# Classification Of Loyality Customer Using K-Means Clustering, Studi Case : PT. Sucofindo (Persero) Denpasar Branch

Charolina Devi Oktaviana Soleman[1*], Nyoman Pramaita[2] and Made Sudarma[3]

[1,2,3] Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
*devicharolina@gmail.com

**Abstract** The success of the company in developing its services can be seen from the number of customers who use these services, customer loyalty in using services can be seen from customer loyalty. Customer loyalty is an important factor in the development of a business in the company, repeated use of services can be used as an indicator in determining the level of customer loyalty, by paying attention to customer loyalty, of course the company will be able to develop customer focus on an ongoing basis. Clarification of customers needs to be done to find out how demographics customers use services, can be seen from how many customers and the level of transactions made. K-Means clustering is one method that can be used for the classification process of customer data through transactions carried out by forming several clusters, this classification process is divided into 5 clusters with the results of which include 1) A few small number transactions with many customers, 2) Many transactions small number with many customers, 3) a few transactions with a medium number of customers begin to decrease, 4) a few large number transactions a few customers, 4) a moderate number of transactions with a number of customers.

*Index of Terms – Loyality, Customer, Classification, K-Means Clustering*

## I. INTRODUCTION

Customer loyalty is one important factor in supporting the company's business development where the customer focus is the company's key in maintaining customer satisfaction with the services provided, Customer loyalty is a form of a result of a trial process in the use of an item or service so as to cause satisfaction for customers to repeat order [1] Customer loyality can also be interpreted as customer loyalty in using products or services repeatedly with different transaction levels, by taking into account customer loyalty issues, the company can determine strategies to improve service to customers.

Transactions carried out by customers can be large or small amounts, both in one transaction or repeatedly, customer transaction data can be used as a data source that is processed to produce information related to customer classification based on the level of transactions made, both from the nominal transaction or from the number of transactions made and helps the company improve its strategy in customer focus.

This customer loyalty classification can be done using the k-means clustering method by utilizing the k-means algorithm the customer classification process will be easier by determining the number of cluster points desired and calculating the closest distance from each data it will produce a grouping of data based on the characteristics of the data so that the data generated will be in accordance with the demographics of customer transactions in the use of services

## II. PURPOSE OF PAPER

The purpose of customer loyalty classification is to obtain information from transaction data related to the level of customer loyalty to the use of services either through nominal transactions or from the level of transactions made at PT. SUCOFINDO (Persero) Denpasar Branch.

## III. LITERATURE REVIEW

### A. State of The Art

According to research from Asri Prameshwari, Rukmi Sri Hartati, Made Sudarma (2018) with research entitled " Analysis of Student Competency Improvement Using the ASSURE-based Learning Model Project-Based Learning"

The use of the genethic k-means algorithm method in research increases student competency by forming optimal clusters using natural selections and produces the 3 best clusters, namely high, medium, and low with the results of research that is the competency of students from the moderate category increases by 31.37% and 14%, for the low category increased by 31.37% and 74% for the high category decreased by 35.29% and 60%. the results of this study can be used by STIKes Wira Medika in improving the teaching and learning process of information technology courses [2].

Other studies by Gustientiedina, M. Hasmi Adiya, Yenny Denelita (2019) entitled "Application of K-Means Algorithm for Drug Data Clustering in RSUD Pekanbaru". In this study, it explains that related to the use of the k-means method in classifying the group of medicine in the general public hospital in Pekanbaru that can be used in decision making in planning and controlling the supply medicine at the hospital by determining 3 clusters of drug use, namely few, moderate and high with a total of 295 data, the result is the level of use of medicine while the average level of medicine use is 18000 units, and the average level of usage is 70000 units while the medicine included in the use of high average medicine demand each year above 70000 pieces [3].

Research by Dwi Ardiada, Pasek Agus Ariawan and Made Sudarma (2010) entitled " Evaluation of Supporting Work Quality Using K-Means Algorithm". In this study, explained related to the application of the k-means algorithm in evaluating the performance of employees at Udayana University with the aim of classifying the quality of contract workers into several clusters to evaluate performance productivity, the data used were 1613 data and cleaning data used to be 544 the data is then iterated to get the optimal number of clusters, so that the data clusters do not change, the results are formed 5 clusters in the iteration process where the performance with a very good assessment there are 201 members, the performance is not good there are 37 members, and the performance with good categories there are 237 members and with poor performance there are 64, there are 5 people with diverse performance assessments, with this grouping data facilitating management in evaluating employee performance for future contracts [4].

Research entitled "Data Mining for Clustering Revenue Plan Expense Area (APBD) by using K-Means Algorithm" by Wahyudin, I Putu Ari Wijaya and Ida Bagus Swamardika, Implementation of data mining application in Bima city government budget data processing in APBD planning, the process is carried out using k-means algorithm and multiple linear regression to form clusters related to the similarity characteristics of budget data to facilitate classification in predicting the value of the APBD that will be requested in the future [5].

Harsemadi and Made Sudarma apply the k-means algorithm for classification of music base on mood [6]. This study aims to obtain information related to the classification of types of music according base on mood using 40-400 data sets where clusters form into 4 that represent the similarity characteristics of the types of music classified, using the unsupervide k-means clustering method. The result shows that the average value of silhoutte coefficient is 0.62688807, that is, each cluster has a good level of compactness.

Sri Utami et all, implementation algorithm k-means and time series method for stock management as stok order [7]. The k-means method is applied to assist in grouping drugs and the right amount of drugs to be purchased so that the management of drug stocks can be carried out correctly and the flow of goods does not arrive at an empty position cluster formation by forming 3 sluters that have been tested using 1000 test data and produce that in cluser 1 there are 13 members with high transaction rates, in cluster 2 there are 29 members with medium transaction levels and in cluster 3 there are 958 clusters with transaction levels low in accordance with the performance of the last 3 months at the pharmacy, from the results obtained, the stock management needs to be considered for purchasing drugs at a higher transaction level.

Rony Setiawan "Application Of Data Mining Using K-Means Clustering Algorithm To Determine The Promotion Strategy New Students" [8] the aim of this study is to assist LP3I polytechnics in determining strategies in the act of promoting the acceptance of new students, the variables used to measure are age, religion, exam status, graduation, registration, sex, registration wave, registration wave, test number, majors, study program, source of information, city of origin, campus location, parents' occupation, type of school and payment of registration fees by forming 4 optimal clusters obtained results namely domination of prospective students from middle and low economies so that the strategies carried out in the sales promotion section It is expected to be able to provide discounts and provide cheaper appeal when registration is opened, a special price is applied for concentrations that are less desirable to attract students in choosing majors.

Sri Rayahu et al implemtation of k-means clustering to determining criminal potential at Banjarbaru City [9]. The aims from ths research is to find out the potential for a crime to be carried out by analyzing crime data owned by the police in Banjarbaru city by forming 3 clusters and paying attention to the attributes of punishment, month and criminal report which are expected to help the police in determining strategies for minimize criminal acts in Banjarbaru.

Anindya Khrisna implementation k-means algorithm for clustering Of Diseases patients In Pekalongan Kajen Public Health [10]. The implementation of the k-means algorithm in this research is by grouping data based on the results of medical records of the health of patients in Pekalongan Hospital, this is done to determine the average level of acute and non-acute illnesses suffered by patients, by dividing into 2 clusters and measure the level of similarity between data through predefined attributes namely age, sex, duration of illness and also diagnosis to analyze the population in health archiving from 1000 data sets that have been determined.

Wrs Nurwidodo and Mochamad Hariadi "Multidimential Decision Support System Using Expensive Based K-Means Clustering Distance"[11]. The implementation of the k-

means algorithm conducted in this study is to design a decision support system in determining potential participants who can be prioritized by using mahalanobis distance, which is measuring the suitability of each cluster member to get the best clustering results, by applying the k-means method. expected to be able to provide an alternative picture in determining the prospective training participants who have met the requirements to be able to attend the training.

Putraman Alkhairi, Agus Perdana "Application of K-Means Cluster in the Potential Areas of Productive Rubber Agriculture in North Sumatra" [12]. the target of this research is to optimize the development and use of rubber land, data obtained from the central statistical agency this is done by measuring the potential of productive rubber producing areas, the measurement is seen from the similarity between data from each cluster / group produced where later this research is expected to be able to contribute in evaluating and increasing the effectiveness of productive rubber land use in each region, from the process achieved by determining 3 clusters, the results of the lowest cluster is a focus of improvement that needs to be done to optimize rubber land.

Ade Bastian et al apply the k-means clustering algorithm for analysis the human infected disease case study in majalengka[13], Testing of the k-means clustering algorithm in this study was carried out to conduct cluster analysis of infectious diseases in humans based on the attributes of each puskesmas owned by each sub-district in the majalengka district, the results obtained were to verify the supply of drugs for the incoming health centers into the cluster has the most patients in handling the spread of cases of human infectious diseases, from the data set that has been determined that is a number of 32 data with the number of clusters formed as many as 6.

Gede aditra pradnyana et al research of "System Of Class Distribution With Students K-Means And K-Nearest Neighbors Methods For Improve Learning Quality" [14]. This study aims to facilitate the university in determining the formation and division of classes more effectively by processing data on each of the attribute requirements in taking new classes, namely grades, midterms, final semester scores and cumulative achievement index (IPK) for each course taken. The results show that the number of clusters formed and the data used is very influential in producing a good cluster or grouping where in this study the data sets used were 100 sets with the number of clusters 10 of 0.534 (*medium structure*).

Sri Tria Siska apply the k-means algorithm method for Determining Selling Water Cubication Based On Customer Grouping [15]. The research is carried out to determine the amount of cubication of water sold from water sales data on each customer, this is done aiming to find out how water is used by customers for follow-up in determining PDAM strategies in water management using 20 customer data sets and rapidminers The software uses 20 sample data sets by forming 3 clusters that define each label, namely Cluster 1 consists of 7 members with moderate water usage, Cluster 2 with 2 customers with high water use levels and Cluster 3

with 11 members with water saving levels .

Lili Rusdiana et al entild the reseach "Modeling Of K-Means On Determining Predicated Graduation Of Palmkaraya Stmik Students" [16]. apply the data mining of k-means clustering for Modeling using k-means grouping is used to determine the predicate of graduation for students at Palangkaraya University by determining some predetermined input attributes based on the amount of study load, GPA and thesis graduation, using 10 sample data sets, from the results obtained carried out by inputs and outputs as calculations using the MAPE formula, obtained 70% value, obtained real results and 30% unapproved real results, collection from the centroid point conducted at the beginning of the study.

Feby Fajrianti et al apply the kmeans clustering method for Analysis And Analysis Discrimination In The Poor Village Grouping In Pangkep District [17]. This study aims to classify poor villages based on several variables that have been determined such as education, occupation density, crude birth rate, gross martyrdom, and the number of family members, data obtained through BPS and the Pangkep district civil registry office with the object of research are all villages that have to the district. the results obtained are grouping villages based on the level of poverty obtained by clustering optimal as many as 3 clusters at 98.06% the results of the calculation of the accuracy of the disciminant analysis.

Very Kurnia and Jatmiko implemenyayion Clustering of Final Project Documents Using K-Means Clustering as Analysis of Implementation of the Gathering System [18]. This aims study to implementation of k-means in this research is to classify the theme of each final project to facilitate the search for documents, before the method used is ranking documents, the results of clustering show that the grouping is quite good and can be applied to the application of the retrieval system with davies values. bouldin index 0.001 thus the k-means method is very good for classifying the final project by theme.

Abdi Hazman et al using Application Of K-Means Clustering Method For Grouping Port And Unload Loading Goods Data In Indonesia [19], This research aims to classify the ports in an area to find out which ports are included in the port that are traded / function properly and not where the attributes used in this study are the name of the port and the number of goods being unloaded at the port, obtained the result is that grouping is done by determining as many as 5 clusters that are formed where in cluster 1 it is obtained which provinces are the centers of the port of the activity of unloading high quantities of goods and in cluster 5 are the ports with a small loading and unloading rate of goods, the factor in port development is not only done by looking at how much the process of loading and unloading of goods is carried out but also based on the geographical layout of each port.

Hardiani P.R et al apply the Analysis of K-Means Claster In Spatial Regression Modeling In Tuberculosis Case In East Java In 2017 [20]. in this study explained that the use of the k-means method for modeling the linear regression modeling process in tuberculosis cases where the merging of these two

methods to get the best results in classifying the spread of tuberculosis in the East Java region, an analysis was carried out on the factors that influence the number of patients with tuberculosis, the results obtained are factors that significantly influence the number of patients with colocosis in East Java, namely the variable population density and healthy homes with R2 and AIC values respectively 87.10% and 586.69.

From some researches that have been done, the application of data mining in utilizing data warehouse obtained from business processes, can help companies in supporting decision making, from several studies that have been carried out in the implementation of the use of k-means also support this research to carry out a classification process with dividing it into several clusters obtained by looking at the characteristics of the data and the number of iterations that occur to produce an optimal classification.

### B. Data Mining

The need to expand the knowledge of the data generated makes the ability of human analysis to be increasingly developed [21], Humans have limitations in managing large amounts of data that require computers as data processing machines [22], with very large amounts of data computers cannot process without an appropriate algorithm in data processing, how humans are able to find an information from a large amount of data collection and make it easy to understand as one of the important information needed. the process of extracting data can be done in an appropriate way to assist humans in finding analysis of data development.

Data mining is a process carried out by extracting basic data into information that has important value, with data patterns that are formed from data, basic data known as Knowledge Discovery in a Database that is searching by using repetitive based learning to find a fixed pattern and easy to use by users in decision making.[23]



Fig 1. Data Mining

Implementation of data mining in this company is able to assist companies in managing data owned by companies that initially have no value into important information to support the decision making in the future by paying attention to data patterns and information that can be obtained through data extraction [24]. Technology in data processing can be used, among others as.

1. Predict a business process where data mining is used to process a large number of databases and obtain information that has added value in it
2. Finding a pattern of relationships between data that was previously unknown in a database.

Data mining and KDD in the process of extracting information in a hidden manner are often used, in fact in both terms have different concepts but are interrelated with each other, The stages of data mining follows .[25]
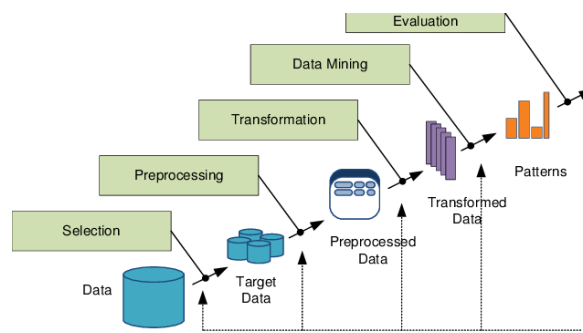


Fig 2. Stages Of Data Mining

1. Data Selection is the process of selection of data contained in an available database that will be carried out extracting information and then stored separately from the database.
2. Pre-Processing / Cleaning is the process that is carried out before the mining process of the data to be processed needs to be done cleansing this data to avoid inconsistent data and data duplication.
3. Transformation is the process of transformation in data mining is the process of processing data that has previously been cleared to obtain information either by coding or by using a system to process the data needed.
4. Data Mining
   Determination of the task of extracting data is an ultimate goal of doing data processing, the results obtained in data mining processing related to the characteristics of data, clustering and cleaning of data that is not inconsistent
5. Interpretation / Evaluation
   the process by which the results of the patterns formed against the data that have been determined can be understood and understood to be used and it is necessary to carry out an evaluation and evaluation of the initial hypothesis that has been determined.

### 6. K-Means Algorithm

K-Means is an algorithm in which K value is a constant value of the number of clusters desired, while means means is the average of a group of numbers in this case defined as a cluster so that K-Means is an unsupervised analysis method and is one of the methods analysis of data with a system partition that is grouping existing data into several groups that have some similar characteristics and some groups that do not have similar characteristics to each other, the k-means algorithm is an algorithm that requires the input parameters

k and divides into groups of *n* objects into *k* clusters , the similarity between the members of the cluster is measured by looking at the closeness of the object between the mean on the cluster or also called the centroid cluster [26].

The following formula for measuring the distance of each centroid *Clustering*

$$d(P,Q) = \sqrt{\sum_{j=1}^{p} \left( x_j(P) - x_j(Q) \right)^2}$$

Information :
D = Document point
P = Data Record
Q = Centroid data

The shortest distance between each centroid and the data determines the cluster position of the data, while the formula of iteration to measure it is as follows.

$$C(i) = \frac{x1 + x2 + x3 + \cdots.. + xn}{\sum x}$$

Information :
$X_1$ = Value record data ke-1
$X_2$ = Value record data ke-2
$\sum X$ = Sum record data

The steps in implementing the k-means algorithm are as follows :
1. Determine K as the number of clusters to be formed
2. Determine the initial K center point cluster (centorid) to be done randomly, the initial centroid determination is done randomly from the available objects as many as K cluster to calculate the centroid cluster k-i
3. Calculate the distance from each object to each centroid from each cluster using euclidean distance
4. Allocate each object into the closest centroid, to allocate the object into each cluster during iteration in general to assert the object as a member of the cluster by measuring the proximity of the properties to the center point
the cluster.
5. Perform iteration and then determine the new centroid position using the equation
6. Repeat step three if the new centroid positions are not the same.

### 7. Clustering
Clustering is one technique in data mining that is grouping a number of data or objects into clusters so that each cluster formed will only have data that have similar characteristics to each other and differ from other objects [26], until now scientists are still trying to find a model which is optimal in determining the number of clusters in the clustering process, there are two known clustering models, including:

1. Clustering hierarchy consisting of complete linkage clustering, single linkage clustering, average linkage clustering and centroid linkage clustering
2. Partitions consist of k-means and fuzzy k-means

### 8. ELBOW Method
Elbow method is a method that can be used to find and determine the best number of clusters by looking at the percentage of comparison results between the number of clusters formed, to get it by calculating the value of the SSE (Sum of Square Error) of each cluster, the greater the number of clusters then the SSE value will be smaller [27], the formula of SSE as follows :

$$SSE = \sum_{K=1}^{K} \sum_{x_i \in S_K} \left\| x_i - c_k \right\|_2^2$$

Information :
K = Total cluster
Xi = Data ke – i
Ck = Centroid Cluster

Stages in the Elbow method algorithm
1. Initialize initial k values;
2. Increase the value of k;
3. Calculating the sum of square error results of each k value;
4. Analysis of the results of the sum of square errors of the value of k experienced
drastic decline;
5. Find and set the value of k in the form of an elbow.

### 9. ELKI System
Elki is data mining software focused in research in algorithm, with an emphasis on unsupervised methods in cluster analysys and outlier detection [28], Knowledge Discovery in Database used for research and learning of database systems . The purpose of this application is to develop sophisticated data mining systems that are sophisticated and have interaction with index structures database, the ELKI framework is built with the Java language and the algorithms that are included in it include grouping through a combination of algorithms including data types, distance functions and data sizes.



Image 2. Example Of Elky System

## IV. METHODOLOGY AND DESIGN

### A. Research Location

The study was conducted at PT. SUCOFINDO (PERSERO) Denpasar Branch in Customer Data Transaction.

### B. Data Transaction

Transaction data is data obtained through recording the receipt of orders from customers by an administrative function that is entered in the Order management system with an Enterprise Resources Planning-based system, where this step is carried out when the customer makes a service request.

The following is the customer data used for the clustering process by taking into account the variable id_customer and the number of transactions carried out. the data used here is transaction data for 2019 PT. Sucofindo (Persero) Denpasar Branch.

Tabel 1. Example data Of Data Transaction

| No | Customer Name | Id Customer | Value Transaction |
|----|---------------|-------------|-------------------|
| 1 | Amman Mineral Nusa Tenggara | 36204 | 213.400.000 |
| 2 | Pama Persada Nusantara | 36602 | 5.500.000 |
| 3 | Tarukalpa Dewata | 128091 | 385.000 |
| 4 | Fitryah Eksport Indah | 56144 | 1.210.000 |
| 5 | Bale Gede Internasional | 58792 | 527.382 |
| 6 | Diva Dwi Putra | 151591 | 385.000 |
| 7 | Mahaka | 125695 | 385.000 |
| 8 | Teak Bali | 78593 | 385.000 |
| 9 | Dua Sahabat | 45226 | 446.839 |
| 10 | Indosena | 56589 | 165.000 |

The data obtained are 2303 data but after cleaning the data 326 data will be generated which will be carried out the classification process to determine customer loyalty in conducting service transactions, then the results are exported into CSV by query.
SELECT name, COUNT (id_transaction), SUM (value_invoice) FROM `fact_transaction` INNER JOIN dim_customer on dim_customer.id_customer = fact_transaction.id_customer GROUP BY fact_transaction.id_customer.

### C. Elky Processing & Transforming

In the process of data processing using the Elki application there are several steps that can be done, among others
1. Data in the form of CSV Then compiled to suit the needs required by ELKI.
2. In the application select DB.in to enter the data just now.

| dbc | Default: FileBasedDatabaseConnection |
|-----|--------------------------------------|
| dbc.in | C:\Users\ANGGA\Desktop\fact_transaction-1.csv |

3. After that select the algorithm we will use (here K-Means) and the desired number of Clusters (Here 5)

| algorithm | clustering.kmeans.KMeansSort |
|-----------|------------------------------|
| kmeans.k | 5 |

4. Press Run Task

| Save | Remove | Run Task |
|------|--------|----------|
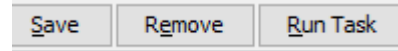
In this research, the k-means clustering process is not done manually but uses Elki Systen where we only need data in the CSV format to do the clustering process and determine the optimal number of clusters to be formed. After selecting 5 clusters on Elki, Elki will automatically set the color along with what will be used. The following are k-means clustering indicators used in ELKI.

**k-Means Clustering**
+ Cluster
× Cluster
○ Cluster
□ Cluster
◇ Cluster

Image 3. Indicator ELKI

Information :
Red    :  Cluster  1
Yellow  : Cluster  2
Blue    :  Clsuter  3
Green   : Cluster  4
Purple   : Clsuter 5

## V. TESTING AND RESULTS

### A. K-Means Clustering Results

The results of grouping transaction data based on id_customer and the number of transactions carried out will get clsutering results as follows :

Tabel 2. Result Clustering

| No | Label | Vector | Cluster Id | Centroid |
|----|-------|--------|------------|----------|
| 1 | Amman Mineral Nusa Tenggara | 77,15869526 | 2 | 77,159 |
| 2 | Pama Persada Nusantara | 1,5 | 0 | 6.743 |

| 3 | Tarukalpa Dewata | 104,5300438 | 0 | 6.743 |
|---|---|---|---|---|
| 4 | Fitryah Eksport Indah | 7,24 | 0 | 6.743 |
| 5 | Bale Gede Internasional | 21,20407864 | 0 | 6.743 |
| 6 | Diva Dwi Putra | 60,3135 | 0 | 6.743 |
| 7 | Mahaka | 168,8367348 | 0 | 6.743 |
| 8 | Teak Bali | 1,35 | 0 | 6.743 |
| 9 | Dua Sahabat | 22,21506267 | 0 | 6.743 |
| 10 | Indosena | 14,34 | 0 | 6.743 |

The following is a graph of clustering results in column 1 or number of transactions. Seen there are the most customers doing transactions from 1-40 transactions a year
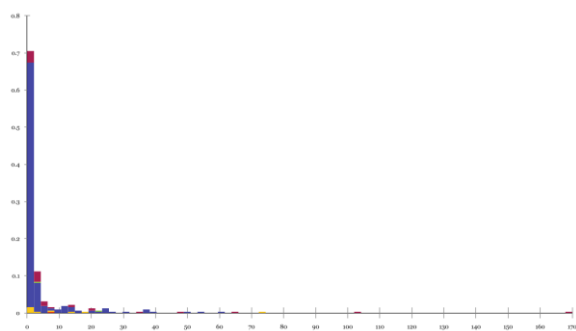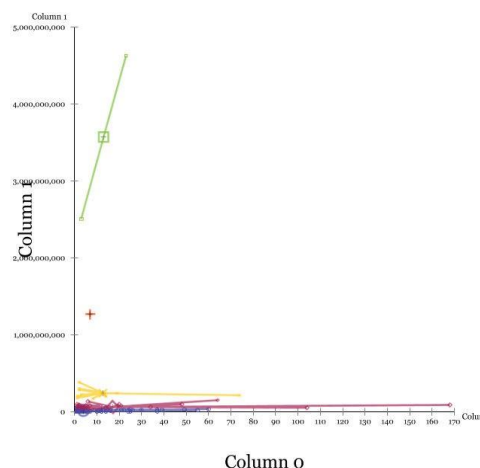


Grafik 1. K-means Mapping

The following is a graph of the results of clustering in column 2 or many transactions. Seen in the chart for many transactions conducted by customers in the range of 0-1 billion in 2019.
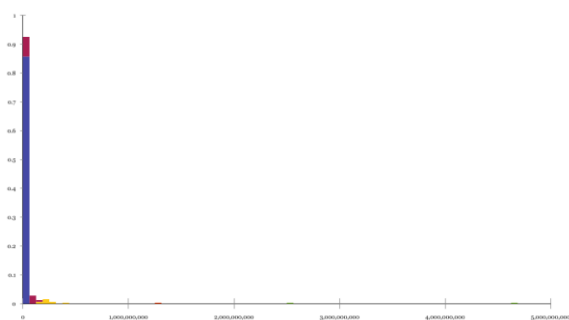


Grafik 2 Transaction Mapping by customer



Grafik 3 Combination Cluster

After being combined into a 2-dimensional graph and 5 clustering as shown next

Blue : Do a few small transactions but many customers there
Red : Many Transactions Small amounts, many customers
Yellow : A few transactions with a medium number of customers begin to decrease
Green : A few large transactions, a few customers
Purple : Medium number of very large transactions, few customers

Based on the results of clustering that has been done by determining the 5 clusters beforehand, it can be obtained information that many customers use services repeatedly, but with a small number of transactions, while for the number of transactions that have a high value tends not to repeat service orders or only make one transaction

## VI. CONCLUSION

In this research, data mining classification is conducted to determine customer loyalty using K-Means Clustering at PT. Sucofindo (Persero) Denpasar branch. The classification process is carried out by dividing into several clusters which are obtained from the data characteristic parameters and the number of iterations that occur to produce optimal classification. The software used for classification is Elki by using company transaction data in the CSV format. The classification results in this study indicate that many corporate customers make repeat transactions but with a small number of transactions. Elki is very suitable to be used in classifying data mining.

REFERENCES

[1] Sinaga, Partua Pramana Hamonangan. 2010. Analisis Pengaruh Kualitas Pelayanan,Kepuasan Pelanggan, Dan Lokasi Terhadap

Loyalitas Pelanggan ( Studi Kasus Pada Warnet Chamber Semarang ), Semarang : Manajemen UNDIP.

[2] Asri, Prameswari, Rukmi Sari Hartati, Made Sudarma (2018). " Analysis of Student Competency Improvement Using the ASSURE-based Learning Model Project-Based Learning " Majalah Ilmiah Teknologi Elektro Vol 17 No 3.

[3] Gustientiedina, M. Hasmil Aditya, and Yenny Desnelita (2019). " Application of K-Means Algorithm for Drug Data Clustering in RSUD Pekanbaru ", Jurnal Nasional Teknologi Dan Sistem Informasi, Vol. 05, No. 01.

[4] Dwi Ardiada, Pasek Agus Ariawan and Made Sudarma. (2018). " Evaluation of Supporting Work Quality Using K-Means Algorithm ", International Journal Of Engineering and Emerging Technology Vol 3, No. 1.

[5] Wahyudin, I Putu Ari Wijaya and Ida Bagus Swamardika, (2017) "Data Mining for Clustering Revenue Plan Expense Area (APBD) by using K-Means Algorithm ", International Journal Of Engineering and Emerging Technology Vol 2, No. 1.

[6] Harsemadi and Made Sudarma (2017). "Classification Of Music Base on Mood using K-Means Method" Konferensi Nasional Sistem dan Informatika, STIMIK STIKOM Bali.

[7] Sri Utami, I Gede Wira Darma, Ni Wayan Sri Aryani (2019). "Stock Management using K-Means and Time Series Method As Stok " International Journal Of Engineering and Emerging Technology Vol 4, No. 1.

[8] Rony Setiawan (2019)."Application Of Data Mining Using K-Means Clustering Algorithm To Determine The Promotion Strategy New Students" Jurnal Lentera ICT Vol. 3 No. 1 / ISSN 2338-3143.

[9] Sri Rahayu, Dodon T. Nugrahadi, and Fatma Indriani (2014). "Clustering Determining Of Potential Crime in Banjarbaru city used K-Means Method" Kumpulan Jurnal Ilmu Kompuyter Volume 01, No. 01.

[10] Anindya Khrisna Wardhani (2016). "K-Means Algorithm Implementation For Clustering Of Patients Disease In Kajen Clinic Of Pekalongan" Jurnal Transformatika, Volume 14, Nomor 1, Juli.

[11] G. Grigoras and F. Scarlatache, (2015). "An assessment of the renewable energy potential using a clustering based data mining method. Case study in Romania," *Energy*, vol. 81, pp. 416–429,.

[12] Putraman Alkhairi and Agus Perdana Widarto "Application of K-Means Cluster in the Potential Areas of Productive Rubber Agriculture in North Sumatra" Seminar Nasional Teknologi Komputer dan Sains Hal. 762 – 767, 2019.

[13] Ade Bastian, Harun Sujadi and Gigin Febrianto "Application Of K-Means Clustering Analysis Algorithm In Human Infected Disease (Case Study Of Majaleng District" *Jurnal Sistem Informasi*, *14*(1), 28-34.

[14] Gede Aditra Pradnyana, Agus Aan Jiwa Permana (2018) "System Of Class Distribution With Students K-Means And K-Nearest Neighbors Methods For Improve Learning Quality" Jurnal Ilmiah Teknologi Informasi, Volume 16, Nomor 1; 59-68.

[15] Siska T.S (2016) "Analysis And Application Of Data Mining To Determine Selling Water Cubication Based On Customer Grouping Using K-Means Clustering Algorithm" Jurnal Teknologi Informasi & Pendidkan Vol 9, No 1.

[16] Rusdiana, Lili; Sam'ani, Sam'ani. (2017) Pemodelan K-Means Pada Penentuan Predikat Kelulusan Mahasiswa STMIK Palangka Raya. Jurnal SAINTEKOM, [S.l.], v. 6, n. 1, p. 1-15, ISSN 2503-3247.

[17] Feby F, Muhammad Najib B, Muhammad Arif T. (2019). "K-Means Cluster Analysis and Discriminant Analysis in the Grouping of Poor Villages in Pangkep Regency", Journal Of Statistics and Its Application On Teaching anf Research, Vol 1, No.2.

[18] Very K.B, Jtmiko I (2017) . "Clustering of Final Project Documents Using K-Means Clustering as Implementation Analysis Retrieval System". Jurnal Ilmiah ManaJEMEN Informatika dan Komputer, Vol 01, No. 01. Pp. 31-34.

[19] Adi Hazman, Purba D,K, Casi S (2019), "Application Of K-Means Clustering Method For Grouping Port And Unload Loading Goods Data In Indonesia". E-proceeding of Engineering Vol 6, No. 1, Page 1450.

[20] Hardani P.R, Wara P, Gangga A, (2020) Use Of Analysis Of K-Means In Claster Modeling Of Spatial Regression In Tuberculosis Case In East Java In 2017. Indonesian Jurnal Of Statistic and Its Applications Vol 4 No. , 164 – 178.

[21] I.C. DeAwi, B.Y. Gautama, and P.A. Mertasana, (2017). "Analysis of Clustering for Grouping of Productive Industry by K-Medoid Method," International Journal of Engineering and Emerging Technology., vol. 2, pp. 26-30.

[22] Mabrur A.G, R. L. (2012). Penerapan Data Mining untuk Memprediksi Program Studi Teknik Informatika Jurnal Komputer dan Informatika ( KOMPUTA ), 1.

[23] Nasari, F., Darma, S., & Informasi, S. (2015). Penerapan K-Means Clustring Pada Data Penerimaan Mahasiswa Baru, 6–8.

[24] Y. Agusta, "K-Means Penerapan, Permasalahan dan Metode Terkait," Jurnal Sistem dan Informatika, vol. 3, no. Pebruari, pp. 47–60, 2007.

[25] Khotimah, T., Teknik, D. F., Studi, P., Informatika, T., & Kudus, U. M. (2014). Pngelompokan Surat Dalam AL QUR ' AN Menggunakan Algoritma K-Means, 5(1), 83–88

[26] Irwanto, et. al (2012). *Optimasi Kinerja Algoritma Klasterisasi K-Means untuk kuantisasi Warna Citra*. Jurnal Teknik ITS, I(1), pp.197-202.

[27] N. Putu, E. Merliana, and A. J. Santoso, "Analisa penentuan jumlah cluster terbaik pada metode K-Means," Seminar Nasional Multi Disiplin Ilmu, pp. 978–979.

[28] Catia Oliveira, Tiago Gumaraes, Filipe Portela, Manuel Santos (2019) " Benchmarking Business Analytics Techniques in Big Data", Procedia Computer Science 160, 690-695.