

## Analysis of the Characteristics of Customers Experiencing Non-Performing Loans (NPL) Using a Data Mining Approach (Case Study at PT. Pegadaian (Persero) Makale Branch Service Unit)

Sylventer Iven<sup>1</sup>, Marchelin<sup>2</sup>, Elisabet Pali<sup>3</sup>

<sup>1</sup>Universitas Kristen Indonesia Toraja,

<sup>2</sup>Universitas Kristen Indonesia Toraja,

<sup>3</sup>Universitas Kristen Indonesia Toraja, Indonesia

Email: [sylventeriven12@gmail.com](mailto:sylventeriven12@gmail.com)

---

### **Abstract**

#### **Keywords:**

Non Performing Loan (NPL),  
Customer Characteristics, Data  
Mining, Clustering, Pawnshops

*This study aims to analyze the characteristics of customers experiencing Non-Performing Loans (NPL) using a data mining approach at PT Pegadaian (Persero) Makale Branch Service Unit. The type of research used is quantitative research with data analysis techniques based on a data mining approach using a clustering model (K-Means algorithm). The data used are secondary data obtained from PT Pegadaian (Persero) Makale Branch Service Unit. The variables analyzed include the characteristics of customers experiencing NPL with several indicators such as gender, age, occupation, income, education level, marital status, loan term, total loan, and credit status. The research stages begin with the data selection process, data cleaning (pre-processing), data transformation, clustering process, and interpretation of the results. The results show that there are several customer clusters in each credit status with different characteristics that have the potential to experience NPL. Each cluster has a certain pattern based on demographic, socio-economic, and customer loan characteristics. These findings can be used as a basis for PT Pegadaian (Persero) Makale Branch Service Unit in improving creditworthiness analysis and developing more effective credit risk mitigation strategies.*

---

## INTRODUCTION

In line with economic development, the public's need for financial services has also increased, both in terms of quality and quantity. Financial institutions include both banking and non-banking financial institutions, although banking institutions dominate the economy. (Eprianti, 2019) Financial institutions, both banks and non-banks, conventional and sharia-compliant, play a significant role in the Indonesian economy. As intermediaries, financial institutions mediate between those with excess funds and those with a shortage. They collect funds from those with surplus funds and then channel them to those with deficit funds in the form of financing. (Amir et al., 2023).

One of the non-bank financial institutions is PT Pegadaian (Persero) which is a State-Owned Enterprise (BUMN) engaged in financial services, specifically pawn-based financing. This company has a strategic role in providing fast and easy financial solutions for the community, by using valuables as collateral. In addition to pawn services, Pegadaian also offers microfinance products, precious metal sales, and other services aimed at supporting financial inclusion in Indonesia. As a non-bank financial institution, PT Pegadaian (Persero) has a vision to "Improve community welfare through the provision of trusted financial solutions". With an

extensive network and service innovation, Pegadaian is committed to helping the community meet their financial needs while driving national economic growth.

The existence of financial institutions, both banking and non-banking, and the public as economic actors, along with regulations and monetary policies, has led to the circulation of money in the economy, where funds are circulated through various mechanisms, such as credit programs provided by financial institutions to the public. Credit provision requires a procedure or system that serves as a guideline for customers from applying for credit to paying installments. Credit provision is a significant source of profit or income for companies. The largest income from credit institutions is loan interest. Credit provision also carries the risk of default and payment delays, which occur because customers neglect to pay their credit installments to the company. Therefore, pawnshops use the principle of prudence in providing credit to prospective creditors and pawnshops must strive to minimize risks in providing credit. (Oktavia, 2022).

Credit is one of the main services in the activities of financial institutions, including PT Pegadaian (Persero). As a financial institution engaged in financing, PT Pegadaian (Persero) plays a crucial role in supporting the community's funding needs, both for consumptive and productive needs. However, in its implementation, providing credit is not free from risks, one of which is the risk of non-performing loans (NPLs). NPLs can have a significant impact on a company's financial stability, both in terms of liquidity, profitability, and customer trust. Therefore, it is crucial for PT Pegadaian (Persero) to have effective strategies and mechanisms to minimize the occurrence of NPLs. In this context, appropriate credit provision, based on comprehensive analysis, and measurable risk management are key to success. In practice, non-performing loans often occur due to careless or insufficient credit analysis in the credit granting process, as well as from unfavorable customer character. The NPL phenomenon is not only caused by the customer's inability to fulfill payment obligations, but can also be caused by a suboptimal credit granting process. Factors such as weak creditworthiness analysis, inaccurate collateral assessments, and inadequate debtor oversight have the potential to exacerbate non-performing loans. To prevent non-performing loans, credit analysts at financial institutions, both banks and non-banks, must be able to make informed decisions about whether to accept or reject credit applications. To determine the feasibility of future credit distribution, accurate analysis is necessary, one method of which is the use of data mining technology.

Non-performing loans (NPLs) are one of the biggest challenges in the financing industry, including at PT Pegadaian (Persero) in Tana Toraja Regency. NPLs not only affect a company's liquidity but also reflect the effectiveness of risk management in assessing the creditworthiness of potential customers. As a pawn-based financing institution, PT Pegadaian (Persero) has a very diverse customer segment, making it crucial to understand the characteristics of customers who are at risk of non-performing loans. With advances in information technology, data mining approaches can be utilized to analyze hidden patterns in historical customer data. Through this technique, the company can be more proactive in managing credit risk and improve the effectiveness of lending.

The following is a list of NPLs obtained by researchers from the pre-research survey.

**Table 1.1**  
**PT Pegadaian UPC Makale Non-Performing Loan Data**  
**Microfinance Products**

Month/Year	Credit Amount	Non-Performing Loans	NPL %
December 2022	Rp. 4,893,557,029	Rp. 170,870,380	3.49%
December 2023	Rp. 9,799,500,030	Rp. 276,035,171	2.82%
December 2024	Rp. 6,858,077,091	Rp. 379,667,711	5.54%
As of April 2025	Rp. 2,968,477,259	Rp. 140,686. 126	2.11%

Source: PT Pegadaian (Persero) Makale Branch Service Unit Office (2025)

Table 1.1 above shows that non-performing loans (NPL) is a problem faced by PT Pegadaian (Persero) Makale Branch Service Unit from at least 2022 to 2024. NPL movements over the past three years reflect fluctuations that indicate the dynamics of customer repayment capacity or the level of smoothness of payments. Therefore, NPL management is crucial for maintaining the financial health of PT Pegadaian (Persero) Makale Branch Service Unit.

Based on the description of the problem, this research is intended to assist credit analysts and creditors in understanding customer characteristics as one of the credit granting processes at PT Pegadaian (Persero) Makale Branch Service Unit in order to minimize and prevent the occurrence of NPL. This research is expected to improve the work and performance of credit analysts in obtaining worthy and appropriate debtors in the credit granting process, provide an overview of the effectiveness of the current credit granting system from the perspective of customer characteristics, and formulate strategic recommendations that can be implemented to improve credit quality and reduce the risk of NPL in the future. Therefore, the author is interested in conducting a study entitled "Analysis of Customer Characteristics Experiencing Non-Performing Loans (NPL) Using a Data Mining Approach (Case Study at PT. Pegadaian (Persero) Makale Branch Service Unit)".

## METHODS

The type of research used in this study is quantitative. This approach systematically and measurably depicts and describes the phenomena in this study using numerical data.

In this study, the type of data and data sources are important aspects that influence the analysis method, collection, processing, and interpretation of the research results. The primary data in this study is information on credit distribution and NPL obtained directly from observations and surveys conducted by researchers by observing activities directly in the field (PT Pegadaian (Persero) Makale Branch Service Unit). In addition to observations and surveys, primary data was also obtained by researchers through direct interviews with Marketing employees, Account Officers (non-pawn products), and Managers (Unit Leaders/Managers). While secondary data in this study are credit report documents and archives as well as historical data on annual and monthly NPL customers obtained by researchers at PT Pegadaian Makale Branch Service Unit. The data sources in this study were obtained by researchers through field research conducted directly to obtain information related to research at PT Pegadaian (Persero) Makale Branch Service Unit.

In this study, the population is defined as the entire subject or object that will have certain characteristics and become the focus of a study. Meanwhile, the sample is a portion of the population that has certain conditions and characteristics that will be studied. The population in this study were active microfinance customers of PT Pegadaian (Persero) Makale Branch Service Unit from 2023-2025 who experienced Non-Performing Loans (NPLs), with a total of 114 customers. The sample in this study was obtained by researchers using a saturated sampling technique. This technique is used in sampling by using the total population as the sample, so the sample in this study was 114 customers.

The variables in this study include the characteristics of customers who experience NPL with several indicators such as customer age, gender, customer occupation, customer income, loan term, marital status, last education, and credit status. The main instrument in this study used historical customer data in the form of documents obtained from the internal system of PT Pegadaian (Persero) Makale Branch Service Unit. The collected data was measured using supporting numerical software (data mining analysis software such as Microsoft Excel and Orange).

The data collection technique used in this study is the documentation technique. Data collection using the documentation technique was carried out by researchers by collecting and analyzing customer data information, credit report documents and archives, as well as historical annual and monthly NPL data obtained at PT Pegadaian (Persero) Makale Branch Service Unit. Data analysis techniques are systematic processes for processing, interpreting, and analyzing collected data so that it can be used to answer questions regarding the research being conducted or to test hypotheses. In this study, the data analysis technique used is a Data Mining-based data analysis technique with a Clustering model, specifically the K-Means algorithm, to group customers who experience NPL based on certain characteristics.

## RESULTS AND DISCUSSION

### Data Description

Based on the results of the analysis and collection of customer information data using documentation techniques on historical NPL data of micro-financing product customers obtained at PT Pegadaian (Persero) Makale Branch Service Unit, the following information data was obtained. The data in this study comprises the characteristics of a total of 114 customers with NPLs. All customers in the dataset are classified into three credit status categories: Special Mention (DPK), Substandard, and Non-Standard. Each observation contains demographic, socioeconomic, and loan characteristics variables relevant to NPL analysis.

#### 1. Customer Demographic Characteristics

- a) Gender. Customers in the dataset are relatively evenly distributed, with both males and females participating in all credit status categories, so NPLs are not dominated by one gender.
- b) Customer Age. Customers range in age from 19 to 63, reflecting that problem loans can occur in the young, productive age group to those approaching retirement. Most customers are in the 30-50 age group, which is the age group most active in economic activity.
- c) Marital Status. Customers fall into three marital status categories: married, unmarried, and divorced. This indicates that the risk of non-performing loans is not entirely influenced by household circumstances.
- d) Last Education. Customers' last education varied from elementary school, junior high school, high school/vocational school, to undergraduate level. The majority of customers had a high school/vocational school education, followed by undergraduate and junior high school education. This distribution indicates that NPLs can affect customers with varying levels of financial literacy.

#### 2. Socio-Economic Characteristics of Customers

- a) Customer Occupation. Customer occupations are dominated by two main categories: self-employed, which is the largest group, indicating that most customers earn income from independent businesses that tend to fluctuate. Farmers/livestock raisers are the second largest group, with relatively lower and unstable income levels according to harvest and production cycles. Customers with both self-employed and farmer/livestock jobs are found across all credit statuses, indicating that income instability contributes to the risk of non-performing loans.
- b) Monthly Income. Customer income ranges from Rp 9,000,000 to Rp 25,000,000 per month. Most customers earn between Rp 9,000,000 and Rp 12,000,000, while those earning above Rp 20,000,000 are predominantly self-employed.

#### 3. Loan Characteristics

- a) Loan Tenor. Loan terms vary between 12 months, 18 months, 24 months, and 36 months. The most common are 24-month and 36-month terms. Longer terms often increase the risk

of default because the longer the loan period, the greater the potential for changes in the customer's economic circumstances.

- b) Total Loans. Total customer loans range from Rp. 5,000,000 to Rp. 35,000,000, with the largest loans falling between Rp. 10,000,000 and Rp. 20,000,000. High-value loans (up to Rp. 35,000,000) are dominated by high-income customers, but they can still experience non-performing loans (NPL).

Customers are grouped into three credit status categories: special mention (DPK), substandard, and non-performing. Each credit status encompasses a diverse combination of characteristics, including demographics, employment, income, education level, and loan amount. Overall, the data shows that the characteristics of customers with non-performing loans are quite varied and come from various backgrounds. Demographic factors such as age, education, and marital status do not directly determine a particular credit status. However, the combination of income fluctuations, loan amount, and employment conditions appear to be the more dominant factors influencing the occurrence of NPL at Pegadaian (Persero) UPC Makale.

## Test Results

### 1. Data Preprocessing

Data preprocessing is a crucial step before data processing and analysis. This step aims to ensure that the data is in a suitable condition, consistent, and meets the requirements of the algorithm used in this study. The data preprocessing steps performed in this study are as follows:

#### a) Converting Categorical Data to Numerical Data

This process is carried out without changing the meaning of the categories that represent the characteristics of the objects. Data conversion aims to ensure that all variables can be calculated and processed computationally by the clustering algorithm. Categorical data, such as education level, type of employment, marital status, gender, and credit status, cannot be directly processed by the clustering algorithm because the algorithm can only process data in numeric form. Therefore, a conversion (encoding) process is carried out from categorical data to numeric data. The categorical data in this study is still divided into two types: nominal and ordinal categorical data, so the data conversion process uses the one-hot encoding technique. Thus, this transformation helps ensure that all information contained in the data can still be processed optimally in the clustering process.

#### b) Data Normalization

Data normalization is performed to standardize the value scale for each variable to avoid excessively large differences in value ranges between variables. This is important because clustering algorithms like the one used by the author (K-Means) use distance measures as the basis for determining similarity or proximity between objects. If a variable has a larger value range than other variables, then that variable will have a greater influence in the cluster formation process. This condition can cause bias and reduce the accuracy of the clustering results. By normalizing the data, all variables are converted to the same scale range (in this study, the researchers used a range of 0-1), so that each variable has a balanced contribution to the analysis process. Thus, normalization ensures that the clustering process produces a more objective and representative grouping of data characteristics. The results of data normalization using Orange Data Mining 3.38 in the Preprocess Widget (Normalize Features), show that all variables have been successfully normalized using the Orange Data Mining 3.38 normalization function where the highest value of each variable is transformed into a value of 1, the lowest value into 0, and the values between the two are in the interval 0-1 according to their proportion. Thus, this normalization result ensures that each variable has a balanced and proportional contribution in the cluster formation process.

### 2. Data Processing Using the K-Means Clustering Approach

After going through the data pre-processing stage, the next step is to carry out the clustering process. This process aims to group customers who experience NPL into several groups (clusters) based on certain similar characteristics according to the indicators that have been explained in this study. In this study, the algorithm used is K-Means Clustering, because this algorithm is an effective unsupervised learning method for grouping large amounts of data and is able to provide results that are easy to interpret. Based on the results of calculations using the Orange Data Mining 3.38 application, the determination of the number of clusters in this study was carried out using the Silhouette Score Method, which is a cluster evaluation technique in the K-Means algorithm that selects the optimal number of clusters based on the highest silhouette value. The results of determining the number of clusters will be used in clustering using the K-Means Cluster algorithm with the optimal number of clusters for each credit status that will determine the customer clusters within it.

### 3. Final Results of K-Means Clustering

The final stage of data processing presents the final results of the clustering process using the K-Means Clustering algorithm that has been carried out on customer characteristics data experiencing NPL at PT. Pegadaian (Persero) Makale Branch Service Unit. The purpose of this analysis is to identify and group customers based on the variable indicators in this study into a number of optimal clusters, where objects in one cluster have high similarities to each other, and significant differences with objects in other clusters, so that certain patterns that have the potential to influence the occurrence of non-performing loans can be identified more clearly. The selection of the number of clusters has been obtained from the highest silhouette score which is the most representative and stable configuration to describe the characteristics of NPL customers. The results of this cluster also serve as a basis for researchers in providing a comprehensive interpretation of the profile of each customer group, the final results of clustering in this study using the Orange Data Mining 3.38 application. Based on the results of data processing using the K-Means Clustering algorithm, three groups of credit status were obtained, namely DPK, substandard, and non-standard, where 3 main clusters (C1, C2 and C3) were obtained for DPK credit status, 2 main clusters (C1 and C2) for substandard credit status, and 2 main clusters (C1 and C2) for non-standard credit status. The three credit statuses with each cluster will represent the grouping of customer characteristics experiencing Non-Performing Loans (NPL). Each cluster shows a different pattern related to socio-demographic, economic, and loan characteristics variables. In general, the cluster results show that NPL customers are not homogeneous, but are divided into several risk profiles. Based on the analysis results from the Orange Data Mining Output 3.38, the following is an explanation of the results of the three credit statuses with each cluster:

#### 1) Special Attention Credit Status (DPK)

- a. *Cluster* Figure 1 (C1) shows a customer cluster with a demographic profile dominated by married women. The majority of customers in this cluster are self-employed, with the highest education level being high school. Despite having the best silhouette coefficient (the most homogeneous), this cluster has a low average normalized total loan, which is in the middle of the other clusters. Therefore, cluster C1 encompasses the characteristics of married self-employed customers with a high school education, who hold a medium portion of loans among DPK customers.
- b. *Cluster* 2 (C2) shows the customer cluster with the most distinct characteristics. The demographic profile in this cluster is dominated by men with the highest average age range. This cluster is the only cluster with a majority of customers as farmers/livestock raisers whose education level is predominantly elementary school. The total customer loans in this cluster are normalized to the lowest. Thus, cluster C2 includes the characteristics of male customers with an age older than the other two clusters, who work as farmers/livestock raisers and have an elementary school education with small-scale loans.

- c. *Cluster3* (C3) shows a customer cluster with a demographic profile dominated by unmarried men. The majority of customers in this cluster are self-employed with a predominantly bachelor's degree (0.67). Cluster C3 has the highest average income but has a low average total loan.

## 2) Substandard Credit Status

- a. *Cluster1* (C1) shows a customer cluster with significant substandard risk due to its loan value, where the demographic profile is dominated by women and the majority are married. The majority of customers in this cluster are self-employed with education ranging from high school to junior high school. This cluster has a low average income and relatively low total loans.
- b. *Cluster2* (C2) shows a cluster with more traditional risks and is similar to C2 in DPK. In this cluster (C2), the demographic profile is heavily male-dominated and tends to be in the relatively older age category. The occupational category is dominated by farmers/ranchers, with a predominantly junior high school education level. This cluster has a high average income with low total loans.

## 3) Non-performing Credit Status

- a. *Cluster1* (C1) shows a demographic profile dominated by relatively young men. Customers in this cluster have middle incomes and mostly work as farmers/livestock raisers. Customer loan terms in this cluster are relatively short with medium total loans, marital status is more varied, with customers spread across all marital statuses, and education levels are also varied, with customers spread across elementary, vocational, and junior high school levels.
- b. *Cluster2* (C2) shows customer characteristics with a demographic profile dominated by women of relatively young age and all working as self-employed. Customers in this cluster have low incomes with loan terms slightly longer than cluster C1, relatively high total loans with the majority being single, and the education levels of customers in this cluster vary, spread across undergraduate, high school, junior high, and elementary school levels.

## Discussion

The discussion in this study is structured based on the stages of data analysis that have been carried out, starting from data pre-processing to interpretation of test results as a basis for forming clusters of customer characteristics experiencing Non-Performing Loans (NPL).

### 1. The Relationship between Data Preprocessing and Algorithm Theory and Needs

In the data preprocessing stage, categorical data were converted to numerical data (encoding) and data normalization were performed. The results of the study indicate that categorical variables such as gender, occupation, marital status, highest education, and credit status cannot be directly used in the K-Means algorithm, requiring an encoding process. This research aligns with data mining theory, which explains that the K-Means algorithm is a distance-based algorithm that can only calculate distances between objects in numerical form (Han & Kamber, 2012). The conversion of categories to numbers in this study aims to maintain the information structure without changing the meaning of the categories. Preprocessing theory also emphasizes that data transformation is necessary to ensure data compatibility with analytical algorithms and maintain the integrity of the information to be processed. The encoding process is an important step to avoid algorithm bias that is unable to read discrete variables. The normalization results in this study indicate that all variables have been converted to a 0-1 scale. This is necessary so that variables with large ranges do not dominate the Euclidean distance calculation process. Normalization is an important step to provide a balanced variable contribution and normalization helps eliminate distortions due to non-uniform variable scales. (Han and Kamber 2012) (Prihastiwi et al., 2023). The results of normalization in this study are similar to previous research by (Shabila Ocktavia, 2025) on credit risk clustering, showing that normalization improves cluster accuracy and stability.

### 2. Discussion of the Profile of Each Cluster

The data grouping process using the K-Means Clustering algorithm produces clusters that represent the characteristics of customers experiencing Non-Performing Loans (NPL) at PT Pegadaian (Persero) Makale Branch Service Unit, both those with DPK status, substandard, and non-standard. These clusters show differences in risk patterns based on variations in demographic, economic, and customer loan characteristics. The following discussion explains the profile of each cluster and its interpretation of the risk of NPL.

### 1) Special Attention Credit Status (DPK)

#### a. Cluster1 (C1)

*ClusterC1* consists of 15 customers with a demographic profile characteristic pattern dominated by women (0.73), adult age (0.58), occupation dominated by self-employed individuals, low income (0.29), medium loan tenor (0.48), low total loan (0.2), marital status dominated by married individuals, and education level dominated by high school (0.87).

*Cluster*This indicates fairly homogeneous characteristics, with women in this cluster showing a preference for productive loans and being in the age category with the most stable financial needs. Their type of employment indicates a need to expand their business in the informal sector, where income is unstable. Loan tenors in this cluster show a fairly flexible preference, with marital status indicating greater financial needs, and secondary education levels, a common characteristic of entrepreneurs who should have sufficient knowledge to manage their businesses. This cluster depicts adult customers who are socially stable but have low incomes. This cluster profile tends to have good repayment commitments but limited financial capacity. In theory, credit management for this cluster is feasible but conservative, so credit must be in accordance with repayment capacity. Therefore, restructuring intervention is necessary, as is handling and monitoring.

#### b. Cluster 2 (C2)

*ClusterC2* consists of only 4 customers with a demographic profile characteristic pattern dominated entirely by male customers, adult age (0.53), occupation dominated entirely by farmers/livestock breeders, low income (0.1), relatively low loan tenor (0.44), low total loan (0.09), balanced marital status between unmarried (0.5) and married (0.5), and the last education dominated by vocational school (0.75).

Cluster C2 is a very small group of farmers/livestock farmers with seasonal income risks. Despite its small size, the pattern it forms is very distinctive compared to the other two clusters, where this cluster is entirely dominated by men, which fits the characteristics of agricultural work, which also indicates a productive age in the agricultural sector. Having a seasonal and fluctuating income pattern, along with the lowest income, indicates limited cash flow. The loan tenor and total loan amount are relatively low due to the consumption patterns of the agricultural sector, which generally take small loans for urgent needs or seasonal capital. Balanced marital status indicates no particular social dominance influencing borrowing behavior. Cluster C2 can be categorized as a moderate-high risk customer due to seasonal income, low loan values, and short tenors indicating short-term needs. Therefore, a more personalized and educational approach is needed, or a special approach such as scheduling installments based on the harvest season.

#### c. Cluster3 (C3)

*ClusterC3* consists of 9 customers with a demographic profile characteristic pattern dominated by men (0.67), young age (0.13), occupation dominated by self-employed individuals, highest income compared to the other two clusters (0.32), and loan tenor which is also longer than the other two clusters (0.5), low total loan (0.21), marital status dominated by unmarried (0.78), and last education dominated by Bachelor's degree (0.67).

*ClusterC3* represents young customers with the highest income capacity and higher education, dominated by men. This age group is oriented towards business development and capacity expansion. Occupations are predominantly self-employed, similar to Cluster C1,

but with different dynamics due to younger age, high income, long loan tenure, and unmarried status, which predominantly reflect more flexible financial conditions and lower household expenditure risks. The education level in this cluster also indicates a higher level of financial literacy, potentially improving the quality of loan management. Despite being categorized as a DPK credit status, Cluster C3 represents young, productive customers with potential for growth. With the highest income and higher education, this cluster has the opportunity to be saved, long-term business orientation, and has better adaptation capacity and risk management. Therefore, it is necessary to develop potential customers into prospective customers for larger financing services, simply with strict monitoring and short-term solutions.

## 2) Substandard Credit Status

### a. Cluster1 (C1)

*ClusterC1* in substandard credit status consists of 43 customers with a demographic profile characteristic pattern dominated by women (0.72%), young age (0.33), customer occupations are entirely dominated by self-employed, low income (0.38), relatively low tenor (0.4), low total loans (0.25), marital status is predominantly married (0.56), and the last education is predominantly high school (0.56).

The characteristics of this cluster indicate that low-income customers working in the self-employed sector are at higher risk of being classified as substandard. The predominance of women and their relatively young age indicate this customer group is in the early stages of household economic development. Although the total loan amount is low, risks remain due to income instability, not the loan amount. Short tenors also typically indicate a relatively rapid repayment capacity, potentially leading to late payments for customers with unstable cash flow.

With the main risk characteristics of unstable income, young age and informal micro-enterprises, business mentoring and simple financial management training, light rescheduling (adjustment of maturity), regular monitoring and intensification of communication are required, and credit products with flexible installment schemes can be provided.

### b) Cluster 2 (C2)

*ClusterC2* in substandard credit status consists of 9 customers with a demographic profile characteristic pattern dominated by men (0.67), adult age (0.56), customer occupation is entirely dominated by farmers/livestock breeders, relatively high income (0.51), low loan tenor (0.33), low total loan (0.27), marital status is dominated by married (0.89), last education is dominated by junior high school (0.51).

The characteristics of customers in this cluster indicate that, despite their relatively higher incomes compared to C1, the risk of substandard loans remains. Customers working as farmers/livestock raisers have strong nominal incomes, but are still affected by seasonality, weather, and commodity price fluctuations. The majority of customers in this cluster are adults and married, indicating the presence of family dependents that can influence income allocation. A relatively low level of education (junior high school) also has the potential to impact financial planning. Compared with cluster C1, this cluster's risk is driven more by income volatility, rather than income level or loan size. The main risk characteristics in this cluster are seasonal income, low financial literacy, and a vulnerable agricultural sector.

Therefore, it is necessary to adjust the installment schedule according to the harvest season (seasonal repayment), provide education on financial literacy and harvest planning, diversify agricultural financing products, and strengthen collateral substitution.

## 3) Non-performing Credit Status

### a. Cluster1 (C1)

*Cluster*C1 in non-performing credit status totaled 11 customers with a demographic profile characteristic pattern dominated by men (0.82), relatively young age (0.35), occupation dominated by farmers/livestock breeders (0.64), relatively middle income (0.5), low loan tenor (0.39), medium total loans (0.5), marital status spread evenly among the three categories (unmarried, married and divorced), education level also spread across the elementary, vocational and junior high categories.

*Cluster*This shows that farming/livestock farming has once again emerged as a dominant factor in credit problems. This reinforces the findings in the substandard cluster that agricultural income is highly vulnerable to external fluctuations. However, in this cluster, the risk increases because the total loan amount is larger than in the substandard cluster, and education is also lower, with variations in marital status. Risks are not only related to family responsibilities but also to the individual's socio-economic stability. As is known in microcredit, customers with high income uncertainty and low financial literacy are at higher risk of default. The main risk characteristics in this cluster are seasonal income, medium-sized loans, and a relatively heavy installment burden.

Therefore, more formal credit restructuring actions (Rescheduling or Reconditioning), assistance in increasing agricultural productivity, group-based credit schemes (group lending), and more intensive field visits are needed.

b. *Cluster*2 (C2)

*Cluster*C2 in non-performing credit status consists of 23 customers with a demographic profile characteristic pattern dominated by women (0.61), relatively young age (0.34), occupation dominated entirely by self-employed, relatively low income (0.41), relatively low loan tenor (0.44), high total loan (0.96), marital status dominated by unmarried (0.56), last education dominated by high school (0.74).

*Cluster*This cluster has the highest risk profile based on a combination of factors: very high total loans, making it highly sensitive to income shocks. Low incomes result in a high debt-to-income ratio, a classic indicator of credit risk. The low-income, unstable income segment of self-employed micro-entrepreneurs poses a higher risk of default. Young and unmarried individuals pose a potential for economic instability and lack of financial experience. A high school education is insufficient to buffer risk when the loan burden significantly exceeds income. Therefore, this cluster reflects the characteristics of borrowers experiencing over-indebtedness, which is theoretically a strong predictor of non-performing loans.

The main risk characteristics are over-indebtedness, unstable cash flow, large installment burdens that require comprehensive restructuring measures (can be by reducing installments and extending the tenor), re-evaluation of business capabilities and potential for recovery, prohibition on providing additional loans (temporary credit freeze), mandatory financial counseling, strict supervision and an early warning system.

## CONCLUSION

Based on the results of the research and discussion that have been conducted in this study, the K-Means algorithm used in the clustering process successfully identified the characteristic patterns of customers who experienced Non-Performing Loans (NPL) at PT Pegadaian (Persersero) Makale Branch Service Unit. The main conclusion that can be drawn in this study is that the K-Means algorithm successfully mapped NPL customers into three credit status groups with 3 clusters in DPK credit status, 2 clusters in substandard credit status, and 2 clusters in non-standard credit status. The explanation of the conclusions of this study is as follows:

1. The K-Means algorithm successfully formed three clusters of NPL customers with different DPK credit statuses, each with different characteristics and risk levels. These clusters consist of:
  - a) *ClusterC1* (adult, female, self-employed, low income). Clients are socially stable but financially weak. The primary risk stems from low income. This group is still worth maintaining with close monitoring.
  - b) *ClusterC2* (small cluster, men, farmers/herdsmen, very low income). This group is at moderate to high risk due to seasonal and fluctuating income. Risk arises from unstable cash flow, even with small loans.
  - c) *ClusterC3* (young male, highest income, higher education, self-employed). The most potential cluster with low-medium risk and growth prospects.
2. The K-Means algorithm successfully formed two clusters of NPL customers with substandard credit status, each with different characteristics and risk levels. These clusters consist of:
  - a) *ClusterC1* (women, young, self-employed, low income). This cluster is most vulnerable due to its informal micro-enterprises, fluctuating income, and short tenors. Risk arises from unstable cash flow, not loan size.
  - b) *ClusterC2* (male, adult, farmer/rancher, moderate but seasonal income). The main risks lie in dependence on the harvest season, low financial literacy, and family responsibilities. Even with sufficient income, volatility leads to late payments.
3. The K-Means algorithm successfully formed two clusters of NPL customers with non-performing loans with different characteristics and risk levels. These clusters consist of:
  - a) *ClusterC1* (male, farmer/herdsman, medium-sized loans, seasonal income). High risk due to higher loan burdens, seasonal income, and low literacy. This cluster reflects the structural vulnerability of the agricultural sector.
  - b) *ClusterC2* (young women, self-employed, low income, with very high total debt). The highest-risk group across all categories. They experience over-indebtedness and unstable cash flow, making them highly vulnerable to permanent default.
4. The results of this study prove that the data mining approach is able to provide strategic insight in identifying risk patterns, so that it can be a basis for pawnshops in formulating NPL mitigation policies more effectively and on target.

### Research Limitations

This research has several limitations that need to be considered as a basis for developing further research, namely:

1. The sample size used in this study was limited. The data used was only 114 NPL customers from one service unit, PT. Pegadaian (Persero) UPC Makale. This relatively small number may not represent the full range of pawnshop customer characteristics more broadly.
2. The variables used are still limited. The variables analyzed only include socio-demographic characteristics and loan data. Other important factors such as previous credit history, repayment behavior, family economic conditions, and psychological factors could not be analyzed due to data limitations.
3. External validity testing has not been carried out on the cluster results where the clustering model was only tested using internal customer data and has not been compared with data from different periods or other branch units.

### Suggestion

Based on the research results and discussions obtained by the author, several suggestions can be used as recommendations for both further research and relevant agencies. The following are the suggestions the author can offer:

1. There is a need to strengthen risk management based on cluster characteristics, not just credit status, and with monitoring and early warning systems implemented more strictly, especially in seasonal income clusters and low-income self-employed individuals.
2. Improving financial literacy for self-employed and farmer customers to reduce the risk of default, while segmenting credit products so that the tenor, installment pattern, and loan amount match cluster characteristics.
3. For third party fund (DPK) credit status, C1 requires regular monitoring, conservative credit provision, and offering light restructuring to maintain smooth operations. C2 requires implementing a seasonal installment scheme and intensive education due to seasonal income risks. C3 requires leveraging potential through greater productive financing while still providing business management support.
4. For substandard credit, C1 requires business mentoring, financial management training, easy rescheduling, and flexible installment plans. C2 requires the implementation of harvest-season-based repayment, agricultural business planning education, and a more comprehensive agricultural risk evaluation.
5. For non-performing loans, formal restructuring (rescheduling/reconditioning), farming assistance, and intensive field visits are required for C1. Comprehensive restructuring, a temporary credit freeze, mandatory financial counseling, and strict monitoring are required for C2 due to the risk of over-indebtedness.
6. For further research, it is recommended to use a larger amount of data, research areas, and add more diverse variables.
7. Further research can use other clustering algorithms, such as Hierarchical Clustering or DBSCAN, to compare with the performance of K-Means and see whether there is a more optimal cluster pattern formation.
8. Integrating the results of data mining clustering into the risk management system or business processes of PT Pegadaian (Persero) UPC Makale, so that the creditworthiness assessment process can be more data-driven, not only relying on manual assessments, or in other words, so that credit granting decisions are not only based on intuition but based on data patterns.

## REFERENCE

- Achmad Fauzi. (2018). The Role of Credit Analysts on NPL at PT X. *JMBI UNSRAT (Scientific Journal of Business Management and Innovation, Sam Ratulangi University)*., 5(2), 9.
- Ameylia Yesica Yosefani Dua Bura, Maria Nona Dince, & Emilianus Eo K. Goo. (2023). Analysis of Credit Granting Procedures at KSP Kopdit Pintu Air, Koting Branch. *Student Scientific Creativity Journal*, 1(1), 376–389.
- Amir, A., Achmad, A., & Faizin, N. (2023). Analysis of Financing Provision Procedures at the East Java BMT NU Asembagus Situbondo Branch Based on the 5C and 7P Principles. *Mazinda: Journal of Accounting, Finance, and Business*, 1(1), 35–41.
- Andrian.Asiva, Noor, Rachmayani.,Sarjana, Sri., et al. (2015). Risk Management. *Bahtera* (2022) 2(1)
- Arifah, Odi Nur., (2017). Analysis of Problematic Mudharabah Financing at BMT Mitra Hasnah Semarang. Vol. 7. 40-46.
- Aufa, A., Putri, A., Rahmah, SA, Informasi, T., Teknik, F., & Dharmawangsa, U. (2024). Implementation of Data Mining with K-Means Clustering Algorithm for Business Analysis in Insurance Companies. 5 (1), 139-152.
- BaRiansori, O., & Wahyuningsih, SD (2018). Analysis of the 5C Principle Assessment in Credit Granting to Non-Performing Loans to Assess the Health Level of Banks at PT BPR Harta Swadiri Pandaan. *Applied Management Research (Penataran)*, 3(1), 54–63.

- Business, J., & Dan, M. (2024). *Entrepreneur*. 5, 31–46.
- Darmawi, Herman. (2016). *Risk Management*. (ed); 2. Bumi Aksara. Jakarta. 225.
- Djuarni, W. (2022). Implementation of the 5C Principles in Determining the Eligibility of Providing Credit to Customers Wenny Djuarni 1,. *Finance and Banking*, 02(02).
- Eprianti, N. (2019). Application of the 5C Principle to the Level of Non-Performing Financing (NPF). *Amwaluna: Journal of Islamic Economics and Finance*, 3(2).
- Fegiriyanti, Y., Kamal, B., & Krisdiyawati. (2021). Analysis of Credit Granting Procedures to Reduce NPL (Non-Performing Loans) at BTPN Syariah Slawi Branch. 1–9.
- Hamonangan. (2020). Analysis of the Application of the 5C Principle in Financing Distribution at Bank Muamalat KCU Padangsidempuan. *MEA Scientific Journal (Management, Economics, and Accounting)*, 4(2), 454–466.
- Handriano, R. (2018). Analysis of Credit Distribution Procedures in an Effort to Reduce the Risk of Problematic Credit in State-Owned Banks in Banjarmasin City. *Economic Dynamics Journal of Economics and Business*, 11(2), 230–244.
- Islami, V., & Islami, V. (2019). Analysis of Home Ownership Credit (KPR) Provision in Order to Reduce Non-Performing Loans (Case Study of PT. Bank Tabungan Negara Tbk. Bogor Branch). *Monetary - Journal of Accounting and Finance*, 6(1), 1–6.
- Journal, CD, Yakub, R., Fajar, M., Hamka, B., Salsabila, NA, Syampturi, E., Kredit, AK, & Kredit, R. (2024). Analysis of the Process of Providing People's Business Credit (KUR) to Minimize NPL (Non-Performing Loan) Customers at PT. Bank Sulselbar Sidrap Branch. 5(5), 9811–9818.
- Jusuf, Jopie. (2010). *Credit Analysis for Account Officers*. Gramedia Pustaka Utama, Jakarta. Volume 4.
- Lestari, Hidayah.T. (2021). Analysis of the Character of Prospective Financing Customers (Study at BSI KCP Ponorogo).
- Manurung, AH (2012). Behavioral Finance Theory. *Economics of Management*, 41(4), 1–13.
- Mariana. (2022). *Accounting Information and Credit Decisions*
- Muhammad. (2002). *Islamic Bank Management*. April 2020.
- Muniarty, Puji., Abbas, Surya, Dirvi., et al. (2020). *Banking Management*. Widina Bhakti Persada Bandung.
- Muryono, Tupan. T., Irwansyah. (2020). Implementation of Data Mining to Determine Credit Eligibility Using the K-Nearest Neighbors (K-NN) Algorithm. *Technology Information*. Vol.6.
- Nugroho, JA, Suardi, Lenny. (2024). Survival Model Approach to Analyze the Event of Current Debtors Becoming Pre-NPL in the Small Credit Segment (Case Study of PT. Bank XYZ Tbk.). *MATHunesa*. Vol. 12. No. 03.
- Octavia, Evi. (2021). Analysis of the Credit Granting Process to Reduce the Risk of Bad Debt at PT. Bank Artha Graha Internasional Tbk Bandung. *Journal of Business and Economic Accounting*. 6(2), 1719-1738.
- Oktavia, Y. (2022). *Economics and Business*, Islamic University of Riau. 30–31.
- Pardjo, Y. (2017). *Practical Guide to Enterprise Risk Management (TG Publishing (ed.); First)*. Growing Publishing.
- Pebriyani, Putri Sahade S, Masnawaty.(2024). Analysis of Credit Granting Procedures at the Berkas Savings and Loan Cooperative (KSP) Tamalate Branch, Makassar City. *Future Academia: The Journal of Multidisciplinary Research on Scientific and Advanced*, 2(4),349-361.
- Pipit Mulyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, T. (2020). No Title No Title No Title. *Journal GEEJ*, 7(2).
- Prihastiwi, DA, Theres, A., Widyastuti, TD, Rosaria, D., Nurlaela, L., Murti, GT, Rosmawati, A., Lestariningsih, M., Dano, D., Anggrayni, L., & Ungkari, MD (2023). *Company*

- Management (C. Edi Pranyoto, SE., MM., CISMA. (ed.); 1st ed.). EUREKA MEDIA LITERATURE.
- Rizky, Muhammad., Umam, Muhammad. IH, Hamzah, Muhammad. L. (2020). Data Mining Application Using the CHAID Method in Determining Credit Status. *Technology and Industry*. Vol. (18). 29-33.
- Rusydi, F. (2023). *Banking Management*.
- Salamah, Ade. (2023). Prudential Banking Principles in Sharia Financing to Reduce the Risk of Non-Performing Loans. *Veritas*, vol. 9. pp. 97-110.
- Shabila Ocktavia1, WTA (2025). Analysis of Financing Customer Segmentation Based on Demographics to Predict Credit Levels Using the K-MEANS Algorithm. *Journal of Information Technology and Science (JINTEKS)* E-ISSN: 2686-3359 Submission Feb 27, 2025 | Revised Mar 05, 2025 | Accepted March 09, 2025 Pg. 394-402 Vol. 7; No. February 1, 2025 ANALYSIS, 7; No. 1, 394–402.
- Shiyamurti, N, R. (2020). *Accountia Journal*. *AccountiaJournal*. 6(2), 31.
- Tengor, RC., Murni, S., & Moniharapon, S. (2015). Implementation of Risk Management to Minimize the Risk of Bad Debt at PT. Bank SulutGo. *EMBA Journal*, 3(4), 345–356.
- Wahyuni, N. (2017). Application of the 5C Principle in Credit Provision as Bank Protection. *Lex Journal: Legal & Justice Studies*, 1(1).
- Widyastuti. (2021). Behavioral Finance in the Decision-Making Process. *LMFE Unpad*, 1–15.
- Wiryaningtyas, DP (2016). Behavioral Finance in Decision Making. *UNEJ E-Proceeding*, 339–344.
- Wunasari., Priatna, Nugraha., Mustika, Wida Prima. (2018). Analysis of Housing Loan Feasibility Using the NAIVE BAYES Classification Method. *Inti Nusa Mandiri*. Vol. 12. 21-28.
- Zahra Yuniar, D., Suherman, E., Epty, D., Buana, U., & Karawang, P. (2022). Analysis of Non-Performing Loans at PT Bank BRI Tbk Article Info Abstract Article history. *Scientific Journal of Accounting and Finance*, 5(5), 2022.