Forecasting Oil Sales Volume with Data Visualization in Sinar Servindo Sakti Company for 1 Year with Knowledge Discovery of Database (KDD) Method

Friska Natalia[1], P.M. Winarno[2], Merfin[3]
{friska.natalia@umn.ac.id[1], pmwinarno@umn.ac.id[2], merfin@student.umn.ac.id[3]}

Faculty of Business, Universitas Multimedia Nusantara, Tangerang, Indonesia[1,2,3]

Abstract. This research is conducted as quantitative research discusses the concept of industry 4.0 by promoting the era of digital technology. This Research is centered on the development of Forecasting in Sinar Servindo company, an oil retail company that purchases orders from their partner companies. The main result is that the company's sales volume predicted to increase quite rapidly and provide excellent profits for the company in the next 1 year with Rapid Miner and Power BI.

Keywords: Forecasting, Rapid Miner, K-Medoids, DBScan, X-Means

1 Introduction

In industry 4.0, there is a rapid development of technology with massive changes giving birth to a new generation, namely Generation Y. There are several main aspects in this era as a sign of the times such as the Internet of Things, Artificial Intelligence, Big Data, Cloud Computing, and so on[1].

In a company, the stock of goods used does not experience many problems, but often this becomes a problem because these raw materials are already scarce. Considering that an item in the company's inventory will incur a large cost, the producer must prepare a stock of oil goods with a capacity of 2 times the previous one. If this is successful, then sales will increase and finances will recover [1].

From this background, the problem statement taken as the main material for this thesis research is the main factors of the influence of sales volume from the price of an oil distribution company in South Tangerang named Sinar Servindo company. This Topic is important because the inventory cost is high and need support to make the balance with stock goods.

The research method to be researched and worked on is quantitative research method with visualization data result of sales price prediction using Rapid Miner and Power BI. Thus, the result of this research is a dashboard visualization of data marts that have been predicted from the company's data in the field of marketing that is linked to sales data at Sinar Servindo company to determine forecasting planning for 3 years (August 2017 – August 2020) along with nominal forecasting and inventory of goods for the next 1 years.

[1] Rizkinaswara, L. (2020). *Revolusi Industri 4.0*. Infrastruktur TIK. https://aptika.kominfo.go.id/2020/01/revolusi-industri-4-0/

ICEBE 2021, October 07, Lampung, Indonesia

DOI 10.4108/eai.7-10-2021.2316143

After knowing the background gap for the selection of the topic, the important problem formulations are how the influence of the main factors of the company's price sales volume based on the application of the *K-medoids, DBScan, and X-means* models in the creation of data visualization of prediction results conducted in the company Sinar Servindo company.

The purpose of this research is to help the process of calculating the predicted sales volume of Sinar Servindo Sakti company for 3 years. The main things that have good benefits to the development of this thesis research is to determine the influence of sales volume prediction process in Sinar Servindo Sakti company based on oil distribution market trends.

.

2 Literature Review

2.1. THEORY REVIEW

2.1.1. Data Mining

Data mining is one of the main methods to perform extract data patterns and detect several parameter relationships in the data. The main purpose of using Educational Data Mining is to develop the model so that it can determine the performance of each available variable. [2]

2.1.2. Data Visualization

A working method for making presentation dashboards that are made as attractive as possible to attract customer attention based on the results of mathematical calculations predicting a problem in the vicinity. [3]

2.1.3. Prediction

The method of work using survey research that occurs around us with the aim of calculating a number that can predict a problem recorded in valid data. [4]

2.1.4. K-Medoids

The difference between the K-Medoids model and the K-Means model is that the K-Medoids model uses representative objects (medoids) as the center of clustering in a research study, while the K-Means model only uses the mean or average value as the center of the process. [5]

2.1.5. DBScan

One of the clustering models is used to cluster a particular problem with irregular data. Research using this model is more efficient than *Clarans* because this model can detect noise or outliers widely, does not need input as in the K-Medoids model, and can recognize some irregular sample data. [6]

2.1.6. X-means

This method enhances the limitations or shortcomings of this K-Means clustering research. This x-means model algorithm will recursively add the number of existing clusters so that the resulting number is the same as the user value above. [7]

2.1.7. Knowledge Discovery of Database

A method where all the knowledge of a data related to an organization or institution that has a very large amount of data and focuses on a broad methodology. An example of application is in the calculation of the FP-Growth Algorithm in an existing pattern. [8]

2.2. CONCEPTUAL FRAMEWORK

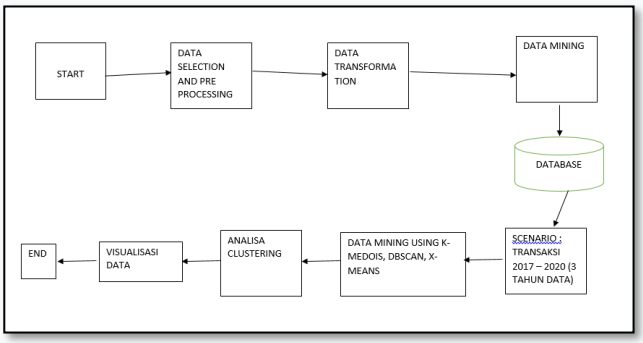

Fig.1. Conceptual Framework Knowledge Discovery Of Database

Figure 1 based on the main journal which is the main reference for determining the factors that influence the volume of sales in the company Sinar Servindo Sakti company, which is the calculation of nominal predictions and inventory of sales for 1 year with the category of hydraulic oil, engine oil, compressor oil and gearbox oil.

3. Methodology and Data Analysis

2.1. RESEARCH PARADIGM

The quantitative research work that will be used is the classification of models from three models at once, namely *K-medoids, DBScan, and X-means*. This research has a positivist paradigm because the thesis research is carried out using a quantitative research approach.

2.2. OBJECT OF RESEARCH

The research object of this research is the influencing factors of the sales volume of company prices and will be examined by means of the forecasting formulation of the *K-medoids, DBScan, and X-means* models using predetermined tools, namely Rapid Miner and Power BI.

2.3. POPULATION AND SAMPLE

The sample and population data, which is about 993 lines of sample data, has been recorded and divided into 3 transaction periods totaling 350 rows, 330 lines, and 312 rows respectively.

2.4. OPERATIONAL VARIABLE

The variable to be studied is the volume of sales prices made in the Sinar Servindo Sakti company for 3 years (August 2017 to August 2020) based on a one-year record of sales transactions that have been recorded properly and are valid and divided into two main sub-variables, inventory and sales.

2.5. DATA COLLECTION TECHNIQUE

The data collection technique in this thesis is the method of observation, which is looking at the raw data to classify the model to find out the main factors of the sales volume of the company price.

2.6. DATA PROCESSING TECHNIQUES

These data will be formed into fresh data through data cleansing. All data that has been cleansed and tidied will be classified using three models, namely *K-medoids, DBScan, and X-means*.

2.7. DATA ANALYSIS TECHNIQUE

The research method that will be used is KDD (Knowledge Discovery in Database) to work on and research using Rapid Miner and Power BI tools to determine the prediction of sales volume of company with three models (*K-medoids, DBScan, and X-means).*

4. Research Result and Discussion

2.1. CHARACTERISTIC

The customer who includes in this research will be the company and individual company. All respondents selected come from individuals who join companies and companies that are partners of Sinar Servindo company. The Forecasting system is not optimal. The application cannot be done without the Acceptance Test because it's still prototype.

2.2. RESULT

The results of the analysis for forecasting research are carried out based on the implementation of the three models and will be compared to make final conclusions and the implications of the analysis results using data visualization.

2.2.1.1. K-Medoids

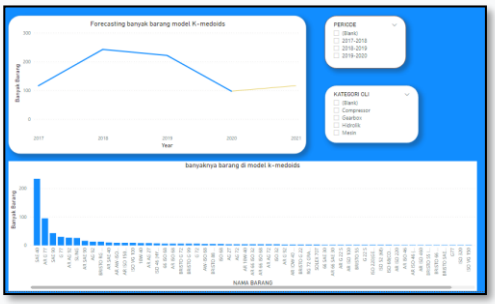

Fig.1. Visualization forecasting K-medoids dashboard for inventory items

The stock of these items is at 116 stock items. This shows that there was a recovery in the stock of goods which had almost doubled from the previous period.

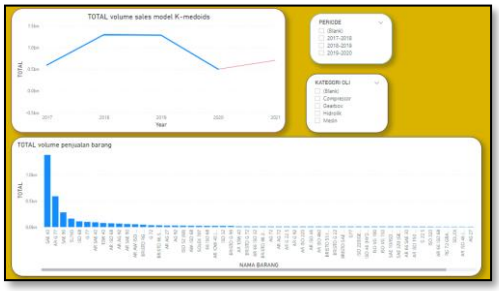

Fig.2. Visualization forecasting K-medoids dashboard for sales items

Sales volume of the Sinar Servindo Sakti company is predicted to increase to IDR 706,548,259 at the end of the next period (August 2020 - July 2021). DBScan

Fig.3. Visualization forecasting *DBSCan* dashboard for inventory items

The forecasting process for the next period shows an improvement in the stock of goods with 130 items predicted to be stored in the company's inventory of Sinar Servindo Sakti.

Fig.4. Visualization forecasting *DBSCan* dashboard for sales items

The company's sales volume is predicted to increase in the next period (August 2020 - July 2021) with the resulting projected value of IDR 816,963,220.

2.2.1.2. X-Means

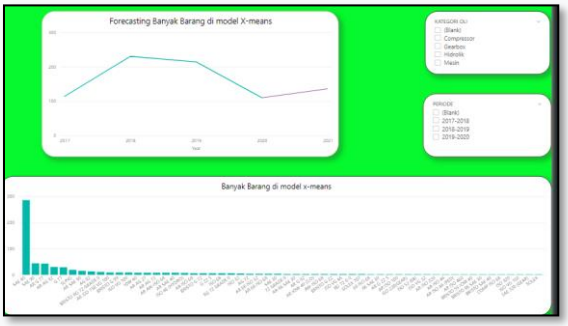

Fig.5. Visualization forecasting X-means dashboard for inventory items

The end of the new period (August 2020 - July 2021) predicts an increase in the stock of goods, which will be around 136 stocks.

Fig.6. Visualization forecasting X-means dashboard for sales items

At the end of the next period (August 2020 - July 2021) it predicts an increase in the company's sales volume figure at IDR 843,880,990.

2.3. DISCUSSION

The main result of this research is that the company's sales volume is predicted to increase quite rapidly and will provide a very good profit for the company in the next 1 year. The main problem in the research that has been made is that the company data has not been audited by the sales company until now. Further research focused on the completeness data of PT. Sinar Servindo Sakti to find out the main factors of the inventory cost of these goods to the customer.

5. Implication and Suggestion for Future Research

2.1. IMPLICATION

The three models in the same KDD method, namely *K-medoids, DBScan, and X-means* both predict that if all stock of goods increases, sales volume will also increase quite rapidly.

2.2. SUGGESTION

Three Suggestion of this research is to check the inventory stock of goods in the company with the aim of reducing the costs that have been budgeted by the company.

Sales companies need to pay attention to and learn about the company's forecasting of Sinar Servindo Sakti. It is better to training for development of the existing forecasting system to improve the quality of sales of goods in the next few years.

It is necessary to do further research on the characteristics of the category of oil goods that are inherent in society with existing data and a broader and more modern forecasting system.

6. Reference

[1] P. Muliady, P. Utomo, and F. Natalia, "Development of Cloud Based Demand Forecasting System : A Case Study in Stationery Industry," in *International Conference of Economics, Business & Entrepreneurship*, 2021, doi: 10.4108/eai.1-10-2020.2305004.

[2] Križanić, S. (2020). Educational data mining using cluster analysis and decision tree technique: A case study. International Journal of Engineering Business Management, 12, 1–9. https://doi.org/10.1177/1847979020908675.

[3] Jagadeesan, K. K., & Ekström, S. (2017). MALDIViz: A Comprehensive Informatics Tool for MALDI-MS Data Visualization and Analysis. SLAS Discovery, 22(10), 1246–1252. https://doi.org/10.1177/2472555217727517.

[4] Skeels, C. L., & Taylor, L. W. (2015). Prediction in linear index models with endogenous regressors. Stata Journal, 15(3), 627–644. https://doi.org/10.1177/1536867x1501500302.

[5] Pramesti, D. F., Lahan, Tanzil Furqon, M., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 1(9), 723–732. https://doi.org/10.1109/EUMC.2008.4751704.

[6] Sari, B. N., & Primajaya, A. (2019). Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang. Jurnal Informatika Dan Komputer, 4(1), 28–34. www.mapcoordinates.net/en.

[7] Wijayanto, A. (2019). Use of the X-Means Clustering Method to Group Potential High Schools in Banyumas Regency. Journal of Informatics, Information Systems, Software Engineering and Applications (INISTA), 2 (1), 80–88. https://doi.org/10.20895/inista.v2i1.99.

[8] Informatika, P. S. T., & Majapahit, U. I. (2019). Pemanfaatan Knowledge Data Discovery (KDD) Pada Pola Permainan Atlet Bulutangkis. Explore IT : Jurnal Keilmuan Dan Aplikasi Teknik Informatika, 11(1), 1–6. https://doi.org/10.35891/explorit.v11i1.1467.