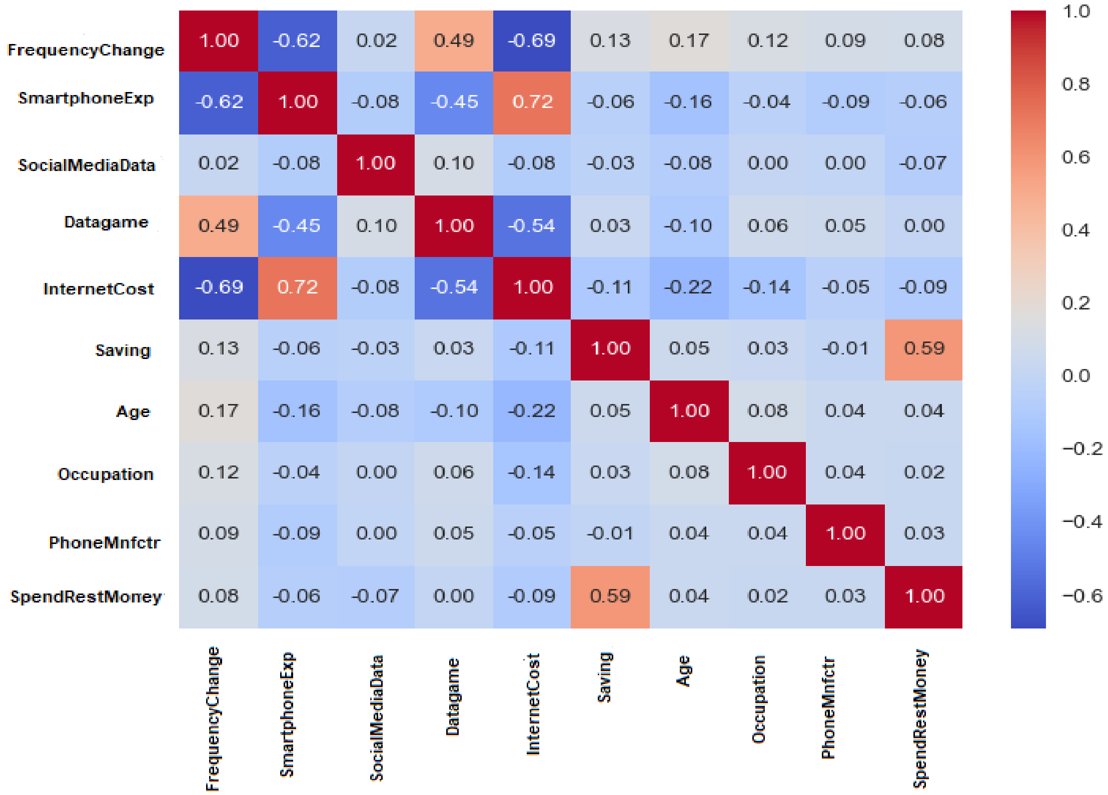

Fig 6: Seaborn or heatmap correlation between attributes

ISSN 1533-9211

DOI: 10.5281/zenodo.10049842

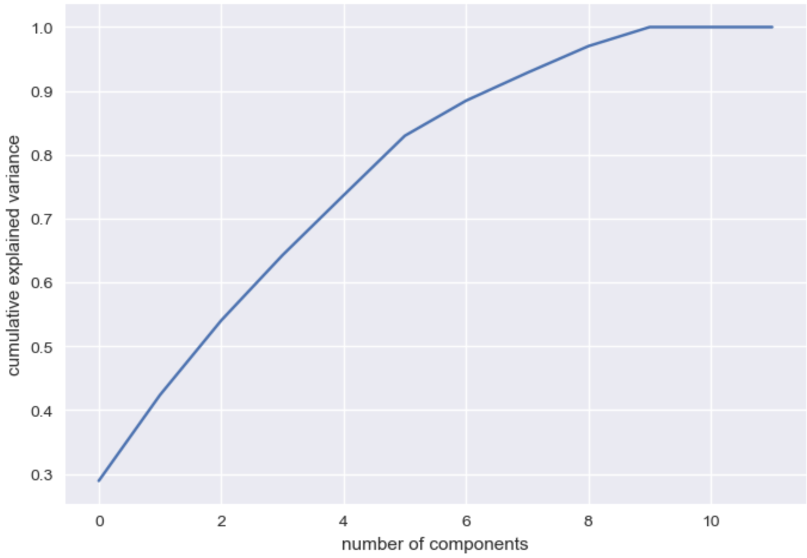

Fig 7: The cumulative variance graph of each component

Performances Evaluation

This research also models the new dataset using four machine learning classification algorithms after the clustering process. The four machine learning classifier algorithms used are Logistic Regression, Decision Tree, SVM, and Naïve Bayes. There are five performance tests carried out on the machine learning classifier algorithms used as described below:

- Accuracy: Expressed as the ratio of True Positive (TP) and True Negative (TN) to True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). The mathematical expression for the accuracy parameter is shown in (1):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

- Precision: Expressed as the ratio of TP to total predicted positive (TP and FP). The mathematical expression for the precision parameter is shown in (2):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

- Recall: Expressed as the ratio of TP to total actual positive (TP and FN). The mathematical expression for the recall parameter is shown in (3):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

ISSN 1533-9211
DOI: 10.5281/zenodo.10049842

- F1-score: Expressed as the balance between the precision and the recall. The mathematical expression for the F1-score parameter is shown in (4):

$$\text{F1 score} = 2 \text{ x} \frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

- AUCROC: This test is to show how well the model can differentiate between classes.

RESULT AND DISCUSSIONS

Result

After calculating the dataset consisting of the twelve questions mentioned above with the PCA and K-means algorithms, three clusters of smartphone users in Indonesia were generated after the COVID-19 pandemic as shown in Table 3 below. Cluster 3 consists of smartphone users who spent the most money on their smartphones compared to others, but their tendency to save and maintain their assets was high. Meanwhile, cluster 2 consists of smartphone users who spent the least money on their smartphones, but their tendency to save and maintain their assets was low. The last cluster is cluster 1 which consists of smartphone users who spent low money on their smartphones, but their tendency to save and maintain their assets was high.

From a total of 803 respondents, the following results were obtained: A total of 31,5% respondents were included in cluster 1. While a total of 27,4% respondents were included in cluster 2 and the remaining 41,1% respondents were included in cluster 3. Furthermore, the new dataset was split into two sections: training and testing, with an 80%:20% split. A training set of eighty percent of the samples chosen at random from the dataset was used to build a predictive model while the remaining twenty percent of samples were used as blind samples to assess the predictive model's robustness.

The four machine learning classifier algorithms used are Logistic Regression, Decision Tree, SVM, and Naïve Bayes. They can build a good machine learning model with accuracy, precision, recall, and F1-score of more than 90% except Decision Tree as shown in Table 4. We can see from Table 4, Logistic Regression shows the accuracy of 0.99, a precision of 0.99, a recall of 0.99, and an F1-score of 0.99. While the Decision Tree with an accuracy of 0.87, a precision of 0.89, a recall of 0.88, and an F1-score of 0.87. Algorithm SVM shows an accuracy of 0.93, a precision of 0.93, a recall of 0.92, and an F1-score of 0.92. While Naïve Bayes has a precision value of 0.92, an accuracy of 0.92, a recall of 0.93, and an F1-score of 0.92.

In addition, the Area Under Curve Receiver Operating Characteristics (AUC ROC) test was also carried out and the results are shown in Table 5. From Table 5, it can be concluded that Logistics Regression has a perfect AUCROC value of 1.0 for the three clusters. Meanwhile, SVM and Naive Bayes have almost perfect scores.

ISSN 1533-9211
DOI: 10.5281/zenodo.10049842

Table 3

Variable name	Cluster 1	Cluster 2	Cluster 3
Age	Young	Oldest	Young
Expense money for smartphone	Low	Low	High
Internet costs	Low	Low	High
Data use for social media	Low	High	Low
The tendency to maintain their assets	High	Low	High
The tendency to not spend the rest of the money	High	Low	High

Table 4

Algorithm	Performance			
	Accuracy	*Precision*	*Recall*	*F1-score*
Logistic Regression	0.99	0.99	0.99	0.99
Decision Tree	0.87	0.89	0.88	0.87
SVM	0.93	0.93	0.92	0.92
Naïve Bayes	0.92	0.92	0.92	0.92

Table 5

Cluster	Algorithm			
	Logistic Regression	*Decision Tree*	*SVM*	*Naïve Bayes*
1	1.0	0.96	0.99	0.98
2	1.0	0.96	0.99	0.99
3	1.0	0.97	1.0	1.0

CONCLUSION

This research work presents the prediction of smartphone users' creditworthiness using machine learning after the COVID-19 pandemic in Indonesia. A total of 1050 respondents participated answering 12 questions consisting of three key categories. With the PCA and K-means computations, our research concludes that there are three clusters of smartphone users in Indonesia during the economic crisis in pandemic times. The three resulting clusters are expected to help financial institutions in Indonesia, especially banks to predict prospective customers' creditworthiness so that banks can reduce the risk of loan default after the COVID-19 pandemic. The study also models four machine learning classifier algorithms and Logistic Regression shows the perfect results of all performances evaluation compared to the other three algorithms. In the future, the work to examine the reasons for the creditworthiness of prospective customers who are also smartphone users based on the performance of the machine learning algorithm models and the important features that change as a result of the influence of the geographical area and economic conditions of a country is an interesting research topic for further discussion.

Research Limitations

As for the limitations of the study: (1) the assessment of all research variables was measured and analyzed at the level of Indonesia smartphone users perception and experience, so that the interpretation of research results may not reflect the actual conditions. (2) Data collection techniques using closed questionnaires allow objective reality to be bound by the questionnaire.

ISSN 1533-9211

DOI: 10.5281/zenodo.10049842

Reference

1) R. R. K. Winahyu, M. Somantri and O. D. Nurhayati, "Predicting Creditworthiness of Smartphone Users in Indonesia during the COVID-19 pandemic using Machine Learning," 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), Jakarta, Indonesia, 2022, pp. 223-227, doi: 10.1109/ISMODE53584.2022.9742831.

2) Wimboh Santoso, Irwan Trinugroho & Tastaftiyan Risfandy (2019): What Determine Loan Rate and Default Status in Financial Technology Online Direct Lending? Evidence from Indonesia, Emerging Markets Finance and Trade, DOI: 10.1080/1540496X.2019.1605595

3) A. Shema, “Effective credit scoring using limited mobile phone data,” ACM Int. Conf. Proceeding Ser., 2019, doi:10.1145/3287098.3287116.

4) Aniceto, M.C., Barboza, F. & Kimura, H. Machine learning predictivity applied to consumer creditworthiness. Futur Bus J 6, 37 (2020). https://doi.org/10.1186/s43093-020-00041-w

5) Kousayri, Husain, Creditworthiness Standards between Theory and Practice (a Field Study on the Syrian Commercial Bank in Lattakia) (May 2, 2020). Available at SSRN: https://ssrn.com/abstract=3591460 or http://dx.doi.org/10.2139/ssrn.3591460

6) Muhyiddin, Muhyiddin, and Hanan Nugroho. "A Year of Covid-19: A Long Road to Recovery and Acceleration of Indonesia's Development." Indonesian Journal of Development Planning, vol. 5, no. 1, 30 Apr. 2021, pp. 1-19, doi:10.36574/jpp.v5i1.181.

7) Muhammad Suhail Rizwan, Ghufran Ahmad, Dawood Ashraf, Systemic Risk: The Impact of COVID-19tn1], Finance Research Letters (2020), doi: https://doi.org/10.1016/j.frl.2020.101682

8) Pemulihan Perekonomian Indonesia Setelah Kontraksi Akibat Pandemi Covid-19. 2022. https://www.djkn.kemenkeu.go.id/kpknl-banjarmasin/baca-artikel/14769/Pemulihan-Perekonomian-Indonesia-Setelah-Kontraksi-Akibat-Pandemi-Covid-19.html (accessed Aug. 10, 2023)

9) Ekonomi Indonesia Tumbuh Tinggi Pada Triwulan II 2023. 2023. https://www.bi.go.id/id/publikasi/ruang-media/news-release/Pages/sp_2521523.aspx (accessed Aug. 10, 2023)

10) L. Alfat, M. Rizkinia, R. F. Sari, and D. M. Romano, “Feature Selection of Credit Score Factor Based on Smartphone Usage using MCFS,” TIMES-iCON 2019 - 2019 4th Technol. Innov. Manag. Eng.Sci. Int. Conf., no. December, 2019, doi: 10.1109/TIMESiCON47539.2019.9024462.

11) Y. Wang, Y. Zhang, Y. Lu, and X. Yu, “A Comparative Assessment of Credit Risk Model Based on Machine Learning ——a case study of bank loan data,” Procedia Comput. Sci., vol. 174, pp. 141–149, 2020, doi: 10.1016/j.procs.2020.06.069.

12) Lainez, Nicolas. (2021). The Prospects and Dangers of Algorithmic Credit Scoring in Vietnam: Regulating a Legal Blindspot.

13) Daniel Björkegren , Darrell Grissen, Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment, The World Bank Economic Review, Volume 34, Issue 3, October 2020, Pages 618–634, https://doi.org/10.1093/wber/lhz006

14) San Pedro, Jose & Proserpio, Davide & Oliver, Nuria. (2015). MobiScore: Towards Universal Credit Scoring from Mobile Phone Data. 195-207. 10.1007/978-3-319-20267-9_16.

15) Migenda, Nico & Möller, Ralf & Schenck, Wolfram. (2019). Adaptive Dimensionality Adjustment for Online “Principal Component Analysis”. 10.1007/978-3-030-33607-3_9.

16) Hossain, Md & Akhtar, Md.Nasim & Ahmad, R.Badlishah & Rahman, Mostafijur. (2019). A dynamic K-means clustering for data mining. Indonesian Journal of Electrical Engineering and Computer Science. 13. 521. 10.11591/ijeecs.v13.i2.pp521-526.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.10049842

17) Ruyu, Bai & Mo, Hai & Haifeng, Li. (2019). A Comparison of Credit Rating Classification Models Based on Spark- Evidence from Lending-club. Procedia Computer Science. 162. 811-818. 10.1016/j.procs.2019.12.054.

18) Omidiora, Elijah & Adeyanju, Ibrahim & Fenwa, Olusayo. (2013). Comparison of Machine Learning Classifiers for Recognition of Online and Offline Handwritten Digits. Computer Engineering and Intelligent Systems. 4. 39-47.

19) Ruiz, Saulo & Gomes, Pedro & Rodrigues, Luís & Gama, João. (2017). Credit Scoring in Microfinance Using Non-traditional Data. 447-458. 10.1007/978-3-319-65340-2_37.

20) M. Óskarsdóttir, et al., The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics, Applied Soft Computing Journal (2018), https://doi.org/10.1016/j.asoc.2018.10.004

21) Hazra, Avijit. (2017). Using the confidence interval confidently. Journal of Thoracic Disease. 9. 4124-4129. 10.21037/jtd.2017.09.14.

22) Jolliffe IT, Cadima J. 2016. Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A 374: 20150202. http://dx.doi.org/10.1098/rsta.2015.0202