Accuracy Analysis of Predictive Value in Transaction Data of Service Company Using Combination of K-Means Clustering and Time Series Methods

Santi Ika Murpratiwi[1]*, Dewa Ayu Indah Cahya Dewi[1], Arik Aranta[2]
[1] Dept Informatics Engineering, STMIK STIKOM INDONESIA
Jl. Tukad Pakerisan No. 97, Denpasar, Bali, INDONESIA
[2] Dept Informatics Engineering, Mataram University
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA
Email: [santiika, cahyadewi]@stiki-indonesia.ac.id, arikaranta@unram.ac.id

Corresponding Author

Abstract **Profit decline is a frightening problem for service companies. The solution to prevent this is by analyzing data transactions using data mining and forecasting. K-Means used to cluster the level of car damage based on the number of panels repaired and the duration of repaired. The results of K-Means used as material for analysis the best time-series method for transaction data. The methods analyzed include the moving average, single exponential smoothing, double exponential smoothing, and winter's method. Single exponential smoothing is the most suitable forecasting method with transaction data. Based on the MAPE value obtained for minor damage of 12.58%, forecasting for moderate damage of 16.83%, forecasting for major damage of 17.31%, and forecasting for overall data of 8.0975%. It concluded that single exponential smoothing can apply with K-Means clustering and the company can use it to make strategies to prepare the number of workers and production materials required.**

Keywords: **Data mining, K-Means, Forecasting, Time-Series, Exponential smoothing**

I. INTRODUCTION

Increasingly tight business competition causes companies to be able to develop strategies to maintain the credibility of their business. The main problem that business actors fear is the decline in company profits. The decline in profit is a measure of the health of a company, especially in the sector that sells services. The decline in profit at service companies caused by many things including internal factors and external factors. Problems that need to be considered first are internal problems such as a decrease in the quality-of-service work, poor workforce management, inappropriate management of tools and materials, setting targets that are too low, lack of company readiness to handle various levels of difficulty, and many other things. This problem is experienced by PT. XYZ is engaged in motor vehicle body repair.

A vehicle repair company or called a body and paint company is a company engaged in the repair of motor vehicles. This company specifically handles problems in car body parts, for example dealing with vehicle damage due to accidents and changing vehicle body parts with official dealer standards.

The number of vehicles repaired each month is fluctuating, making it difficult for the company to prepare a business strategy to maximize the profit it can get. The emergence of many competitors is a concern for the company so that the company must be able to develop a strategy based on existing data.

To make a strategy regarding the number of transactions needed transaction data. With the background of a large amount of company transaction data, data mining was chosen to assist in the data processing and analysis process. Even though the analysis process can using data mining only, but to increase the accuracy of the analysis, forecasting added. Forecasting is expected to help predict the sustainability of a company quantitatively [1]. Judging from the existing transaction data, confusion arises regarding the forecasting method used so this study focused on the selection best of forecasting methods by existing transaction data.

Based on the available transaction data, the data mining method used is K-Means Clustering. K-means are chosen because K-Means can classify transaction data based on the level of damage [2]. The variables used in the clustering process are the number of panels and the length of work. Based on these two variables, it is expected that the transaction data will be able to classify into 3 (three) groups, namely the minor damage group, the moderate damage group, and the major damage group seen from the number of panels repaired and the length of work.

Furthermore, data from the clustering process is used as a material to analyze the forecasting method that has the highest accuracy value for existing transaction data. Judging from the pattern of transaction data, the time series forecasting method chosen because the time series method refers to past transaction data with a certain time series [3]. Several time-series methods will be implemented namely the moving average, single exponential smoothing, double exponential smoothing, and the winter's method. Because forecasting does not always produce accurate forecasting results, some of the time-series methods compared to see the error value, so that the most appropriate forecasting method for transaction data at service companies grouped based on the level of damage.

II. LITERATURE REVIEW AND BASIC THEORY

A. Literature Review

Data mining is an alternative to reading data trends that occur. Data mining is capable of classifying, clustering, and forecasting large amounts of data. With data mining, many companies and institutions are helped to analyze the data in their company.

Data mining can be used in market analysis applications that have been developed to help analyze the level of purchase of a product to obtain purchase patterns that can help sellers in developing strategies, it is I Shopping application: Intelligent Shopping and Predicate Analysis System Using Data Mining. The amount of existing transaction data must be used to assist the company's managerial team in making decisions [4]. Besides, data mining can also be implemented to conduct sentiment analysis on positive and negative opinions from sentences in tweets related to tourism in Lombok [5].

Clustering data is an attempt to divide the data set into a related cluster so that all clusters have nothing in common. Clustering data is widely used to segment certain data. In a study by Venkatkumar [6], a comparison was made of the K-means clustering method, BIRCH (Balanced Iterative Reducing Clustering using Hierarchies), DBSCAN (Density-Based Spatial Clustering of application with Noise), and STING (Statistical Information Grid). It can be concluded that the K-means method is the simplest method that can provide good clustering results and is suitable for simple clustering with the number of clusters according to the user's wishes. Besides, K-Means is also compared with the FCM method to determine the school's marketing promotion strategy. The results show that K-Means can clustering data into the number of clusters according to user requests [7].

Based on the effectiveness and flexibility of using the K-Means clustering method, many studies have implemented clustering for various cases. K-Means is used to analyze the pattern of student enrollments at various universities in India based on the applicant's region of origin. The results of this study are expected to assist the government in making the distribution of higher education in India [8]. K-Means is also implemented in the retail company PT. Indomarco Palembang to get product sales

patterns [9]. Almost all studies using K-Means aim to determine certain patterns of data. This pattern is used to make decisions, in other words, the data pattern from the clustering results can be used to predict future conditions.

Accurate predictions can be made by forecasting using certain methods. Forecasting can be applied in many sectors, such as in determining the prediction of the number of tourist visits to West Nusa Tenggara (NTB) using past data within a certain period using the backpropagation method [10]. Besides, the time series method is widely used for forecasting regarding past data. The implementation of time series forecasting is carried out to predict the sales of freight forwarding services [11]. The time series method is also implemented to forecast earnings in XYZ company [12] and to predict the inflation rate in Indonesia [13]. From this research, it was found that the exponential smoothing method had the highest accuracy value.

Based on the literature review, there is found a research gap there is no research that discusses the use of clustering and time series forecasting method simultaneously. By using time series forecasting, it is hoped that it can increase the accuracy of the prediction results of data patterns to help management to simplify strategy formulation. Because several time series methods are available, in this study the most appropriate time series method is chosen from the accuracy value that matches the transaction data.

B. Basic Theory

B.1 K-Means Clustering

K-Means is a method for clustering data that is very popular in data mining. The principle of clustering is to maximize the similarity between members of one cluster and minimize similarity between clusters. Clustering can be done on data that has several attributes that are mapped as multidimensional spaces. K-Means is an unsupervised machine learning algorithm where K-means is done by determining the number of clusters to be formed and determining the initial centroid and calculating the distance of each data from the centroid and grouping them based on the closest distance [14].

The following are the steps for clustering using the K-Means method.
1. Determine the value of k as the number of clusters.
2. Allocate data into clusters randomly.
3. Calculate the centroid or average of the data in each cluster.
4. Allocate each data to the centroid (nearest average).
5. Return to the third step, if there are still data moving clusters or changes in the centroid value.

K-Means is considered quite efficient because it can cluster relatively quickly even though there is quite a lot of data and clusters formed. The K-Means clustering process stops when the centroid has stabilized and the number of iterations is as desired. The clustering was stable if no values have changed [15].

B.2 Forecasting

Forecasting is very important in any business aspect because forecasting can be used to estimate the uncertainty in the future. To get forecasts with high accuracy values, many forecasting methods can be used. The choice of forecasting method can be by the objectives and form of the available data. Broadly speaking, forecasting is divided into two, namely qualitative forecasting and quantitative forecasting [1].

Quantitative forecasting is more widely used because it uses and produces definite data. The quantitative method is a method of forecasting that relies heavily on historical data patterns from a certain period. This quantitative forecasting is used if there is a condition where information is available about the past, this information can be quantified in the form of data, this information can be assumed that the past pattern continues. One example of quantitative forecasting is forecasting production, income, and forecasting risk. Forecasting is very important because it is used as a work basis and a basis for decision making for the company. By knowing the results of the forecast, the company can make a business strategy following company goals [16]. The quantitative method that is widely used is the time series method.

The time series method in predicting the future based on the value of a variable in the past or mistakes made previously. It aims to examine data patterns used to predict and extrapolate to the future. This forecasting method uses the basic time series of forecasting the actual data that will then be predicted to determine the required data pattern [1]. Forecasting methods using time series, namely:

1. Moving Average

Moving average is a method of forecasting that is done by taking a group of values within a certain period and looking for the average value of these values to determine the forecast value for the next period. Historical data is taken in a certain period, the usual duration is 3, 6, and 12 months. If the historical data is taken 3 months moving average, then the forecast for the fifth month will only be made after the fourth month has finished/ended. The longer the moving average is, the more visible the effect of the slippage will be in the forecast results. The appearance of the equation must be shown as presented in Eq. 1.

$$S_{t+1} = \frac{X_t + X_{t-1} + \cdots + X_{t-n+1}}{n} \qquad (1)$$

Where the forecast value for the period t + 1 is denoted in. Forecast value is obtained by adding up the real value of the period t to t-1 then divided by the number (n) limits in the moving average.

2. Single Exponential Smoothing

Exponential smoothing is a weighted time series forecasting technique. Where data is given weight by an exponential function. The exponential smoothing method follows the pattern of data fluctuation observed in a period for future forecasts by smoothing or what is called smoothing and reducing the fluctuation of the forecast.

The single exponential smoothing method only requires two data points to predict the value that will occur in the future. The appearance of the equation must be shown as presented in Eq. 1.

$$F_{t+1} = \alpha X_i + (1 - \alpha)F_{t-1} \qquad (2)$$

The forecast value for period t + 1 can be calculated by multiplying the value by the real value in period t plus the result of minus 1 by multiplying by the forecast value in period t-1. This method requires a value as the value of the smoothing parameter. With the appropriate parameter value provides an optimal forecast with the smallest error value. The value of α is carried out by comparing using smoothing intervals between $0 < \alpha < 1$, namely α (0.1 to 0.9). This method is only able to provide forecasts for one period ahead and is suitable for data containing stationary elements. Because if it is applied to a data series that has a consistent trend, the forecast will always be behind the trend. Besides, this exponential method also gives a relatively higher weight to the most recent observed values than the values of the previous period.

3. Double Exponential Smoothing

Double exponential smoothing is a linear method proposed by Brown. This method is used when the data shows a trend. The trend is a smoothed estimate of the average growth at the end of each period [17].

The rationale for Double Exponential Smoothing from Brown is similar to Double Moving Average because both Single Smoothing and Double Smoothing values lag behind the actual data whenever there is a trend element. The difference between the Single Smoothing and Double Smoothing (-) values can be added to the single smoothing (S) values and adjusted for the trend.

The double exponential smoothing method is a prediction that only requires three data values and α values. The equations used in this method shown as presented in Eq. 3, Eq. 4, and Eq. 5

$$S'_t = \alpha X_t + (1 - \alpha)S'_{t-1} \qquad (3)$$

$$S''_t = \alpha S'_t + (1 - \alpha)S''_{t-1} \qquad (4)$$

$$S_{t+m} = a_t + b_t m \qquad (5)$$

To be able to use the formula, the value and must be available. But when t = 1, these values are not available. Since these values must be determined at the beginning of the period, to solve this problem can be done by setting and equal to the value (actual data) [18].

4. Winter's Method

Winter's method is a time series method that can be implemented if there is a seasonal pattern in addition to the trend data pattern. The Winters method uses three exponential equations with three parameters, namely stationery, trend, and seasonal elements. Each parameter is denoted in the form of alpha, gamma, and beta. The equations used in this method shown as presented in Eq. 6, Eq. 7, Eq. 8, and Eq. 9.

a. Overall Smoothing

$$S_t = \alpha \frac{X_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad (6)$$

b. Trend Smoothing

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \qquad (7)$$

(7)

c. Seasonal Smoothing

$$I_t = \beta \frac{X_t}{S_t} + (1-\beta)I_{t-L} \qquad (8)$$

d. Forecast Value

$$F_{t+m} = (S_t + b_t m)I_{t-L+m} \qquad (9)$$

Where L is the seasonal length, for example, the number of months or quarters in a year, b is the trend component, I is the seasonal adjustment factor, and is the forecast for the period m.

The Winters method also has drawbacks, the main drawback that hinders its widespread use, namely that it requires three smoothing parameters (alpha, beta, gamma) which can be valued between 0 and 1, so many combinations must be tried before the optimal parameter value is determined. An alternative method that can reduce doubts about the optimal value is to find a better initial value estimate, then assign small values to the three smoothing parameters (about 0.1 to 0.3). A value of 0.1 makes the forecast too cautious, while a value of 0.3 gives a more responsive system. Because of this narrower set of value choices, this method is usually seen as a method that is easier to use [19].

B.3 Forecast Error Measurement

Testing is carried out to determine the level of error in forecasting which is commonly called forecast error. Several methods can be used to calculate the forecast error. The following is a test of the forecasting accuracy carried out in this study.

1. Mean Absolute Deviation (MAD)

Mean Absolute Deviation (MAD) or absolute mean deviation is a measure of the overall forecast error for a model. The following are the equations used to calculate the MAD value.

$$MAD = \frac{\sum|A_t - F_t|}{n} \qquad (10)$$

Where A_t is the actual value of data and F_t is the value of forecasting data. Subtract the actual value from the predicted value divided by the amount of data.

2. Mean Squared Deviation (MSD)

MSD is the average of the difference between the squared and observed values. Here is an equation for calculating the Mean Squared Deviation (MSD).

$$MSD = \frac{\sum(A_t - F_t)^2}{n} \qquad (11)$$

Where A_t is the actual value of data and F_t is the value of forecasting data. Subtraction from the actual value and the forecast value is then squared and divided by the amount of data.

3. Mean Absolute Percent Error (MAPE)

MAPE is calculated as the average of the absolute differentiation between the predicted and actual values, expressed as a percentage of the actual value. MAPE is calculated using the following formula [20].

$$MSE = \frac{100 \sum_{i=1}^{n}(|A_t - F_t|)/A_t}{n} \qquad (12)$$

The feasibility level of using a forecasting method is seen from the resulting MAPE value. Table I shows the range of feasibility levels based on the MAPE value [20].

TABLE I. RANGE OF MAPE VALUE

Range MAPE	Meaning
< 10 %	The ability of very good forecasting model
10 - 20 %	The ability of a good forecasting model
20 - 50 %	The ability of proper forecasting model
> 50 %	The ability of bad forecasting model

III. RESEARCH METHOD

A. The Research Approach

The approach used in this research is quantitative. The reason for using a quantitative approach is that the output of this study is in the form of numbers. However, to assist the translation process of the research results it is equipped with a descriptive research method. The numbers that have been obtained are then searched for the relationship between all elements so that a conclusion can be drawn and described for analysis and forecasting studies on company transaction data. With a descriptive approach, it is hoped that the results of the analysis can provide new knowledge that can help the development of the field of forecasting.

B. Location and Time Research

The location of the research was carried out in a vehicle repair company, namely PT. XYZ is located in Denpasar, Bali. The subjects of this study are a set of company transaction data, namely data from January 2014 - December 2018. The stages of the research consisted of literature studies, data processing, calculation methods, analysis studies, conclusions, and report writing.

C. Research Flow and Data Collection Techniques

The flow of this research is divided into three stages as follows.

1. The First Stage of Research Flow

Following is the first stage of research flow, which is the process of understanding business and preprocessing data.

a. Collecting company business information by observing and searching for a list of data used for the analysis process.
b. Selection of the method used for the analysis process and adjusted to the availability of data in the vehicle body repair company.
c. Literature studies and data collection related to data processing using K-Means for data clustering, and time series methods consisting of moving averages, single exponential smoothing, double exponential smoothing, and winter's method for forecasting.
d. Sorting the necessary data by grouping data according to the purpose of the analysis, removing data redundancies, and sorting data according to priority.

2. The Second Stage of Research Flow

Following is the research flow in the second stage which is the method implementation process.

a. Extracting data into Microsoft Excel.
b. Applying the concept of clustering several k clusters to form data patterns using the K-Means method.
c. The results of the clustering process are transaction data clusters based on the level of damage determined by the number of damaged panels and the length of time to finish the vehicle repair. The forecasting process is carried out using the time series method, namely Moving Average, Single Exponential Smoothing, Double Exponential Smoothing, and Winter's method.
d. Read the pattern of data processing results to draw conclusions and represent them in the form of a report.

3. The Third Stage of Research Flow

The third stage research flow discusses the process of calculating the accuracy of the forecast results. The following is the third stage of the research flow.

a. Analyze the results of data processing and forecasting from reports generated by the previous process.
b. Perform data accuracy calculations using the forecast error method.
c. Knowing the accuracy value of the analysis results for each of the time series methods used.

d. Conclude, the time series method is most suitable to be applied to corporate data transaction data.

IV. RESULT AND DISCUSSIONS

The number of vehicle repair transactions in a period of 5 years from January 2014 to December 2018 shows an increasing trend. Good quality and service are the benchmarks for the increase in vehicle repair transactions from an internal perspective. From the external side, it can be seen that the increase in the number of vehicles circulating in the community is also the reason for the increase in the number of existing transactions. Judging by the stable upward trend, a strategy is needed to help increase transactions.

The data analysis in this study used 2.957 transaction data from 2014 - 2018. From this data, it can be seen that the data tends to increase steadily but experienced a decline in 2017. Fig. 1 is a graph of the number of vehicle repair transactions.
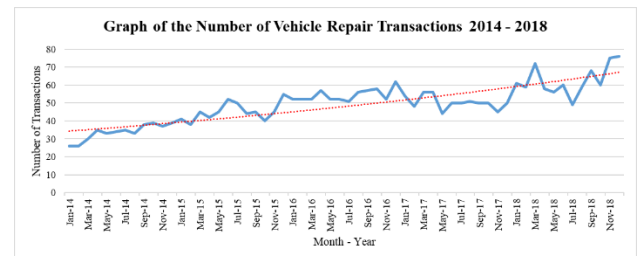

Fig. 1. Graph of the Number of Vehicle Repair Transaction 2014–2018

Based on the number of transactions that occurred, the company wants the transactions to be analyzed based on the level of damage. In the vehicle body repair company, the provision of vehicle repair tools and materials requires a large amount of capital. So that to minimize the capital that must be spent, it is necessary to classify the level of damage. The level of damage is classified using the K-Means method. The grouping is based on the number of vehicle repair panels and the duration of repairs.

The number of clusters formed was three, namely minor damage, moderate damage, and major damage. The iteration process is carried out 8 times. In the 8th iteration, the data no longer changes so the clustering process is stopped. Table II is the resulting centroid cluster.

TABLE II. RESULT OF CENTROID CLUSTER

Variable	Cluster 1	Cluster 2	Cluster 3	Grand Centroid
Duration of repairs	5,7574	18,4344	45,254	16,9838
The number of panels	3,0993	11,333	23,9839	9,4893

Based on the centroid cluster, it can be analyzed that the vehicle is said to be minor damaged if it has some repair panels 3 and 6 days processing time and thereafter for moderate and major damage. Table III shows the results of clustering using K-Means.

TABLE III. K-MEANS CLUSTERING RESULTS

Total Data	Cluster 1 (Minor Damage)	Cluster 2 (Moderate Damage)	Cluster 3 (Major Damage)
2.957	1.521	877	559

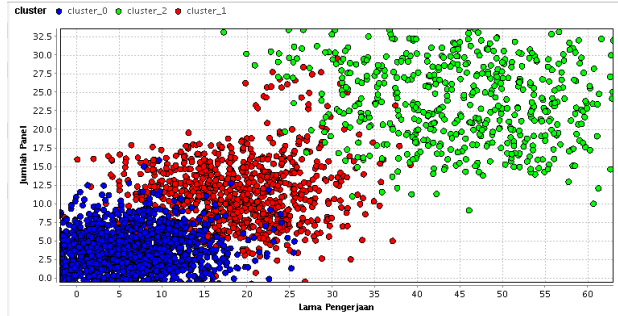The distribution of cluster members can be seen in Fig. 2 below.

Fig. 2. Graph of Cluster Distribution

Each data has its own cluster identity. The results of grouping based on the level of damage are used as material for forecasting. Figure 3 is a graph of the distribution of the number of transactions based on the level of damage.
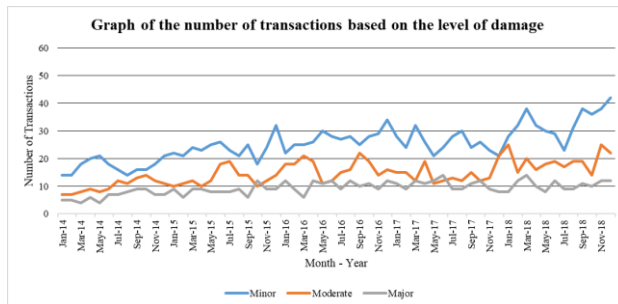
Fig. 3. Graph of the transaction based on the level of damage

The analysis is continued by forecasting data using the time series method. The following are the results of the analysis of each periodic series method.

1. Moving Average

In this study, the moving average method of the last three periods was used. Each of the results of the damage level classification is analyzed using moving average forecasting. In Fig. 4 is a graph of the results of forecasting the overall data, Fig. 5 is a graph of the results of forecasting data with a minor level of damage, Fig. 6 is the result of forecasting moderate damage level data, and Fig. 7 shows a graph of the results of forecasting data on the level of major damage.
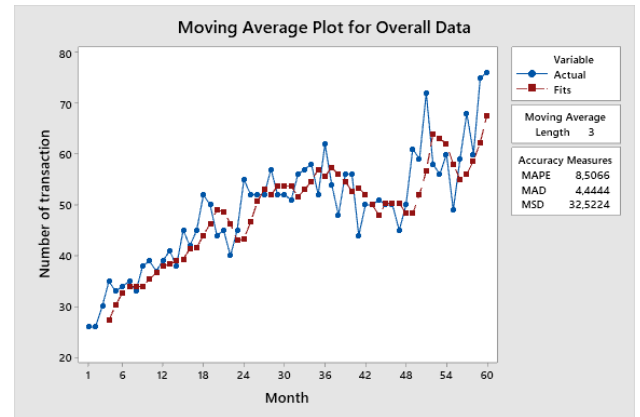
Fig. 4. Moving Average for Overall Data
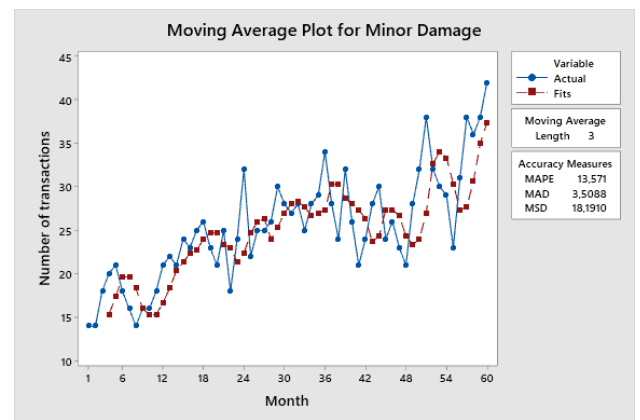
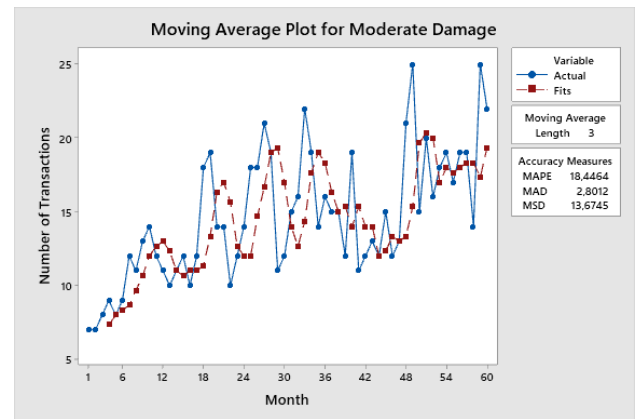Fig. 5. Moving Average for Minor Damage Level

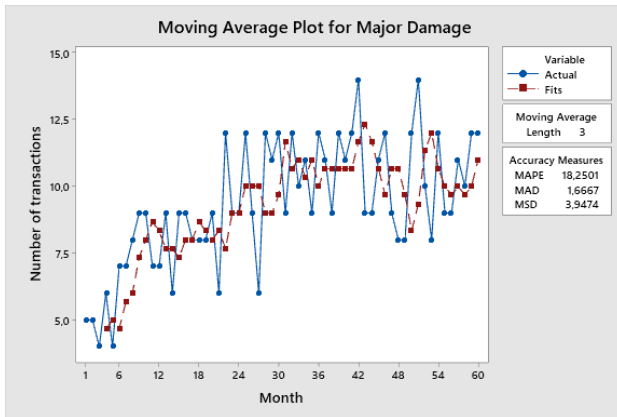Fig. 6. Moving Average for Moderate Damage Level

Fig. 7. Moving Average for Major Damage Level

From the results of forecasting using the three-period moving average method, it can be seen that the highest MAPE value is obtained in forecasting the overall transaction data with a value of 8.5066 indicating that this method is very good to be implemented in data forecasting as a whole. However, the MAPE value for forecasting based on the level of damage ranges from 13 - 18, it is still in the good category so the three-period moving average method can be considered.

2. Single Exponential Smoothing
Forecasting using the single exponential smoothing method requires an alpha constant. The determination of alpha constants is assisted by using the ARIMA optimization provided by the Minitab application. In Figure 8 is a graph of the results of forecasting the overall data with the single exponential smoothing method, Fig. 9 is a graph of the results of forecasting the level of minor damage, Fig. 10 is a graph of moderate damage forecasting results, and Fig. 11 is the result of forecasting the level of major damage with single exponential smoothing.
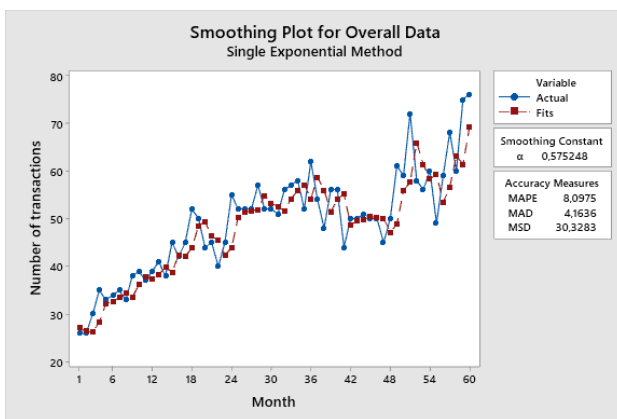

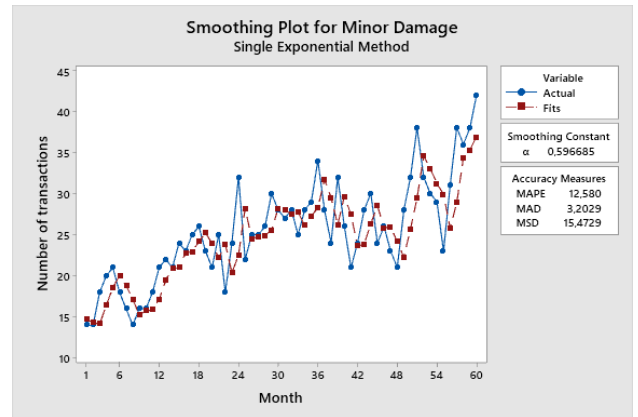Fig. 8. Single Exponential Smoothing for Overall Data

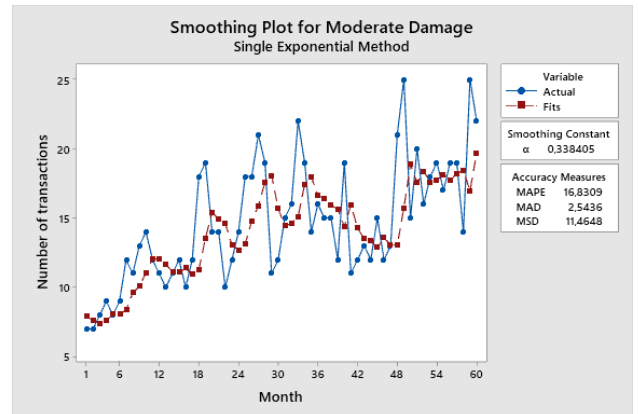Fig. 9. Single Exponential Smoothing for Minor Damage Level

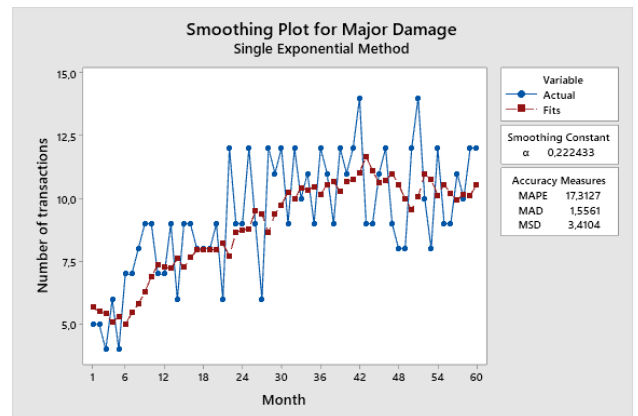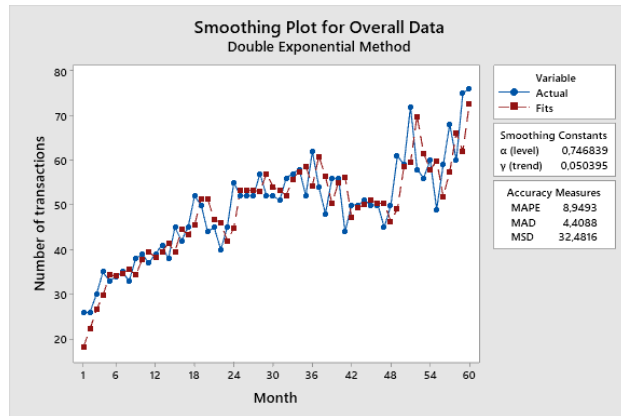Fig. 10. Single Exponential Smoothing for Moderate Damage Level

Fig. 11. Single Exponential Smoothing for Major Damage Level

Forecasting results with Single Exponential Smoothing when compared with the Moving Average, the results are better at forecasting Single Exponential Smoothing. This can be seen from the decrease in MAPE value in data forecasting based on the level of damage. The results of the analysis using the Single Exponential smoothing method are said to be good for forecasting vehicle repair transaction data, especially for forecasting the overall data when viewed from the MAPE value obtained.

3. Double Exponential Smoothing

In forecasting using the Double Exponential Smoothing method, two smoothing constants are needed, namely alpha and gamma. This method requires a trend. If the available data have a trend then this method is worth considering. In Fig. 12 is a graph of the results of forecasting Double Exponential Smoothing for the overall transaction data, Fig. 13 is the forecasting result for minor damage, Fig. 14 is the forecasting result for moderate damage, and Fig. 15 is the forecasting result for major damage.

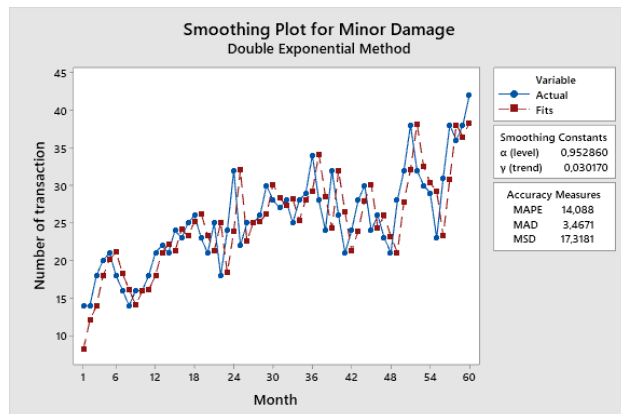Fig. 12. Double Exponential Smoothing for Overall Data

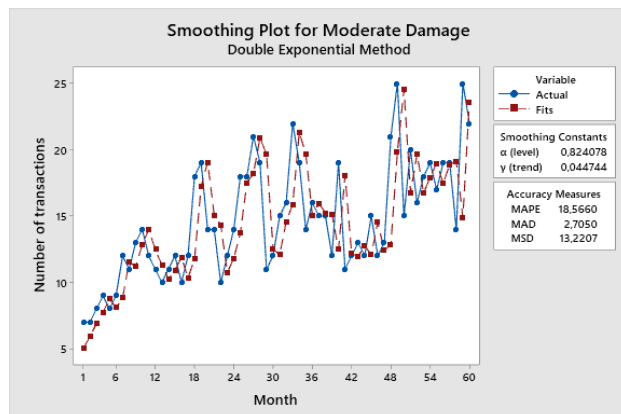Fig. 13. Double Exponential Smoothing for Minor Damage Level

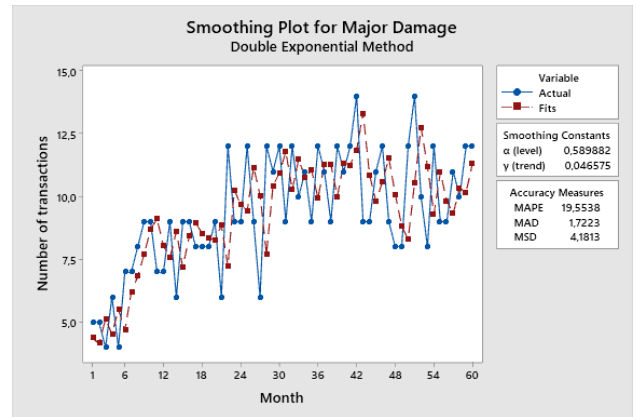Fig. 14. Double Exponential Smoothing for Moderate Damage Level

Fig. 15. Double Exponential Smoothing for Major Damage Level

The results of forecasting analysis using the Double exponential smoothing method can be said that this method is a suitable method for existing data. It can be seen in Fig 1 that the transaction data has trend data. The analysis is carried out by looking at the graph, wherefrom the 4 graphs the forecasting results show a relatively small MAD value even though the resulting MAPE value is good for forecasting based on the level of damage and is very good for overall forecasting.

4. Winter's Method

The winter method is suitable if the data has a trend and is seasonal. To see whether transaction data is seasonal, Winter's method is implemented as a comparison. In the Winters method, three smoothing constants are needed, namely alpha, gamma, and beta. In Fig. 16 is a graph of forecasting the overall data using the Winters method, Fig. 17 is the result of forecasting minor damage, Fig. 18 is the result of forecasting moderate damage, and Fig. 19 is the result of forecasting major damage.
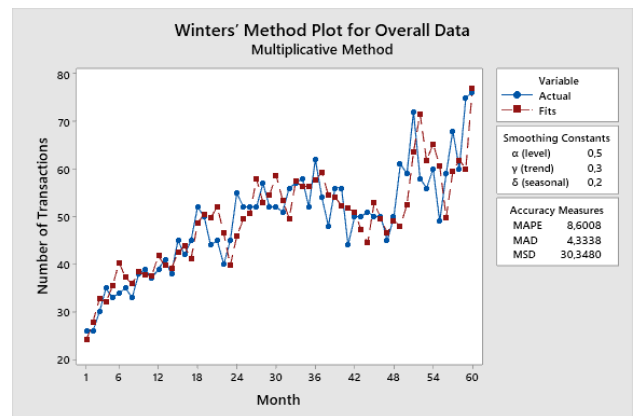

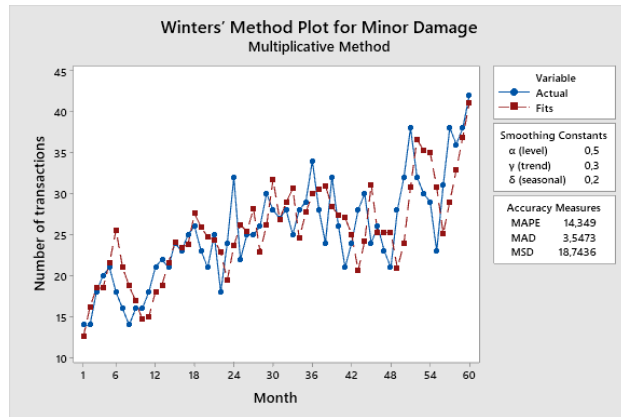Fig. 16. Winters Method for Overall Data

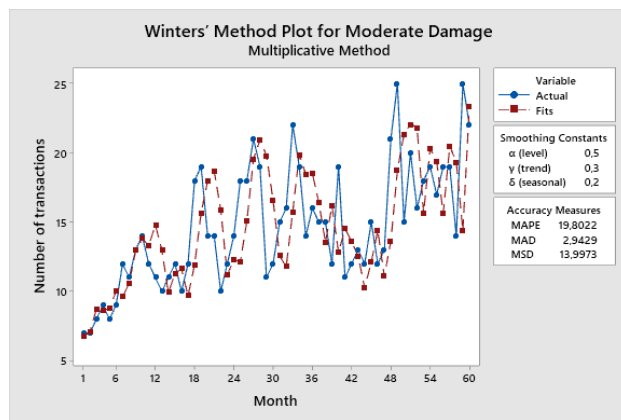Fig. 17. Winters Method for Minor Damage Level

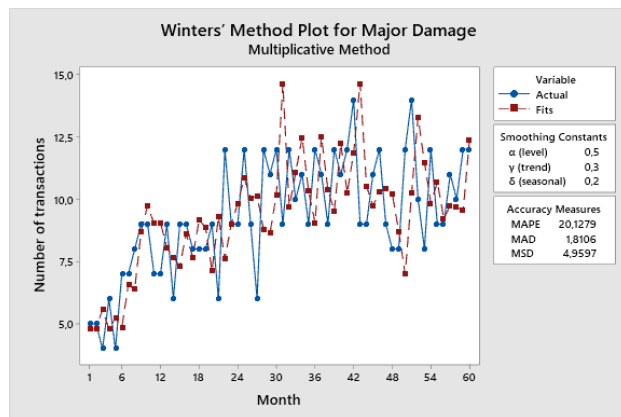Fig. 18. Winters Method for Moderate Damage Level

Fig. 19. Winters Method for Major Damage Level

Judging from the MAPE value obtained from forecasting using the Winters method, the MAPE value for forecasting major damage is above 20% which means it is feasible. For forecasting, the overall data has a small MAPE value and is in the very good category.

Table IV is a summary of the results of the MAPE, MAD, and MSD values from the forecasts that have been carried out. The selection of the right forecasting method can be seen from the percentage error obtained from each method. Based on Table IV, it can be seen that exponential

smoothing has a very good MAPE value for each forecast including forecasting the overall data, the level of minor damage, the level of moderate damage, and the level of major damage.

TABLE IV. PERCENTAGE ERROR OF TIME SERIES METHOD

Forecasting Method	Percentage		
	MAPE	MAD	MSD
Moving Average For Minor Damage	13,5718	3,5088	18,191
Moving Average For Moderate Damage	18,4464	2,8012	13,6745
Moving Average For Major Damage	18,2501	1,6667	3,9474
Moving Average for Overall Data	8,5066	4,4444	32,5224
Single Exponential Smoothing for Minor Damage	12,58	3,2029	15,4729
Single Exponential Smoothing for Moderate Damage	16,8309	2,5436	11,4648
Single Exponential Smoothing for Major Damage	17,3127	1,5561	3,4104
Single Exponential Smoothing for Overall Data	8,0975	4,1636	30,3283
Double Exponential Smoothing for Minor Damage	14,0887	3,4671	17,3181
Double Exponential Smoothing for Moderate Damage	18,566	2,705	13,2207
Double Exponential Smoothing for Major Damage	19,5538	1,7223	4,1813
Double Exponential Smoothing for Overall Data	8,9493	4,4088	32,4816
Winters Method for Minor Damage	14,3495	3,5473	18,7436
Winters Method for Moderate Damage	19,8022	2,9429	13,9973
Winters Method for Major Damage	20,1279	1,8106	4,9597
Winters Method for Overall Data	8,6008	4,3338	30,348

But if seen in Figure 1 that the available data has a trend, the best forecast to choose is double exponential smoothing with two smoothing constants.

V. CONCLUSION AND SUGGESTION

Based on the analysis of company transaction data, the use of the K-Means clustering method can assist companies in dividing transactions based on the level of damage, namely minor, moderate, and major. Transaction data with a total of 2,957 data were successfully divided into 3 groups for forecasting. Furthermore, based on the experimental application of the periodic series method to transaction data, the following conclusions can be drawn.
1. By comparing the values of MAPE, MSD, and MAD for the four methods with the smallest value is exponential smoothing, especially single exponential smoothing. However, if it is seen from the data trend

that the transaction data has double exponential smoothing been an alternative option.
2. To get the appropriate forecasting results, trial and error experiments are needed in determining the smoothing constant.
3. The selection of the right forecasting method according to the data conditions and the homogeneity of the available data where different data conditions can produce different forecasting results.
4. Combining data clustering and forecasting can be implemented to help break down data so that forecasting results are more accurate.

REFERENCES

[1] S. Wang and W. Chaovalitwongse, "Evaluating and Comparing Forecasting Models," *Wiley Encycl. Oper. Res. Manag. Sci.*, no. February 2011, DOI: 10.1002/9780470400531.eorms0307.

[2] N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction," *J.ICT Res. Appl. ITB Publ.*, vol. 11, no. 2, pp. 168–184, 2017, DOI: 10.5614/ itbj.ict.res.appl.2017.11.2.4.

[3] D. Fang, Y. Zhang, K. Spicher, and A. Revinda, "Forecasting Accuracy Analysis based on Two New Heuristic Methods and Holt-," 2011.

[4] N. D. Hettiarachchi, "I Shopping : Intelligent Shopping and Predicate Analysis System Using Data Mining," *IEEE Int. Conf. Data Sci. Data-Intensive Syst.*, pp. 1–8, 2015, DOI: 10.1109/DSDIS.2015.75.

[5] M. A. Ulfa, B. Irmawati, and A. Y. Husodo, "Twitter Sentiment Analysis using Naive Bayes Classifier with Mutual Information Feature Selection," *J-COSINE*, vol. 2, no. 2, pp. 106–111, 2018, DOI: https://doi.org/10.29303/jcosine.v2i2.120.

[6] I. A. Venkatkumar and S. J. K. Shardaben, "Comparative study of Data Mining Clustering Algorithms," in *2016 International Conference on Data Science and Engineering (ICDSE)*, 2016, DOI: 10.1109/ICDSE.2016.7823946.

[7] D. T. Kusuma and N. Agani, "Prototipe Komparasi Model Clustering Menggunakan Metode K-Means Dan FCM untuk Menentukan Strategi Promosi : Study Kasus Sekolah Tinggi Teknik-PLN Jakarta," *J. TICOM*, vol. 3, no. 3, pp. 1–10, 2015.

[8] R. Ahlawat, S. Sahay, S. Sabitha, and A. Bansal, "Analysis of factors affecting enrollment pattern in Indian universities using k-means clustering," *Int. Conf. Inf. Technol. - Next Gener. IT Summit*, pp. 321–326, 2016.

[9] Sutrisno, Afriyudi, and Widiyanto, "Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Studi Kasus PT. Indomarco Palembang," *J. Ilm. Teknol. Inform. Ilmu Komput.*, vol. x, no. x, pp. 1–11, 2013.

[10] K. T. N. Lestaru, M. A. Albar, and R. Afwani, "Penerapan Metode Backpropagation Dalam Memprediksi Jumlah Kunjungan Wisatawan Ke Provinsi Nusa Tenggara Barat (NTB)," *J-COSINE*, vol. 3, no. 1, pp. 39–48, 2019, doi: https://doi.org/10.29303/jcosine.v3i1.236.

[11] R. D. Syahputra, Suharyono, and Supriono, "Peramalan Penjualan Jasa Freight Forwarding Dengan Metode Single Moving Averages, Exponential Smoothing, Dan Weighted Moving Averages," *J. Adm. Bisnis*, vol. 55, no. 2, pp. 113–121, 2018.

[12] M. A. Maricar, "Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ," *J. Sist. dan Inform.*, vol. 13, no. 2, pp. 36–45, 2019.

[13] Y. Rismawanti and M. Y. Darsyah, "Perbandingan Peramalan Metode Moving Average dan Exponential Smoothing Holt Winter Untuk Menentukan Peramalan Inflasi di Indonesia," *Pros. Semin. Nas. Mhs. Unimus*, vol. 1, pp. 330–335, 2018.

[14] T. M. Kodinariya and D. P. R. Makwana, "Review on determining the number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, November 2013, 2013.

[15] S. Shukla and S. Naganna, "A Review ON K-means DATA Clustering Approach," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 17, pp. 1847–1860, 2014.

[16] S. Xu and X. Qi, "A comparison of time series methods for forecasting container throughput," *Int. J. Logist. Res. Appl.*, vol. 22, no. 3, pp. 294–303, 2019, DOI: 10.1080/13675567.2018.1525342.

[17] H. V. Ravinder, "Forecasting With Exponential Smoothing – What's The Right Smoothing Constant?," *Rev. Bus. Inf. Syst.*, vol. 17, no. 3, pp. 117–126, 2013.

[18] E. Pujiati, D. Yuniarti, and R. Goejantoro, "Peramalan Dengan Menggunakan Metode Double Exponential Smoothing Dari Brown (Studi Kasus : Indeks Harga Konsumen (IHK) Kota Samarinda)," *J. EKSPONENTIAL*, vol. 7, 2016.

[19] I. Sungkawa and R. T. Megasari, "Penerapan Ukuran Ketepatam Nilai Ramalan Data Deret Waktu Dalam Seleksi Model Peramalan Volume Penjualan PT Satriamandiri Citramulia," *ComTech*, vol. 2, no. 2, pp. 636–645, 2011.

[20] A. H. Hutasuhut, W. Anggraeni, and R. Tyasnurita, "Pembuatan Aplikasi Pendukung Keputusan untuk Peramalan Persediaan Bahan Baku Produksi Plastik Blowing dan Inject Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average) di CV. Asia," *J. Tek. Pomits*, vol. 3, no. 2, p. A-169-A-174, 2014, doi: https://dx.doi.org/10.12962/j23373539.v3i2.8114.