www.bircu-journal.com/index.php/birci
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)
Budapest Institute

Implementation of Data Mining to Determine Payment Delays for Mall Shopping Center Tenants Using K-Means Clustering Method

Ary Suryo Bimantoro[1], Rafi Athallah[2], Naufal Ifsan Al-Azhar[3], Nilo Legowo[4]
[1,2,3,4]*Information Systems Management Department, Binus Graduate Program-Master of Information Systems Management, Universitas Bina Nusantara, Indonesia*
ary.bimantoro@binus.ac.id, rafi.athallah@binus.ac.id, naufal.alazhar@binus.ac.id, nlegowo@binus.edu

Abstract

With the rapid growth of their business, XYZ Company wanted to utilize their data to the upmost maximum. One of the methods to make the most out of their data is to do data mining. With the available data and RapidMiner as processing tool, researcher can build a clustering model to determine the accuracy of payment for their clients. It was discovered that the dataset can be divided into 3 clusters, namely on time, late, and very late. From the cluster discovered, a suggestion can be made on how to handle the payment for each group, so that there will be no more late payment in the future by applying a penalty for the tenants that are late paying their bills.

Keywords
data mining; clustering; CRISP-DM; prescriptive analysis; k-means

I. Introduction

In the current digital era, technologies are developing very quickly following the development of science. Data is an important element in the use of systems to support a company business. With the development of methods to process data into strategic information for companies, the company need a method that can extract data to become a new knowledge for the company. XYZ Retail is a company engaged in retail space rental, In order to obtain steady sources of supply and demand and to maximize their mutual profits, many vendors and clients would prefer to develop long-term cooperative relationships in a highly competitive commercial market (Chen & Kang, 2010). The company currently undergoing digital transformation in its IT department. As with other areas of life, technology is used to make changes, so also with the legal system as technology in making changes (Hartanto, 2020). The problem is that management is unable to optimize data, which prevents the organization from making the best use of a lot of potentially important information. The issue that is frequently seen is tenant who frequently pay the building's rent late. This can lead to new issues down the road, such as recording errors, as well as the potential loss of revenues because there are no clear standards about rent payments.

What is being done is to develop a strategy based on data from the mall's operational activities. It's main business activity is to provide a place for tenants to sell in the mall area, the IT division wants to take advantage of operational data by utilizing data mining methods that produce prescriptive analysis to assist company decision making. The prescriptive analysis approach is utilized by the firm since the descriptive analysis method is regarded insufficient to meet the objectives of the company's top level management, who desire a system that can forecast what will happen in the future and decide the activities that must be taken to achieve company and strategic goals.

The analysis of the tenant's payment delay needs to be done in depth in order to obtain the pattern of each tenant payment behavior and to assist XYZ Retail in making decision on what needs to be done to the tenants that have been neglecting their responsibilities to the company. The right approach in this case is by using the Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM is a method that is easy to apply in this sort of case because every phase or stage is well defined and structured with a complete and well-documented data mining methodology (Khumaidi, 2020).

II. Review of Literature

2.1 Business Intelligence

Business intelligence (BI) is a process that uses various data, information, and knowledge owned by the organization as raw material in the decision-making process to boost the company's competitive advantage (Miranda, 2008).

Business Intelligence (BI) is a process to increase the company's competitive advantage through the utilization (Zhang et al., 2021), The Business Intelligence method enables to disclose reserves, increase sales, reduce costs and achieve higher profits within(Václav et al., 2021). BI, an analytical process supported by technology, collects and turns fragmented data from businesses and marketplaces into knowledge or information about the goals, opportunities, and capabilities of an organization (Wieder & Ossimitz, 2015).

Unlike a number of other applications with similar objectives that were previously introduced, the BI concept emphasizes the application of 5 information functions for specific business purposes, which are:

1. Data Sourcing: Relates to the ability of the system to access various data and information from a number of sources with different formats.
2. Data Analysis: Related to the ability of the system to perform analysis and information owned by the company or agency with the aim of assisting the knowledge creation process.
3. Situation Awareness: Related to the ability of the system to be able to find and provide data and information that is relevant to business needs at a certain time.
4. Risk Analysis: Related to the ability of the system to analyze and calculate the risk ratio that can be faced by the company in relation to certain conditions.
5. Decision Support: Related to the ability of the system to assist company management in providing recommendations for quality business decisions by taking into account internal and external data and information that have been calculated and processed.

2.2 Data Mining

Data mining is a method for calculating massive data sets to find patterns and extract relevant information (Gladju et al., 2022). According to (Sikumbang, 2018), data mining is a method that utilizes one or more computer learning techniques (machine learning) to automatically analyze and extract knowledge. Data mining is an iterative and interactive process that explores a big database for new patterns or models that are excellent, useful, and understandable (massive database) (Larose & Larose, 2014). Data mining is an iterative and interactive process to find new patterns or models that are perfect, useful and understandable in a very large database (massive database) so it can used to make very important business decisions.

Data mining is divided into several groups based on the tasks that can be performed, namely Description, Estimation, Prediction, Classification, Clustering, and Association.

There are various learning methods in the Data Mining algorithm, namely (Damayanti, 2006):

1. Supervised Learning: Most data mining algorithms (estimation, prediction/forecasting, classification) are supervised learning. The target variable/label/class is determined, the algorithm performs the learning process based on the value of the target variable associated with the value of the predictor variable, to be able to predict and explain the value of a target variable is the purpose of Supervised Learning (Laperrière-Robillard et al., 2022).
2. Unsupervised Learning: Data mining algorithms look for patterns from all variables (attributes). The variable (attribute) that is the target/label/class is not specified (none), the clustering algorithm is an unsupervised learning algorithm, Unsupervised learning- based techniques demand a lot of work during model training and selecting the suitable threshold (Choi et al., 2022).
3. Association Learning: The learning process in the association rule is somewhat different because the goal is to find attributes that appear together in one transaction. Usually used for shopping transaction analysis to find out the items purchased simultaneously. In a shopping center that has many search products that require high costs, the A Priori Algorithm can solve this problem efficiently

2.3 Prescriptive Analysis

Prescriptive Analytic utilizing data and mathematical formulas, prescriptive analytics suggests precise actions (Brandt et al., 2021), Prescriptive analytics, which is both more advanced and less developed, leverages the results of other analytics to take the best possible action (Mosavi & Santos, 2020). Prescriptive analytics evaluate the results of several choices, enabling the choice of the most appropriate current course of action(Tinoco et al., 2021). Prescriptive analytics is the most complex sort of business analytics and may provide firms with the most intelligence and value. Its goal is to advise (prescribe) the optimal decision possibilities in order to capitalize on the expected future using massive volumes of data (Šikšnys et al., 2016).

2.4 CRISP-DM

According to the research of (Schröer et al., 2021), CRISP-DM is a data mining process model that is independent from the industry. From business knowledge to implementation, there are six iterative phases (see Table 1). Based on the CRISP-DM user guide, Table 1 briefly describes the core ideas, tasks, and outputs of these phases

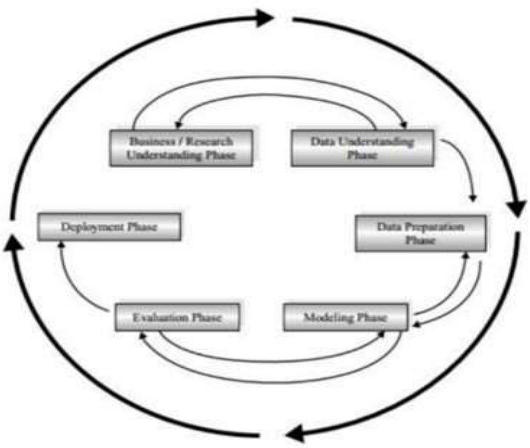

Figure 1. *Crisp DM Methodology Source: (Larose & Larose, 2014)*

1. Business Understanding: To determine the resources that are available and those that are needed, the business's existing situation should be investigated. Choosing the data mining objectives is one of the most crucial elements of this process. First, it is important to describe the type of data mining (such as classification) and the success criteria for data mining (such as precision). It is essential to create a project plan.
2. Data Understanding: This step requires collecting data from multiple sources, investigating and summarizing it, and verifying the quality of the data. In other words, the user guide explains how to perform data description tasks by performing statistical analysis and defining attributes and their order.
3. Data Preparation: In order to complete this phase, data must be gathered from various sources, investigated, summarized, and the accuracy of the data must be confirmed. In other words, the user guide describes how to perform statistical analysis, define attributes, and arrange them in order to describe data.
4. Modelling: The construction of test cases, models, and the choice of modeling techniques are all necessary steps in the data modeling process. Any data mining strategy is acceptable. The choice is supported by data in general and business difficulties specifically. How you explain your choice is more critical. Before the model can be built, several parameters must be established. When assessing a model, it's a good idea to compare it to other models.
5. Evaluation: The outcomes are compared to the established business objectives during the evaluation step. As an outcome, it is important to analyze the findings and specify the next steps. Another consideration is that a thorough examination of the process is necessary.
6. Deployment: The user guide provides a broad description of the deployment phase. This might be a software module or a final report. The deployment phase, according to the user guide, involves planning, monitoring, and maintenance.

2.5 Rapid Miner

is a user-interactive setting for operations like data mining and machine learning. It is a free, open-source project that was developed in Java. It is a modular operator concept that enables the building of intricate stacked operator chains for a wide range of learning problems, representing a modular approach to designing even very complex issues (Naik & Samant, 2016). To help customers get insights and make the best decisions, RapidMiner employs a number of descriptive and predictive algorithms. (Aprilla Dennis, 2013).

2.6 Related Works

Table 1. Related Works

Author	Title	Conclusion
(Fan & Xiao, 2017)	Assessment of building operational performance using data mining techniques: a case study	This paper describes the operational performance of the building. The case was taken at the university building in Hong Kong, using the decision tree data mining method. the result is a pattern of operation and energy conservation opportunities.

(Mauritsius et al., 2019)	Bank marketing data mining using CRISP-DM approach	In this paper, it is explained that to perform data mining on the Bank's marketing, the CRISP-DM framework is used by using logical regression algorithms and multilayer perceptron. This paper aims to create a predictive model that can label data related to bank marketing into 2 predefined classes, namely Yes and No.
(Achenbach & Spinler, 2018)	Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights	In this paper, it is explained that the use of data mining and prescriptive analytics can be used to create models that can assist airlines in predicting flight arrival times so that they can optimize the cost index, especially on short-haul flights. This study found that the optimal cost index level is highly dependent on flight distance, fuel costs and delay costs.
(Desineedi et al., 2020)	Developing driving cycles using k-means clustering and determining their optimal duration	In this paper, it is explained that the use of data mining using K-Means clustering can be used to create a grouping of driving cycles data using travel time to better understand driving patterns and vehicle emission estimates. This study uses K-Means to classify micro-trips and Markov modelling for cluster sequencing.
(Choudhari & Potey, 2018)	Predictive to Prescriptive Analysis for Customer Churn in Telecom Industry Using Hybrid Data Mining Techniques	This paper describes using a hybrid classification technique, namely decision tree and logistic regression to predict customer turnover in the telecommunications industry. In addition, by using this technique, prescriptive analysis can be carried out to overcome the significant decline in revenue due to customers leaving the company.

III. Research Method

This research will use CRISP-DM methodology, this strategy offers a life cycle approach for projects including applied artificial intelligence (Solano et al., 2021). as stated in the previous chapter, the flow of CRISP DM is consisting of:

1. Business Understanding: Understanding the purpose of the company and the business that the company are operating to better visualize the deployment method after evaluating all data.
2. Data Understanding: Ensure the dataset which is the source of information of the tenant is accurate and complete in order to prevent miscalculation from the data mining process.
3. Data Preparation: Prepare the data in Rapid miner, during this process there are some adjustments regarding the dataset, such as change data type, and remove blank row.
4. Modelling: Create data model to desired result, in order to obtain valuable information, there are various methods such as clustering to split the customer into different cluster.
5. Evaluation: Explore the dataset furthermore by analyzing the result of modelling which are the cluster of tenants who are late paying rent and tenants who pay the rent on time.

6. Deployment: Make action plan to reduce the number of tenants that are late with paying the rent in the future.

The output of each CRISP-DM steps will be discussed in the following chapter, with a simple visualization to accompany the explanation.

IV. Discussion

4.1 Business Understanding

- Determine Business Objectives: The purpose of this business goal is to find out tenants who are late paying at XYZ Mall, so that follow-up actions can be carried out on these tenants
- Assess the Situation: The application discussed in this project is XYZ Mall tenant payments which are used for payments to both tenants in one mall, this application is used to record payments made by tenants
- Determine Data Mining Goals: The purpose of applying data mining to the XYZ mall payment application is to find out the number of tenants who are late paying, this number of tenants will be divided into several clusters for classification according to the payment time made by tenants as a basis for decision making

4.2 Data Understanding

- Data Source: The source of this dataset is obtained through the XYZ Retail tenant payment application database. This data is extracted from SQL data and then will be exported to Excel format. The amount of data in this dataset reaches approximately 3103 data, with a fairly good data quality because the data in this dataset has been completely filled and sorted within a certain period. This data consists of 9 attributes, each of which describes various information about tenant payments. By looking at the amount of data obtained with the business intelligence project to provide input to the mall as a basis for making decisions regarding units and tenants renting at the mall. So it can be said that the data owned is sufficient to be a reference in this project.
- Data Dictionary:

Table 2. Data Dictionary

No	Attribute	Data Type	Field Size	Example	Description
1.	Kode Cabang	Varchar	4	PBTO	Branch code
2.	Kode Tenant	Integer	5	00054	Tenant code
3.	Kode Tipe Bisnis	Integer	5	40501	Business type code
4.	Luas Booth	Float /Integer	10	45	Area rented by tenant
5.	Kode Tipe Pembayaran	Varchar	2	60	Payment type code
6.	Jumlah Biaya Tagihan	Integer	20	12150000	Billing amount
7.	Jumlah Hari Terlambat	Varchar	5	173	Number of days late from due
8.	Ketepatan Bayar	Varchar	15	Terlambat	Payment accuracy
9.	Cluster	Integer	1	1	Cluster of payment

4.3 Data Preparation

- Data Pre-Processing: At the data pre-processing stage, our group got a fairly complete and large dataset. This led us to ensure the data to be used and complete. Because it was discovered that there are 22 empty rows, therefore it is necessary to delete the empty rows using remove value. Another reason is that there was an error in inputting data by Retail company XYZ. In this pre-processing process, some attributes that are still of polynomial type are converted to numerical to make data processing easier. Pre-processing of this data is done through several software according to the functions provided by each application.

Table 3. Dataset After Pre-Processing

No	Attribute	Data Type	Field Size	Example	Description
1.	Kode Cabang	Varchar	4	1,2,3,4	Branch code
2.	Kode Tenant	Integer	5	1	Tenant code
3.	Kode Tipe Bisnis	Integer	5	1	Business type code
4.	Luas Booth	Integer	10	45	Area rented by tenant
5.	Kode Tipe Pembayaran	Integer	2	1,2,3	Payment type code
6.	Jumlah	Integer	20	12150000	Billing amount
	Biaya Tagihan				
7.	Jumlah Hari Terlambat	Integer	5	173	Number of days late from due
8.	Ketepatan Bayar	Integer	15	1,0	Payment accuracy
9.	Cluster	Integer	1	1,2,3	Cluster of payment

4.4 Modelling

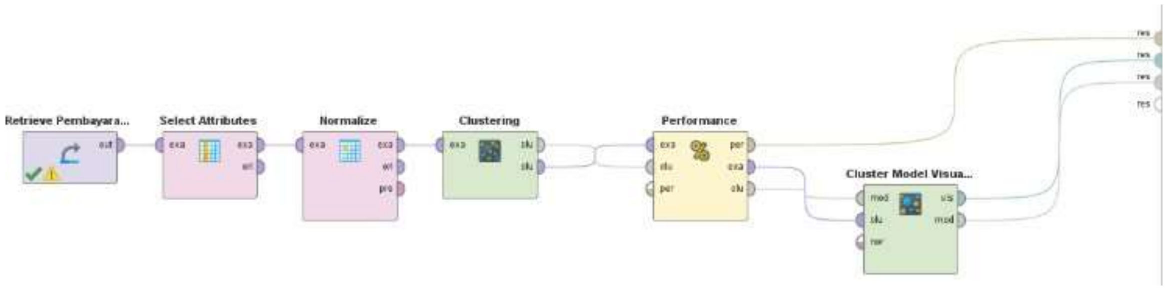

Figure 2. *Data Modelling*

The picture above is a data mining model generated by utilizing several operators in such as date to numerical, select attribute, clustering, cluster model visualizer, and performance.

- Select attribute: Used to determine what data is used for the modelling process
- Clustering: This operator is a clustering algorithm that is used to create several clusters from existing datasets
- Cluster model visualizer: This operator is used to view cluster details generated by the k-means algorithm, some of the information displayed is a cluster overview, heatmap, centroid chart, centroid table, and scatter plot
- Performance: This operator is used to see the capabilities of the model that has been made, some of the information displayed is
- Normalization: This operator is used to perform value scaling so that the values in the dataset can be divided into several ranges.

4.5 Evaluation

- Dataset Exploration

Row No.	Kode Tenant	Kode Cabang	Kode Tipe B...	Luas Booth	Kode Tipe P...	Jumlah Bay...	Jumlah Hari ...	Ketepatan B...	Cluster
1	1	8	0	45	3	12150088	173	3	0
2	1	8	0	45	2	8850000	35	3	1
3	1	8	0	45	3	8300000	17	3	1
4	1	8	0	45	2	8850000	21	3	1
5	1	8	0	45	3	8379328	1	3	1
6	1	8	0	45	2	8850000	17	3	1
7	1	8	0	45	3	7841000	25	3	1
8	1	8	0	45	2	8850000	35	3	1
9	1	8	0	45	3	8772161	16	3	1
10	1	8	0	45	2	8850000	29	3	1
11	1	8	0	45	3	8877848	10	3	1
12	1	8	0	45	2	8850000	20	3	1
13	1	8	0	45	3	7857275	7	3	1
14	1	8	0	45	2	8850000	21	3	1
15	1	8	0	45	3	8213888	16	3	1
16	1	8	0	45	2	8850000	29	3	1
17	1	8	0	45	3	8231835	3	3	1
18	1	8	0	45	2	8850000	54	3	1
19	1	8	0	45	3	8625176	16	3	1
20	1	8	0	45	2	8850000	73	3	1
21	1	8	0	45	3	8299521	8	3	1
22	1	8	0	45	3	8850000	42	3	1
23	1	8	0	45	3	8799634	2	3	1
24	1	8	0	45	2	8850000	26	3	1
25	1	8	0	45	3	10564565	9	3	1

Figure 3. *Dataset Exploration*

The following picture is a Dataset taken from a tenant payment application. There are 9 attributes, with a total of 3103 data, this dataset is tenant payment data rangingfrom 2012 to 2020.

- Davies Bouldin Evaluation

Figure 4. *Result of Davies Bouldin*

Several experiments to determine clusters have been carried out using the clustering evaluation method. By looking at the results of Davies Bouldin's calculations, by applying several experiments with different numbers of clusters, it was found that the best number for Davies Bouldin is in the 3-cluster experiment because it has a value closest to 1.

- Data Selection and Clustering

Row No.	id	Kode Tenant	cluster	Ketepatan B...	Jumlah Hari ...	Cluster
1	1	1	cluster_1	-0.136	173	0
2	2	1	cluster_0	-0.136	35	1
3	3	1	cluster_0	-0.136	17	1
4	4	1	cluster_0	-0.136	21	1
5	5	1	cluster_0	-0.136	1	1
6	6	1	cluster_0	-0.136	17	1
7	7	1	cluster_0	-0.136	25	1
8	8	1	cluster_0	-0.136	35	1
9	9	1	cluster_0	-0.136	16	1
10	10	1	cluster_0	-0.136	29	1
11	11	1	cluster_0	-0.136	10	1
12	12	1	cluster_0	-0.136	20	1
13	13	1	cluster_0	-0.136	7	1
14	14	1	cluster_0	-0.136	21	1
15	15	1	cluster_0	-0.136	16	1
16	16	1	cluster_0	-0.136	28	1
17	17	1	cluster_0	-0.136	3	1
18	18	1	cluster_0	-0.136	54	1

ExampleSet (3,103 examples, 3 special attributes, 3 regular attributes)

Figure 5. *Data Clustering*

The picture above is a dataset whose attributes have been selected and divided by the clustering method. Of the 9 attributes that were generated, after going throughthe results of the discussion, our group decided to use the 6 attributes used, namely tenant code (label), payment type code, payment accuracy, business type code, number of days late, and cluster, because the focus of information is What is desiredis the division of cluster tenants who pay late and tenants who pay on time.

- Overview Cluster Model Visualizer

Figure 6. *Overview Cluster Model Visualizer*

The following is an overview of the results of the visualizer model cluster operator, it can be seen that the dataset is divided into 3 clusters with an explanation of the percentage on the average number of days late attribute and payment type code.

- Centroid Table dan Chart

Cluster	Ketepatan Bayar	Jumlah Hari Terlambat	Cluster
Cluster 0	0.022	28.478	1.018
Cluster 1	-0.136	121.600	0.397
Cluster 2	-0.136	293.098	0

Figure 7. *Centroid Table*

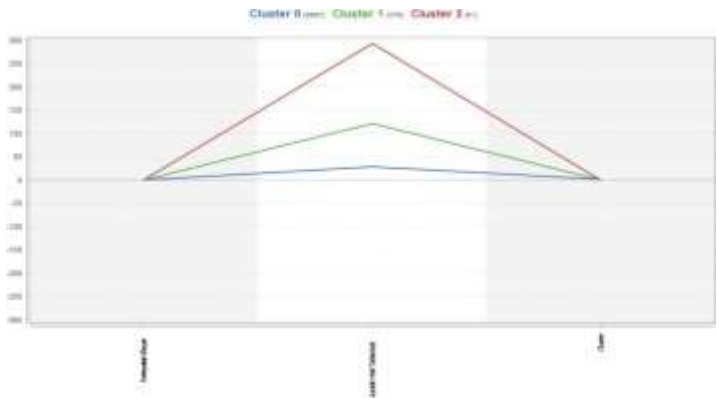

Figure 8. *Centroid Chart*

The following is an overview of graphs and tables of centroids for modeling results that describe the distribution of data and the average of each cluster with the attributesused in data processing.

- Cluster Model and Performance Vector

Table 4. Cluster Model

Cluster Model	
Cluster 0	2667
Cluster 1	375
Cluster 2	61
Total Number of Items	3103

Table 5. Performance Vector

Performance Vector	
Avg. within centroid distance	-657.874
Avg. within centroid distance_cluster 0	-351.814
Avg. within centroid distance_cluster 1	-1380.085
Avg. within centroid distance_cluster 2	-9599.367
Davies Bouldin	-0.546

The table above is a cluster model and performance vector which is the output of the k- means and performance operator. For the cluster model table contains an explanation regarding the number of values contained in the cluster consisting of Cluster 0: 2667, Cluster 1: 375, Cluster 2: 61 so in total, this amount is in accordance with the total data contained in the dataset.

While in the performance vector table, there is an explanation of the average distance from the centroids in each cluster, with Davies Bouldin value of -0.546, the higher the centroid distance number, the farther the data distance between the cluster centroids is.

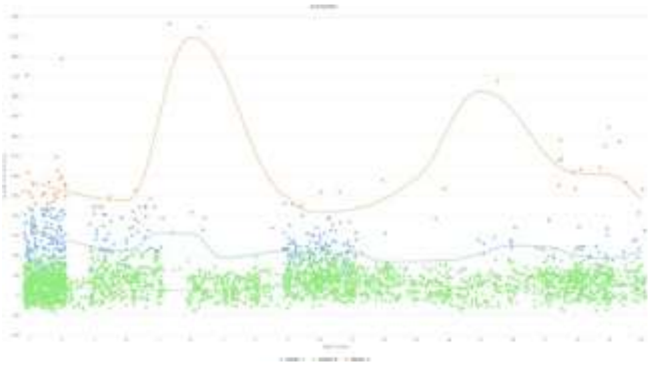

Figure 9. *Cluster Visualization*

For each cluster can be categorized as follows:

- Cluster 0 = Those who pay under 76 days (On time)
- Cluster 1 = Those who pay from 76 – 200 days (Late)
- Cluster 2 = Those who pay more than 210 days (Very late)

4.6 Deployment

From the analysis, a mitigation plan can be generated for each cluster, so that there will beno more late payment in the future:

Table 6. Mitigation Plan

Cluster	Mitigation Plan
0	-
1	Moderate fine
2	Heavy fine

Based on the clusters above, the mitigation measure to prevent similar cases of late paymentsby tenants in the future is by applying a penalty system which will get worst the longer the tenant stalls the payment from the due date. Cluster 0 is tenants that pay on time so they willnot be penalized, cluster 1 in tenants who are late paying with up to 200 days will be subjectto medium fines, and cluster 2, namely tenants who are late paying with a time range of morethan 210 days will be subject to heavy fines.

V. Conclusion

This project used RapidMiner as a tool to process the dataset. The dataset is taken from the XYZ mall tenant payment application to find out and classify the payment times made by XYZ mall tenants. In the modeling stage, the method used in this research is K-Means clustering as an algorithm in making data clustering. The results obtained are 3 data clusters, each of which has a different time range. The results obtained are expected to help companies in managing tenants by using data from data mining as a basis for making decisions on tenants who experience late payments and from this research the company's management can make regulations regarding payments for a service used by tenants. Due to the short time and the availability of inadequate data, of course this research is not a perfect study and hopefully that there are wide opportunities for further research in the future by other researchers.

References

Achenbach, A., & Spinler, S. (2018). Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights. Operations Research Perspectives, 5(July), 265–279. https://doi.org/10.1016/j.orp.2018.08.004

Aprilla Dennis. (2013). Learn Data Mining with Rapidminer. Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2, 5(4), 1–5. http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf

Brandt, T., Wagner, S., & Neumann, D. (2021). Prescriptive analytics in public-sector decision-making: A framework and insights from charging infrastructure planning. European Journal of Operational Research, 291(1), 379–393. https://doi.org/10.1016/j.ejor.2020.09.034

Chen, L. H., & Kang, F. Sen. (2010). Integrated inventory models considering permissible delay in payment and variant pricing strategy. Applied Mathematical Modelling, 34(1), 36–46. https://doi.org/10.1016/j.apm.2009.03.023

Choi, Y., Yoon, G., & Kim, J. (2022). Unsupervised learning algorithm for signal validation in emergency situations at nuclear power plants. Nuclear Engineering and Technology, 54(4), 1230–1244. https://doi.org/10.1016/j.net.2021.10.006

Choudhari, A. S., & Potey, M. (2018). Predictive to Prescriptive Analysis for Customer Churn in Telecom Industry Using Hybrid Data Mining Techniques. Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018, 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697532

Damayanti, S. E. (2006). Analisis Dan Implementasi Framework Crisp-Dm (Cross Industry Standard Process for Data Mining) Untuk Clustering Perguruan Tinggi Swasta. Jurnal STT Bandung.

Desineedi, R. M., Mahesh, S., & Ramadurai, G. (2020). Developing driving cycles using k- means clustering and determining their optimal duration. Transportation Research Procedia, 48(2018), 2083–2095. https://doi.org/10.1016/j.trpro.2020.08.268

Fan, C., & Xiao, F. (2017). Assessment of Building Operational Performance Using Data Mining Techniques: A Case Study. Energy Procedia, 111(September 2016), 1070–1078. https://doi.org/10.1016/j.egypro.2017.03.270

Gladju, J., Kamalam, B. S., & Kanagaraj, A. (2022). Applications of data mining and machine learning framework in aquaculture and fisheries: A review. Smart Agricultural Technology, 2(April), 100061. https://doi.org/10.1016/j.atech.2022.100061

Hartanto, D. (2020). Sociology Review of Social Phenomenon, Social Rules and Social Technology. Budapest International Research and Critics Institute-Journal (BIRCI-Journal) Vol 3, (2): 1175-1184.

Khumaidi, A. (2020). Data Mining for Predicting the Amount of Coffee Production Using CRISP-DM Method. Jurnal Techno Nusa Mandiri, 17(1), 1–8. https://doi.org/10.33480/techno.v17i1.1240

Laperrière-Robillard, T., Morin, M., & Abi-Zeid, I. (2022). Supervised learning for maritime search operations: An artificial intelligence approach to search efficiency evaluation. Expert Systems with Applications, 206(April), 117857. https://doi.org/10.1016/j.eswa.2022.117857

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. Wiley.

Mauritsius, T., Braza, A. S., & Fransisca. (2019). Bank marketing data mining using CRISP- DM approach. International Journal of Advanced Trends in Computer Science and Engineering, 8(5), 2322–2329. https://doi.org/10.30534/ijatcse/2019/71852019

Miranda, E. (2008). Pengembangan Business Intelligence Bagi Perkembangan Bisnis Perusahaan. CommIT (Communication and Information Technology) Journal, 2(2), 111. https://doi.org/10.21512/commit.v2i2.501

Mosavi, N. S., & Santos, M. F. (2020). How prescriptive analytics influences decision making in precision medicine. Procedia Computer Science, 177, 528–533. https://doi.org/10.1016/j.procs.2020.10.073

Naik, A., & Samant, L. (2016). Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science, 85(Cms), 662–668. https://doi.org/10.1016/j.procs.2016.05.251

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181(2019), 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Šikšnys, L., Pedersen, T. B., Liu, L., & Özsu, M. (2016). Prescriptive analytics.

Encyclopedia of Database Systems, 1–2.

Sikumbang, E. D. (2018). Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. Jurnal Teknik Komputer AMIK BSI (JTK), Vol 4, No.(September), 1–4.

Solano, J. A., Lancheros Cuesta, D. J., Umaña Ibáñez, S. F., & Coronado-Hernández, J. R. (2021). Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test. Procedia Computer Science, 198(2020), 512–517. https://doi.org/10.1016/j.procs.2021.12.278

Tinoco, J., Parente, M., Gomes Correia, A., Cortez, P., & Toll, D. (2021). Predictive and prescriptive analytics in transportation geotechnics: Three case studies. Transportation Engineering, 5(May), 100074. https://doi.org/10.1016/j.treng.2021.100074

Václav, C., Gabriel, F., Blanka, K., Libor, K., & Michal, T. (2021). Utilization of business intelligence tools in cargo control. Transportation Research Procedia, 53(2019), 212–223. https://doi.org/10.1016/j.trpro.2021.02.028

Wieder, B., & Ossimitz, M. L. (2015). The Impact of Business Intelligence on the Quality of Decision Making - A Mediation Model. Procedia Computer Science, 64, 1163–1171. https://doi.org/10.1016/j.procs.2015.08.599

Zhang, L., Qi, Z., & Meng, F. (2021). A Review on the Construction of Business Intelligence System Based on Unstructured Image Data. Procedia Computer Science, 199, 392– 398. https://doi.org/10.1016/j.procs.2022.01.048