J-COSINE (Journal of Computer Science and Informatics Engineering)
Vol. 7, No. 2, December 2023 E-ISSN:2541-0806
Accredited Sinta-4 by RISTEKDIKTI Decree No. 79/E/KPT/2023 P-ISSN:2540-8895

Implementation of an Information System for Transaction Data Analysis Using the Holt Winters Method

Santi Ika Murpratiwi[1]*, Rukmi Sari Hartati[2], Made Sudarma[2]

[1]Dept Informatics Engineering, Mataram University
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA
[2] Dept Electrical Engineering, Udayana University
Jl. Raya Kampus Unud, Jimbaran, Bali, INDONESIA

Email: santiika@staff.unram.ac.id, *:* [rukmisari, msudarma]@unud.ac.id

Corresponding Author

Abstract **Errors in decision making are very fatal in company operations and often result in reduced profits. In making the right decisions, in-depth analysis of internal and external data is required. One effort to increase the accuracy of decision making is to apply a forecasting algorithm for existing transaction data. In this research, data mining and forecasting are used for car repair companies. The information system being built is expected to be able to display a list of satisfied customer data, group transaction data based on the level of damage, and provide forecasts for the number of panels, number of units, and total revenue for the next period. From this data, it will be easier for management to determine the number of workers, additional tools and stock of materials for the vehicle repair process. This information system was built by applying three algorithms, namely C4.5, K-Means, and Holt Winter Multivariative. The C4.5 algorithm will help group customer types, K-Means helps to cluster vehicle damage levels, and Holt winters is used for experience. The tests carried out included testing the accuracy of each algorithm, testing the black box system, and testing the system time so that the accuracy results were 100% for the C4.5 classification algorithm, 92.28% for the accuracy of K-Means which was tested with the purity measure. Meanwhile, for forecast testing, the MAPE value was between 9 – 11% for each forecast. So, the forecasting system built already has good accuracy**

Key words**: Holt Winters, K-Means, C.4.5, Forecasting**

I. INTRODUCTION

In Information Technology, data processing is carried out to enrich, mine and clarify information obtained using certain algorithms. One example is the development of decision-making systems, expert systems, data analysis, and data forecasting. So currently many companies or agencies are competing to process data with the aim of obtaining information that is useful for the sustainability of the company. One of them is a car repair company which was used as the object of this research.

A car repair company is a company that operates in the field of repairing motorized vehicles, especially cars. In car repair companies there are three types of customers, namely personal customers, insurance customers and partner customers. Private customers are customers whose repair costs are borne independently by the individual, while insurance customers are customers who have joined vehicle insurance and all repair costs are borne by the insurance company in accordance with the agreement and claims that have been made. Partner customers are customers who have become partners. the company so that it gets special treatment and prices in accordance with the agreement that has been made. Examples of partner customers are rental car, travel and hotel companies.

All business processes are carried out computerized through a claims information system provided by insurance and manual recapitulation by the repair shop. However, in its implementation, discrepancies often occur and data storage is still semi-structured because the composition of transaction data is incomplete and there are many empty fields and data redundancies. Car repair company transaction data is stored in conventional archive form, where transaction data is only used for reporting monthly and annual income. The data available in companies is very large, but companies are still not utilizing the data optimally. This can be seen from the company's difficulties in planning business development activities, difficulties in making decisions, and difficulties in evaluating company performance. In addition, the company's income level often experiences ups and downs, causing the company's sustainability to falter. So, external problems also occur, such as many transactions being rejected and being put on a waiting schedule due to limited production space and limited workers, which causes disappointment and many customers move to other companies. So, there are many complaints regarding the availability of spare parts and workmanship results that are less than optimal, causing a decrease in customer satisfaction levels which affects the company's credibility. With internal and external problems, an analysis system is needed that is equipped with complex data processing processes. With a very large amount of transaction data, data mining was chosen to help with data analysis which

is equipped with data forecasting to help predict the company's sustainability quantitatively.

Based on the semi-structured data available at the company, the first thing to do is classify customer satisfaction data using the C4.5 method. The results of the classification obtained data on satisfied customers and dissatisfied customers. Satisfied customer data will be used for clustering vehicle damage levels using the K-Means method. Furthermore, the K-Means results will be used as material for forecasting using the Holt Winters method.

The C4.5 method was chosen because it has an accuracy rate of more than 88% and has a greater level of truth than the naïve Bayes method [1]. K-means was chosen because it is included in partitioning clustering and is very suitable for making certain predictions [2]. K-Means also has a high level of accuracy, namely 89.4% [3], and 95.8% accuracy when combined with classification [4]. Holt Winters was chosen for forecasting because it is a lightweight forecast, has constant results, and can be used to see the trend, level and seasonality of a prediction [5]. In addition, Holt winters are suitable for long-term and seasonal predictions [6].

In this way, the analysis system built is expected to be able to assist in analyzing and predicting the number of vehicles, the number of repair panels, and the amount of income each month based on the level of vehicle damage (light, medium, and heavy) so that it can help company management.

II. Literature Review

Data Mining is a semi-automatic process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify potential and useful knowledge information stored in large databases [7]. Data Mining is a process of automatic or semi-automatic exploration and analysis of large amounts of data with the aim of finding meaningful patterns or rules [8].

There is research that discusses the creation of market analysis software called I Shopping. This software functions to analyze transaction data based on the level of purchase of a product with the aim of obtaining purchasing patterns that can help sellers. This research is motivated by the development of an e-commerce sales model where sellers must pay attention to consumer purchasing patterns and market desires in order to gain consumer trust [9].

The implementation of K-Means is used to analyze the pattern of student enrollment at various universities in India based on the registrant's city of origin with the aim of knowing trends in total enrollment, trends in enrollment from foreign students, and trends in community enrollment. The results of this research are expected to help the government in implementing equality. higher education in India [10].

Apart from processing data by means of Clustering, in Data Mining there is also a classification process. Use of the classification method, namely the C4.5 Algorithm, to determine disease in fish. This research was conducted with sample data originating from the Kediri City Fisheries Service. The amount of data used in this research was 1,120 data which was divided into 6 types of disease. The accuracy of the C4.5 algorithm can be seen from the accuracy value produced in this study, the lowest was 55.3% and the highest was 88.4% [11].

Combining two data processing processes is often carried out in an analysis by utilizing Data Mining to increase the accuracy of the analysis results. K-Means clustering is the first thing to do to divide data or actions into several clusters. Next, the cluster will be classified into normal or dangerous using the C4.5 Algorithm. In the accuracy testing process, comparisons were made with similar methods such as ID3, Naïve Bayes, KNN, SVM, TCM-KNN, and K-Means + C4.5 Algorithm. The results of the accuracy test showed that K-Means clustering which was supported by the C4.5 algorithm with classification obtained a score of 95.8% [3].

In the field of analysis, Forecasting is very necessary to answer an analysis quantitatively or qualitatively. Forecasting methods can be used to predict difficult situations in the development of the business world through internet transactions. Customer patterns also need to be considered [5].

III. Research Methodology

A. Research Approach

The approach used in this research is a quantitative approach because the output of this research is in the form of numbers. However, to help the translation process it is carried out using descriptive methods. The numbers that have been obtained are used to look for relationships between elements so that conclusions can be drawn. With a descriptive approach, it is hoped that the results of the analysis can provide new knowledge that can help the company progress.

B. Research Sites

The research location was carried out at a vehicle repair company called CV. Primadadi Motor located in Denpasar, Bali. The subject of this research is a collection of company transaction data, namely data from January 2013 – December 2016.

C. Types and Source Data

The types of data used in this research are as follows.

1. Primary Data

The primary data used is questionnaire data, transaction data and vehicle data obtained from the company within a certain period of time and the company's Standard Operation Procedure (SOP) data which is used to understand business processes within the company.

2. Secondary Data

In this research, the secondary data used is data from studies or research which is usually in the form of documents, literature, archives, or journals, and all files

related to the research topic and can be used as material for consideration..

D. Research Flow

The research flow is a flowchart to describe the data flow process in system design and creation. The research flow is divided into four stages as follows.

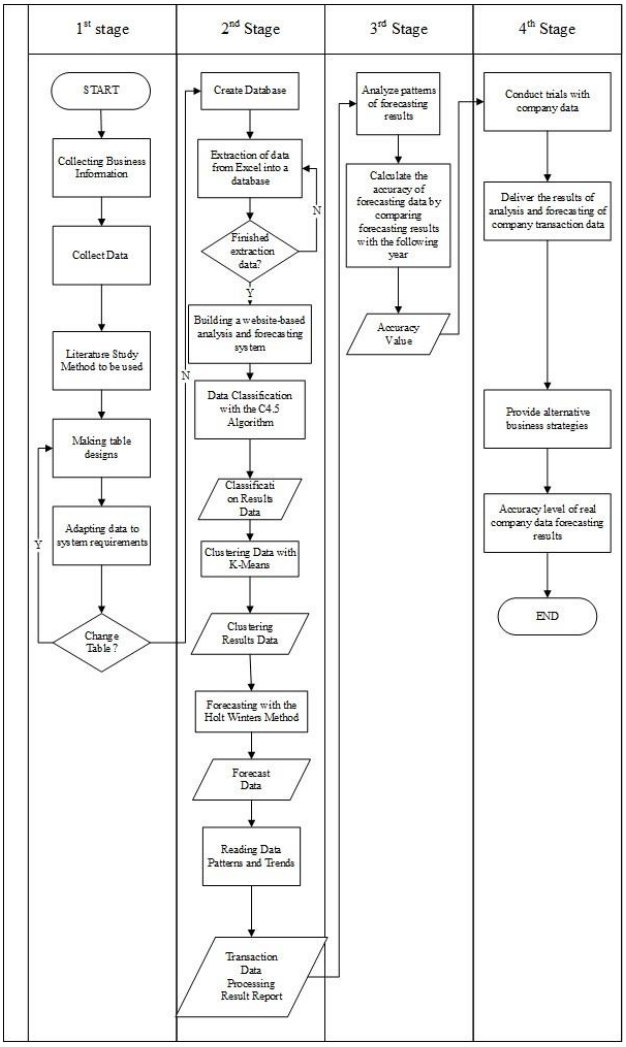

Fig. 1. Research Flow

1. Stage 1 Research Flow

The first stage was carried out to obtain data and information starting with collecting company business information, selecting a method based on literature studies related to the C4.5, K-Means, and Holt Winters methods, then designing a database to match the database design of the previous system.

2. Phase 2 Research Flow

The second stage is the information system implementation stage which begins with building a database using MySQL based on the design in stage 1, followed by extracting data from Excel files into a database, building an analysis and forecasting system interface using PHP and Bootstrap 5. In the information system The C4.5 algorithm was built to classify customer satisfaction data, the classification data was then clustered in a number of k clusters to form a data pattern using the K-Means method, the results of the clustering were used as forecasting material using the Holt-Winters method to predict the number of units, panels , and company revenue.

3. Stage 3 Research Flow

The third stage is the analysis stage which is carried out by analyzing the results of processing and forecasting money data generated by the system by calculating the accuracy of each algorithm to ensure the correct use of the algorithm.

4. Stage 4 Research Flow

The fourth stage is testing the system that has been built. Tests carried out include testing method accuracy, testing system functionality using the black box method, and testing data processing time. In the testing stage, the process of submitting the analysis results along with alternative business strategies that must be carried out by the company is carried out based on the results of the data analysis.

E. System Analysis and Design

General overview of the transaction data analysis information system is a system that provides analysis, forecasting and reporting needs for vehicle repair companies. This information system is used to process and store all questionnaire data and transaction data which were previously stored in excel format files. The format of the questionnaire data and transaction data used has been determined to match the data format. In the data pre-processing process, questionnaire data and transaction data are first checked for completeness of the data fields and suitability of the data format, where then valid questionnaire data and transaction data are imported into the system database that has been built. Customer satisfaction questionnaire data is classified using the C4.5 algorithm. Customer satisfaction questionnaire data is divided into two types, namely training data and test data. Customer satisfaction questionnaire data from 2013 will be used as training material to build a decision tree and data from 2013 to 2016 will be used as test data. The output in this data classification process, customer satisfaction questionnaire data is divided into two, namely satisfied and dissatisfied transaction questionnaire data. The flowchart for implementing the C4.5 algorithm is as shown in Figure 2.

Next, customer satisfaction questionnaire data with satisfied status from the data classification process will be grouped using K-Means Clustering. Satisfied customer transaction data was chosen because satisfied customers have a greater possibility of returning to make transactions with the company in the future. Transaction data from satisfied customers is grouped into 3 (three) clusters based on the level of vehicle damage, namely light, heavy and moderate. Determining the level of damage uses the work time field and number of panels in the transaction data. This is because the number of panels and the length of work can indicate the level of damage to the vehicle. The output of

the clustering process is data grouped into three based on the level of damage, namely light transaction data, medium transaction data and heavy transaction data. Figure 3 is a flowchart of the K-Means algorithm.

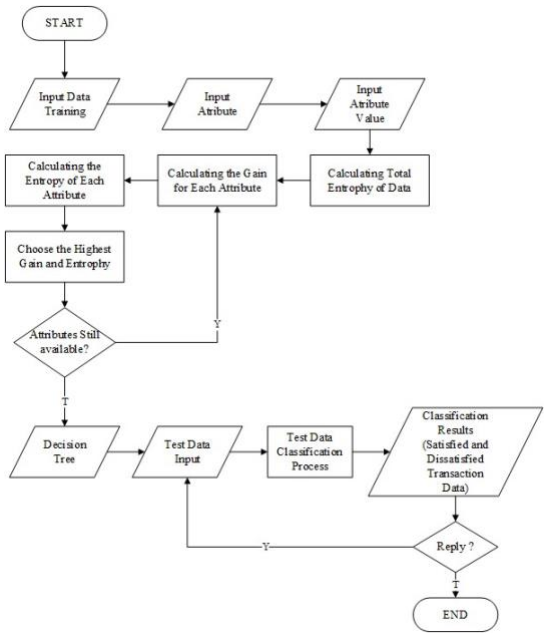

Fig. 2. C4.5 Classification Algorithm Flowchart

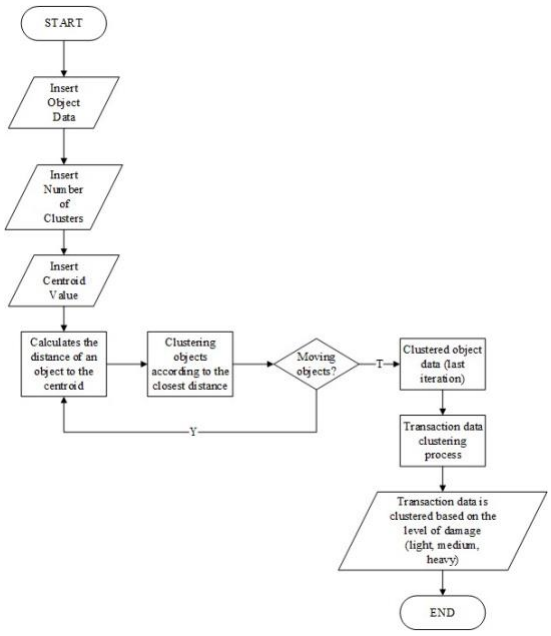

Fig. 3. Clustering K-Means Algorithm Flowchart

The clustering data is then used as material for forecasting using the Holt Winters method which is described in Fig. 4. Forecasting includes forecasting the number of vehicle panels, number of vehicle units, and company revenue. The output of this forecasting process is the result of forecasting the number of units, number of panels, and amount of income with light, medium, heavy, and overall damage.

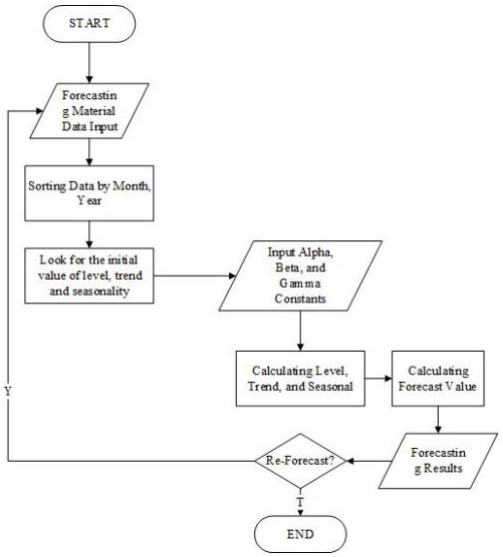

Fig. 4. Flowchart Metode Holt Winters

To make it easier for companies to carry out analysis of vehicle repair transaction data, this analysis and forecasting system is equipped with reports which include transaction data reports, satisfied customer data reports, vehicle damage level reports, and forecasting results reports.

The developer of this analysis and forecasting system uses the Agile Model software development methodology which has stages, namely conception, initiation, analysis, design, construction, testing and deployment. Where the stages in the Agile Model are more practical because processes can be carried out simultaneously without waiting for other processes to complete.

Testing in this research is divided into two, namely results testing and system testing. Results testing was carried out by calculating the accuracy of data processing with the C4.5 Algorithm, K-Means clustering, and Holt Winters forecasting. Meanwhile, system testing is carried out using black box testing to measure system functionality.

F. Systems Architecture

Business processes are represented in the form of Data Flow Diagrams (DFD). In Figure 5 is DFD Level 0.

Based on the data flow, there are several data that are input into this information system, namely questionnaire data, transaction data, vehicle data, training data, attribute data, attribute value data, centered data, and forecasting data. Meanwhile, the output of this information system is transaction data reports, satisfied customer data reports, vehicle damage level data reports, and data forecasting results reports. Next, a table relationship diagram is created which is used in the Transaction Data Analysis information system as in Figure 6.

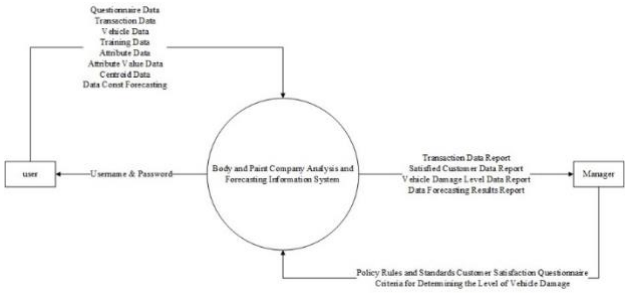

Fig. 5. Data Flow Diagram Level 0

Fig. 6. Diagram relasti tabel

The list of tables used in this information system are User Table, Questionnaire Data Table, Vehicle Data Table, Transaction Data Table, Training Data Table, C4.5 Attribute Table, C4.5 Attribute Value Table, Iteration Table, Mining Table, Decision Tree Table , Classification Results Table, Object Data Table, Centroid Data Table, Light Data Table, Medium Data Table, Heavy Data Table, Unit Data Table, Panel Data Table, Income Data Table, Clustering Results Table, Forecasting Constants Table, Panel Forecasting Results Table, Income Forecasting Results Table, and Unit Forecasting Results Table.

G. Interface Design

The interface design is created to help represent the appearance you want to create. Figure 7 is the interface design for the home page and Figure 8 is the interface design for the data import page.

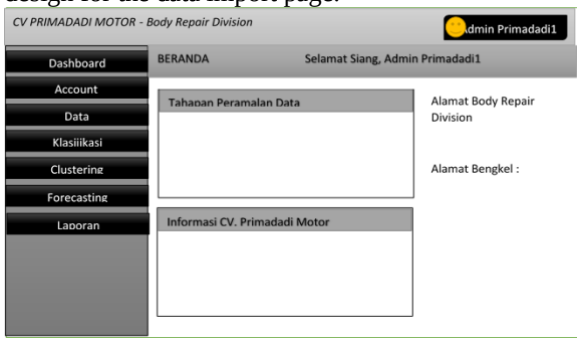

Fig. 7. Homepage interface design

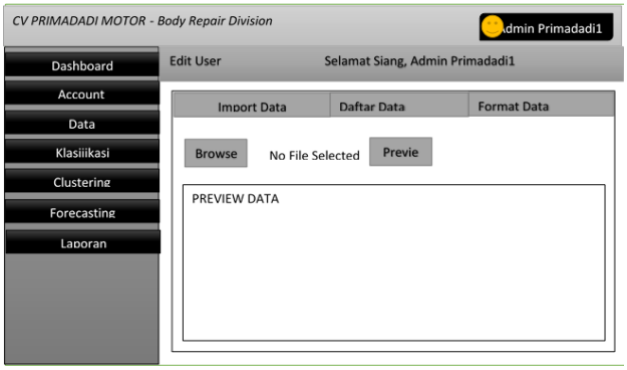

Fig. 8. Page of Import Data Design

IV. Result And Disccussion

A. Data Description

This research used 5,129 questionnaire data and vehicle repair transaction data at CV. Primadadi Motor, namely from the period January 2013 to December 2016. Questionnaire data for the period January to December 2013 will be used as training data to classify customer satisfaction data and transaction data from January to December 2017 will be used as validation of forecasting results.

Questionnaire data is used as material to create a decision tree for the C4.5 method to help divide transaction data from satisfied and dissatisfied customers. Furthermore, satisfied transaction data is used for the data clustering process. Transaction data is divided into 3 clusters based on the level of vehicle damage. The variables used to group transaction data are the number of panels and processing time.

Next, according to the group of vehicle damage levels, forecasting is carried out using the Holt Winter's multiplicative method which is adapted to the time series data plot. For example, data on the number of vehicle repair panels is symbolized by Xt where t is the period in months as in Figure 9.

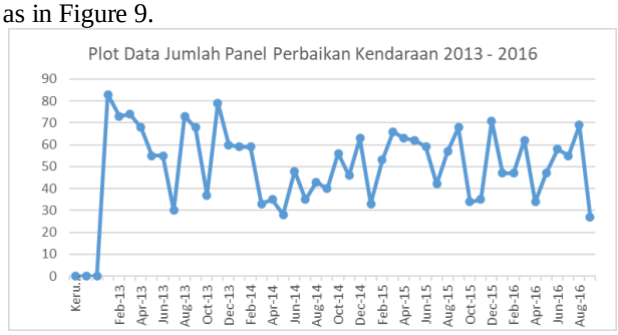

Fig. 9. Plot data on the number of vehicle repair panels

In Figure 10 it can be seen that the number of panels experienced a high increase in December and January, apart from that in the holiday months the number of vehicle repair panels continued to increase and was above average. For example, during Nyepi time which always occurs in March, the number of vehicle repair panels increases relatively much. Likewise with the months of Idul Fitri, Galungan and Kuningan. Figure 10 shows that the data is

influenced by patterns and trends as well as seasonal patterns because the time series plot shows fluctuations in the number of panels which increase and decrease in certain months each year. For example, the number of panels was stable in March, repeated the following year, and continued to increase. Data fluctuations tend to be unstable because they are influenced by seasonal factors.

B. Classification Method C4.5

Training data is used for learning in the data mining process or decision tree formation process. In this research, the training data used was 2013 customer questionnaire data, totaling 1090 data. In Table I below is a table of total training data results.

TABLE I. TOTAL OF DATA TRAINING

Total Case	Amount (Satisfied)	Amount (Unsatisfied)
1182	1090	92

The attributes used in determining customer satisfaction are Service, Price, Facilities, Results, Recommendations, Loyalty and Speed. The following are the stages of the process of implementing the C4.5 classification method to divide customers into 2 classes, namely satisfied and dissatisfied customers.

The first process in classifying data using Entropy is determining the total case Entropy value. The formula for finding Entropy is Formula 2. The following is the calculation of the Entropy value for the total cases.

$$Entropy\ (S) = \left(-\left(\frac{Sum\ (Puas)}{Total}\right) x\ \log_2\left(\frac{Sum\ (Puas)}{Total}\right)\right) + \left(-\left(\frac{Sum\ (Tidak\ Puas)}{Total}\right) x\ \log_2\left(\frac{Sum\ (Tidak\ Puas)}{Total}\right)\right) \quad (1)$$

$$Entropy\ (1090, 92) = \left(-\left(\frac{1090}{1182}\right) x\ \log_2\left(\frac{1090}{1182}\right)\right) + \left(-\left(\frac{92}{1182}\right) x\ \log_2\left(\frac{92}{1182}\right)\right)$$

$$= 0.3945$$

TABLE II. ENTROPY CALCULATION

Total Case	Amount (Satisfied)	Amount (Unsatisfied)	*Entropy*
1182	1090	92	0.3945

After the Entropy value for the entire data is known, the next step is to calculate the Gain value for the Result attribute using the following formula.

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy\ (S_i) \quad (2)$$

Where S is a set of cases, A is amount of fitur, n is amount of partision of A, $|S_i|$ is proportion of S_i to S, and $|S|$ is the number of cases in S. So the following results are obtained.

$$Gain\ (S, A) = Entropy\ (total) - \left(\left(\frac{sum\ (total)}{jumlah\ h\ kasus}\right)\right) \quad (3)$$

$$Gain\ (S, A) = 0.3945 - \left(\left(\frac{1090}{1182}\right) x\ 0\ + \left(\frac{92}{1182}\right) x\ 0\ \right)$$

$$Gain\ (S, A) = 0.3945$$

The calculation stops here because the Entropy value is 0, it can be concluded that the customer is satisfied if the workmanship is good. Because the Gain value of the result attribute is the same as the Entropy value, the calculation process is stopped so that a decision tree is created. Figure 10 is a decision tree formed from the data classification process using the C4.5 Algorithm.

TABLE III. CALCULATION OF GAIN RESULTS

Atribute	Atribute Value	Total Case	Amount (Satisfied)	Amount (Unsatisfied)	*Entropy*	*Gain*
Total		1182	1090	92	0.3945	
Results	Good	1090	1090	0	0	0.3945
	Bad	92	0	92	0	0.3945

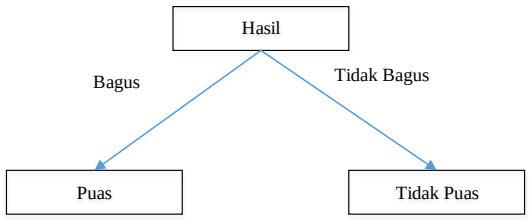

Fig. 10. Plot data on the number of vehicle repair panels

C. K-Means Clustering Method

The input data used is 4,854 customer transaction data with satisfied status. In carrying out clustering with K-Means, the object data used is the number of panels and the processing time from vehicle repair transaction data. In Table IV is the clustering process object data in the form of 2-dimensional data.

TABLE IV. DATA OBJEK

No.	No. Transaction	Month, Year	Work duration, Number of Panels
1	P130001	1,2013	9,2
2	P130003	1,2013	3,3
3	P130004	1,2013	9,5
4	P130005	1,2013	9,4
5	P130006	1,2013	8,4
6	P130007	1,2013	50,17
…	…	…	…
1151	P140071	1,2014	14,5
1152	P140072	1,2014	11,3
1153	P140073	1,2014	11,9
1154	P140074	1,2014	15,13
…	…	…	…
4850	P161341	12,2016	36,15
4851	P161342	12,2016	20,9
4852	P161343	12,2016	15,3
4853	P161344	12,2016	14,9
4854	P161345	12,2016	35,20

The object data is then divided into 3 groups with centroids as in Table V.

TABLE V. INITIAL CENTROID

Centroid	Work duration, Number of P
1	3.4
2	10.7
3	15,13

Object data is grouped according to the initial centroid data with 2 iterations as in Table V. The final centroids of this clustering process are Cluster 1 (4.75, 2.65), Cluster 2 (11.42, 4.84), and Cluster 3 (29.11, 12.045).

TABLE VI. CLUSTERING ITERATION


ITERASI 1

Objek	Iterasi	Data 1	Data 2	Cluster 1	Cluster 2	Cluster 3
P130001	1	9	2	Null	OK	Null
P130003	1	3	3	OK	Null	Null
P130004	1	9	5	Null	OK	Null
P130005	1	9	4	Null	OK	Null
P130006	1	8	4	Null	OK	Null
P130007	1	50	17	Null	Null	OK
P130008	1	11	4	Null	OK	Null
P130010	1	5	3	OK	Null	Null
P130011	1	9	2	Null	OK	Null
P130012	1	17	6	Null	OK	Null
P130013	1	8	4	Null	OK	Null
P130014	1	30	15	Null	Null	OK
P130015	1	3	2	OK	Null	Null
P130016	1	9	6	Null	OK	Null
P130017	1	50	15	Null	Null	OK
P130019	1	17	7	Null	Null	OK

Cluster 1 -> 4.7509778357236 2.6544980443286 Cluster 2 -> 11.423309178744 4.842391304347 8 Cluster 3 -> 29.114767585356 12.045248868778

From the process of implementing the K-Means algorithm on 4,854 data, the results showed that 2,492 transaction data fell into cluster 1, which means it had a light level of damage, 609 transaction data had a moderate level of damage, and 1,753 data had a heavy level of damage. The data from the clustering results will then be used to forecast the number of units, number of panels, and total company revenue based on the level of vehicle damage, namely light, medium, and heavy.

D. Holt Winters Method

The exponential smoothing used in this research is (Multiplicative Seasonal Method) with the initial values of the calculation process being as follows.

$$S_1 = \frac{83}{(83+73+74+68+55+55+30+73+68+37+79+60)} = \frac{83}{755} = 0.1099$$

$$T_1 = \frac{1}{12}\left(\frac{(59-83)+(59-73)+(33-74)+(35-68)+(28-55)+(48-55)+(35-30)+(43-73)+(40-68)+(56-37)+(46-79)+(63-60)}{12}\right)$$

$$T_1 = \left(\frac{(-24)+(-14)+(-41)+(-33)+(-27)+(-7)+(5)+(-30)+(-28)+19+(-33)+3}{12}\right) = -17.5$$

$$N_1 = \frac{83}{0.1099} = 755.23$$

After obtaining the initial value, it is then substituted into Formula to calculate the value of the forecast results. Below is an example of a forecasting calculation using the Holt Winters Multiplicative method.

$$N_2 = 0.5\ \frac{73}{0.0967} + (1-0.5)\big(755.23+(-17.5)\big) = 746.32$$

$$T_2 = 0.3\ (746.32-755.23) + (1-0.3)(-17.5) = -14.92$$

$$S_2 = 0.2\ \frac{73}{755.23} + (1-0.2)(0.0967) = 0.01933 + 0.07736 = 0.0967$$

Thus, the forecast value for the next period with m=1 is obtained as follows. And the forecasting process will be repeated continuously over a certain period of time.

$$F_{t+m} = (746.32+(-14.92))0.0967 = 70.73$$

E. System Implementation Results

The transaction data analysis information system is created in the form of a website where after logging in to the system the first page that can be seen is the Dashboard as in Figure 11. This homepage is the default and functions as the main page. The main page provides information about the forecasting stages and a glimpse of the CV. Primadadi Motor.

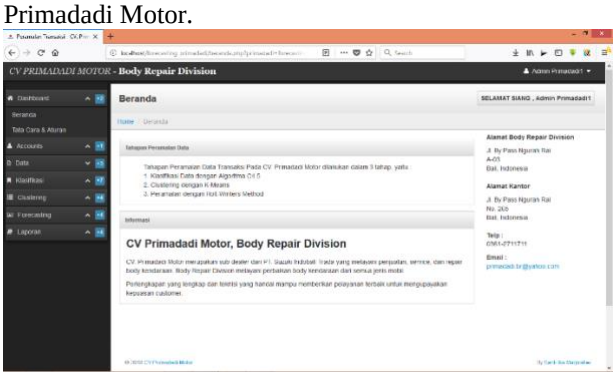

Fig. 11. Homepage

The Data menu has 3 sub menus, namely Questionnaire Data, Vehicle Data, and Transaction Data which are displayed as in Figure 12. Next, the data that has been imported is carried out in a classification process on the Classification menu page which has 7 (seven) sub menus consisting of Training Data, Attribute Settings, Attribute Value Settings, Classification Process, Decision Tree, Data Classification, and Classification Results. The output results of data processing using the C4.5 algorithm are shown in Figure 13 and Figure 14.

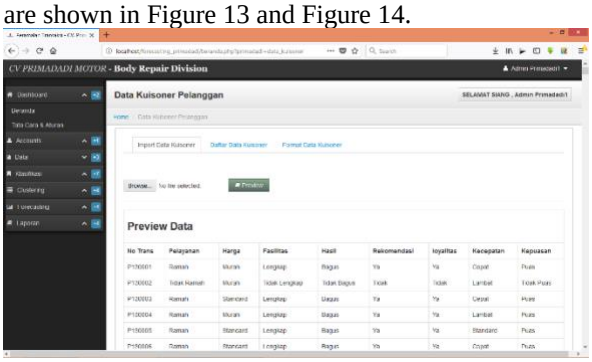

Fig. 12. Kuesioner Data Import Page

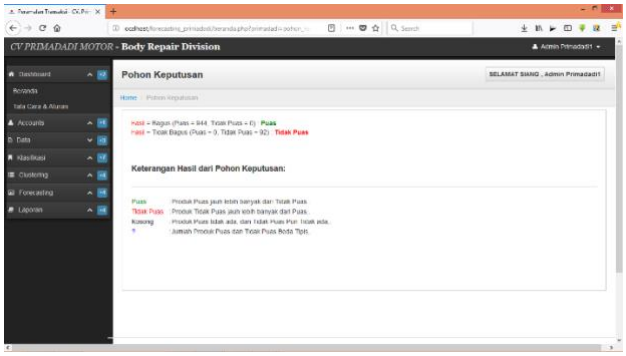

Fig. 13. Decision tree page of C4.5 algorithm results

Based on the classification results, the data is then processed in the clustering menu which provides 4 sub menus which can be used to process data clustering using the K-Means method. There are 4 (four) sub menus in the clustering menu, namely Object Data, Centroid Settings, Data Clustering, and Clustering Results. Figure 15 is the

input for the clustering process and the output is shown in Figure 16.

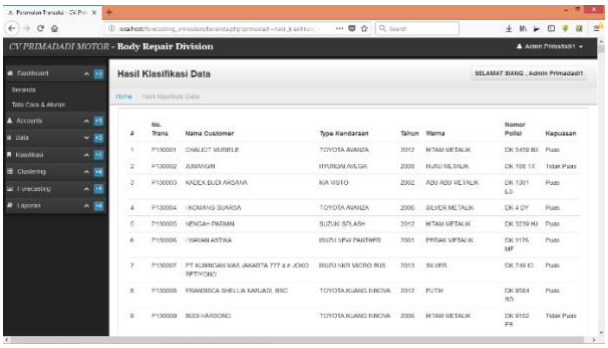

Fig. 14. Clasification Results Page

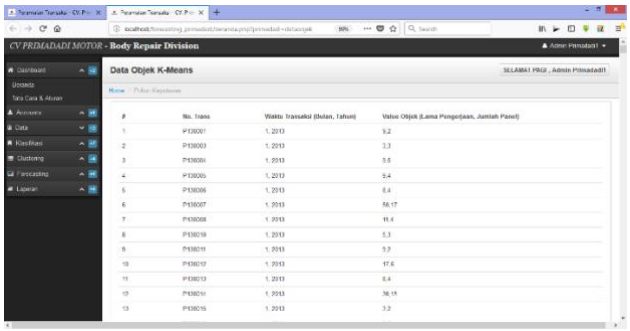

Fig. 15. Clustering Input Page

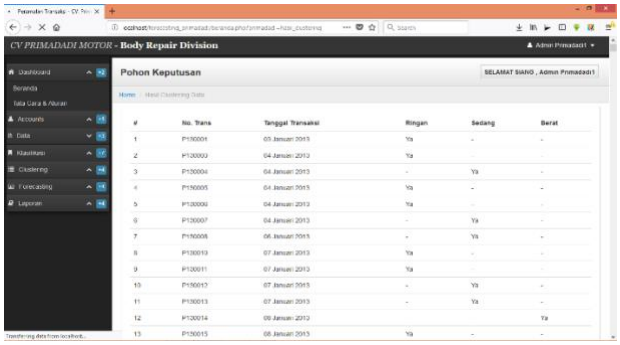

Fig. 16. Output of clustering Page

After obtaining the vehicle clustering results, proceed to the forecasting menu. Forecasting in this research was carried out on three variables, namely the number of panels, the number of units, and the amount of income. Each data will be predicted based on the level of damage and the total amount of data. The input data used for the forecasting process comes from transaction data resulting from the clustering process. In the forecasting menu there are 4 sub menus, namely forecasting data, forecasting variable settings, data forecasting, and forecasting graphs. Forecasting data displays data that will be processed using the Holt Winters method. The forecasting panel data page is shown in Figure 17.

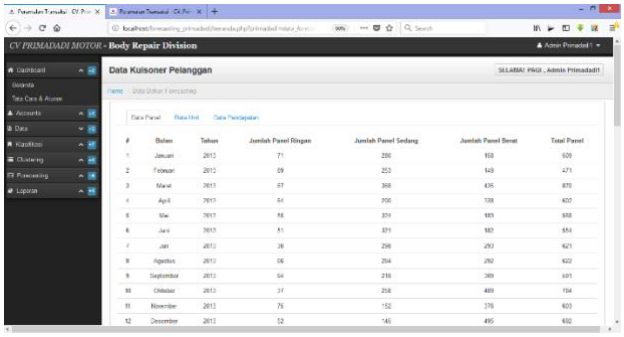

Fig. 17. Panel Actual Data Page

The Holt Winters method has 3 calculation constants, namely level, trend and seasonal. When forecasting data with Holt Winters, constants must be determined first. The forecast constant settings menu is provided to assist users in setting forecast constants. In Figure 18 is the display of the forecasting constant settings page and in Figure 19 is the output of the forecasting results for the number of panels.

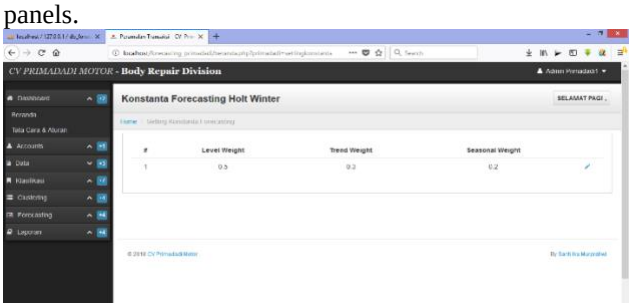

Fig. 18. Page for Setting Constanta of Holt Winters Method

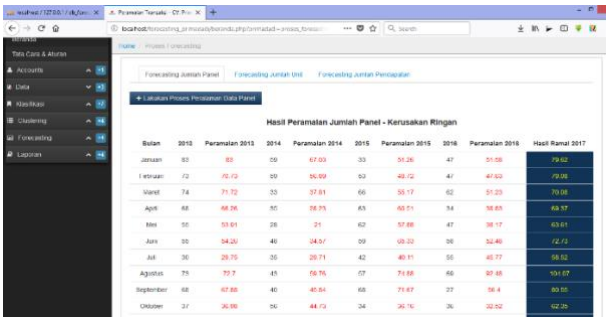

Fig. 19. Result Page number of prediction transactions

To help and simplify the data analysis process, a graph was created from the forecasting results for the number of panels, data for the number of units, and data for the amount of income. In Figure 20 below is a graph of the results of forecasting the number of panels using the Holt

Winters method.

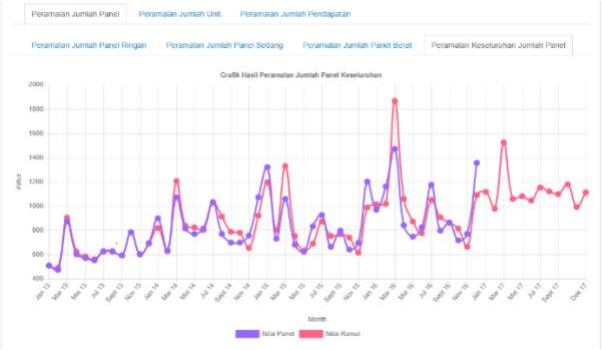

Fig. 20. Graph page Comparison of actual data and forecasting results

In this information system, a report menu is also created to assist the process of reporting transaction data by users. The report menu is equipped with date filtering to help select data within a certain time period. There are 4 (four) reports provided, namely the Transaction Data Report, Satisfied Customer Data Report, Vehicle Damage Level Report, and Transaction Data Forecasting Results Report. In Figure 22 the report output results.

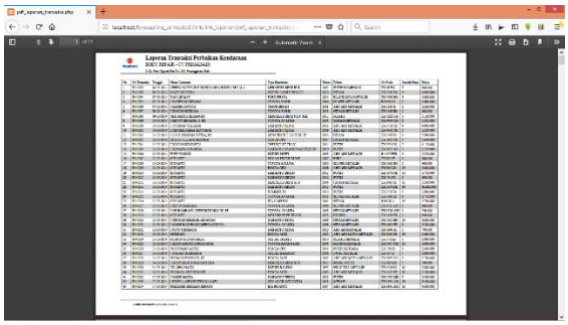

Fig. 21. Transaction Data Report Print Page

F. System Testing

System testing is carried out by carrying out black box testing and system computing time testing with the following details.

1. System time testing has been carried out on each algorithm in processing transaction data and the results are obtained in Table VII .

TABLE VII. SYSTEM COMPUTING TIME TESTING RESULTS

No	Process	Trial				
		I	II	III	IV	V
1	Import Data	1 minute 52 s	2 minute 8 s	2 minute 16 s	2 minute 26 s	3 minute 3 s
2	Classification	2 minute 12 s	2 minute 25 s	2 minute 58 s	3 minute 18 s	3 minute 27 s
3	*Clustering*	28 minute 12 s	29 minute 3 s	29 minute 19 s	30 minute 46 s	32 minute 31 s
4	Forecasting	1 minute 53 s	1 minute 53 s	1 minute 53 s	1 minute 53 s	1 minute 53 s
Total Time		34 minute 9 s	35 minute 29 s	36 minute 26 s	38 minute 23 s	40 minute 54 s

From the system time testing results, it can be seen that the entire data processing process in this research took more than 30 minutes. With a total of 5,109 data, the processed transaction data took a total of 40 minutes 54 seconds. The longest time required is the data clustering process. This is because the clustering process using K-Means requires several iterations until the data is clustered according to the level of damage. Apart from that, the clustering process with two-dimensional data, namely the processing time and number of panels, requires a longer processing time because the clustering process is carried out on two objects at once. So it can be concluded that the amount of data and diversity of data in the clustering process affects the time required,

Meanwhile, for data forecasting, the time required is relatively stable, namely 1 minute 53 s for all amounts of data. This is because in data forecasting, transaction data is grouped based on the month and year in which the transaction took place. So the transaction data is divided into 48 months for the number of panels, number of units, and, amount of income which is then carried out a data forecasting process.

2. Blackbox testing is carried out by testing the system functionality and is shown in Table VIII below.

TABLE VIII. BLACKBOX SYSTEM TESTING RESULTS

No	Component	Scenario	Result
1	Administrator authentication	A user has an admin username and an admin password	√
2.	Update administrator password	A user changes the password, before entering a new password	√
3.	Import Questionnaire Data	*User* can import quesionnaire data from excel file	√
4.	Import Vehicle Data	User can import vehicle data from excel file	√
5.	Import Transaction Data	User can import transaction data from excel file	√
6.	Import Transaction Data	User can reply import transaction data from excel file	√
7.	Import Training Data	User can import training data from excel file	√
8.	Import Training Data	User can reply import training data from excel file	√
9.	Insert Data Atribute	User can insert data atribute to system	√
10.	Delete Data Atribute	User can delete data atribute to system	√
11.	Edit Value Atribute Data	User can modify value of data atribute to system	√
12.	Delete Value Atribute Data	User can delete value of atribut data	√
13.	Clasification data training	User can classification data training in the system	√
14.	View Decision Tree	User can view decision tree as result of classification data	√
15.	Process of Data Classification	System can process data classification	√
16.	Insert Data Object	Insert data object with work duration and number of panel	√
17.	Insert Data Centroid	User can processing input data centroid	√

No	Component	Scenario	Result
18.	Delete data centroid	User can delete data centroid	√
19.	Process of *clustering* data	System can process data clustering	√
20.	*Show results of clustering* data transaction	System can show the result of clustering data	√
21.	Insert data setting variable Faorecasting	User can input variable forecasting (alpha, betha, gamma)	√
22.	Edit variabel forecasting	User can modify value of variable forecasting	√
23.	Forcasting Data Unit	System can process forecasting data units	√
24.	Forecasting data panel	System can process forecasting data panel	√
25.	Forcasting Income data	System can process forecasting data income	√
26.	Export report to .pdf	System can save report as pdf file	√
27.	Displays Customer Satisfaction Level Reports within a certain period of time	User can setting certain period of time in the customer satisfaction reports	√
28.	Displays a vehicle damage level report	User can setting certain period of time in the customer vehicle damage level report	√
29.	Logout of the sistem	User can exit from the forecasting system	√

V. Conclusion And Suggestion

The system for analyzing and forecasting clustering company transaction data begins with designing and building the system until it is implemented and tested. Based on the discussion of the results achieved, the following conclusions can be drawn:

1. The clustering company transaction data analysis and forecasting system was successfully created and implemented using a combination of the C4.5 Algorithm classification, K-Means clustering, and Holt Winters forecasting with real company data totaling 5,109 transaction data.
2. Implementation of classification and clustering provides good accuracy values, where the classification accuracy with the C4.5 Algorithm is 100% and the clustering accuracy with K-Means is 93.28%.
3. Forecasting transaction data based on the level of vehicle damage was implemented using Holt Winter Multiplicative and produced data forecasting results with the majority of MAPE values below 20% with an average for forecasting the number of panels of 10.82%, forecasting the number of units of 9.59%, and forecasting revenue of 10.82%. 11.46%. The unit forecasting MAPE value has the lowest value because the number of units entering the company each month is limited due to limited space so that the number of units entering is more regular and stable in number, resulting in a higher accuracy value. So, the forecasting system built already has good accuracy.

The suggestions for developing a system for analyzing and forecasting company business transaction data are as follows.

1. The diversity of questionnaire data should be added to produce more complex rules.
2. Improved appearance for reports so that they can be exported in various file types.
3. Experiment in using Holt Winters forecasting constants in order to find forecasting constants that are suitable for company transaction data so as to produce higher forecasting accuracy values

References

[1] Lakshmi, B. N., Indumathi, T. S. and Ravi, N. (2015) 'A Comparative Study of Classification Algorithms for Predicting Gestational Risks in Pregnant Women', in *International Conference on Computers, Communications, and Systems*, pp. 42–46.

[2] Venkatkumar, I. A. and Shardaben, S. J. K. (2016) 'Comparative study of Data Mining Clustering Algorithms', in *2016 International Conference on Data Science and Engineering (ICDSE)*. Cochin, India: IEEE. doi: 10.1109/ICDSE.2016.7823946.

[3] Muniyandi, A. P., Rajeswari, R. and Rajaram, R. (2012) 'Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree Algorithm', *Procedia Engineering*, 30(2011), pp. 174–182. doi: 10.1016/j.proeng.2012.01.849.

[4] Fadlil, J. (2007) 'Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering', *Kursor*, 3(1), pp. 45–66.

[5] Fang, D., Zhang, Y., Spicher, K. and Revinda, A. (2011) 'Forecasting Accuracy Analysis based on Two New Heuristic Methods and Holt-', in *2016 IEEE International Conference on Big Data Analysis (ICBDA)*. Hangzhou, China: IEEE. doi: 10.1109/ICBDA.2016.7509814.

[6] Tratar, F. L. and Strmcnik, E. (2016) 'The Comparison of Holt Winters method and Multiple regression method: A case study', *Emergy*, 109, pp. 266–276. doi: 10.1016/j.energy.2016.04.115.

[7] Turban, E., dkk, 2005, Decicion Support Systems and Intelligent Systems, Andi Offset

[8] Larose, D. T. (2005) *Data Mining Methods And Models*, *Spring*. doi: 10.1177/0002716209351510.

[9] Hettiarachchi, N. D. (2015) 'I Shopping : Intelligent Shopping and Predicate Analysis System Using Data Mining', pp. 1–8. doi: 10.1109/DSDIS.2015.75.

[10] Ahlawat, R., Sahay, S., Sabitha, S. and Bansal, A. (2016) 'Analysis of Factors Affecting Enrollment Pattern in Indian Universities Using K-Means Clustering', pp. 3–8.

[11] Taufiq, E. L. (2016) 'Classification Method of Multi-class on C4 . 5 Algorithm for Fish Diseases', pp. 5–9.