

Data Visualization and Sales Prediction of PD. Asia Agung (Ajinomoto) Pontianak in 2019

Willy Yunus¹, Ririn Ikana Desanti², Wella³

^{1,2,3} Information Systems Department, Universitas Multimedia Nusantara, Tangerang, Indonesia

¹ willy.yunus@student.umn.ac.id

² ririn.desanti@umn.ac.id

³ wella@umn.ac.id

Accepted on 09 October 2019

Approved on 18 June 2020

Abstract—PD. Asia Agung Pontianak is the only official distributor of Ajinomoto in the West Kalimantan region. Every year this company needs to find out the amount of turnover that will be obtained in the coming year. Unfortunately, the company only makes predictions using the average income from each year which is very less accurate. This research is conducted to create visualizations and predictions using multiple linear regression methods to predict the turnover obtained in the coming year. Multiple linear regression is a regression analysis method that can use more than 2 variables in the prediction process which is divided into 2 parts, namely the dependent variable and the independent variable. The results obtained in this research are prediction results in 2019 using data from 2010 to 2018 as a basis. Prediction results show that the longer the data used the smaller the error rate obtained. The original data from the company is visualized using a dashboard on tableau software so that the data could be easier to analyze by the company.

Index Terms—data visualization, multiple linear regressions, prediction, tableau

I. INTRODUCTION

PD. Asia Agung is the only official sole distributor that sells products from AJINOMOTO in the West Kalimantan region [1]. To find out the history of the company, interviews were conducted with the company. Based on the result of interview with one of the admin and finance heads, it is known that the company had financial information about company turnover data for 2010-2018. This company has been established since the 1960s and started marketing AJINOMOTO's official products in the 1970s. PD. Asia Agung is a family business and now the owner has entered the 3rd generation, under the leadership of Mr. Wiyan as the son of Mr. Agus Amin. Under the leadership of Mr. Wiyan, the company is rapidly growing which makes company's revenue increase quite rapidly too. Mr. Wiyan also wanted to find out the turnover for each year, where the admin and finance section made predictions based on the average

increase in total turnover each year but the company felt it was less accurate.

Therefore, research will be conducted in the form of visualization and prediction of sales turnover data from PD. Asia Agung Pontianak. Based on the data obtained from the company, this study will predict using the Multiple Linear Regression method which is usually used in prediction or forecasting which is prepared on the basis of the pattern of relevant data relationships in the past which is assisted by using SPSS tools in data processing [2]. In the regression method generally predictable variables such as the sale or demand for a product, expressed as the dependent variable, this variable is influenced by the magnitude of the independent variable [2]. In order to make data easier to analyze, this research will visualize using Tableau software.

II. THEORETICAL BASIS

A. Definition of Prediction

Prediction is a process of systematically estimating something that is most likely to occur in the future based on past and present information that is owned, so that errors can be minimized [3].

Forecasting is a combination of art and science in estimating the situation in the future, by projecting the data of the past into the future using mathematical models and subjective estimates [4].

Based on the techniques used to predict, predictions can be divided into two parts, qualitative predictions and quantitative predictions [5]:

1. Qualitative Prediction

Qualitative predictions are based on qualitative data in the past. Qualitative methods are used if past data from variables that are predicted to be non-existent, not enough or less reliable. The results of predictions are very dependent on the

individual who composed them. This is important because the results of the prediction are determined based on judgment thoughts or opinions, knowledge and experience from the constituents. Therefore, this qualitative method is also called judgment, subjective, intuitive.

2. Quantitative Prediction

Quantitative predictions are based on quantitative data in the past. Predictions made are very dependent on the method used in the prediction. With different methods different prediction results will be obtained. The thing that needs to be considered from the use of these methods is whether or not the method used is very good and is determined from the deviation between the results of the prediction and the reality that occurs. A good method is a method that provides possible values of differences or deviations. Quantitative predictions can only be used if there are three conditions as follows: There is information about other circumstances, the information can be quantified in the form of data, it can be assumed that the previous pattern will be sustainable in the future.

B. Prediction Methodology

Multiple Linear Regression can be used in prediction or forecasting which is arranged on the basis of the pattern of relevant data relationships in the past. In regression methods generally predictable variables such as sales or demand for a product, expressed as dependent variables, this variable is influenced by the magnitude of the independent variable [2]. Multiple linear regression can be calculated based on:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i \quad (1)$$

\hat{Y} = Predict the value of dependent variable

β_0 = Constant (fixed value)

β_i = Regression weight (coefficient) for the I independent variable

X_i = I Independent variable

B.1 Decision Tree

Decision Tree is a tree structure, where each tree node represents the tested attribute, each branch is a distribution of test results, and leaf nodes represent certain class groups. The top level node of a Decision Tree is the root node (root) which is usually the attribute that has the greatest influence on a particular class. In general, Decision Tree performs a top-down search strategy for the solution. In the process of classifying unknown data, the attribute value will be

tested by tracking the path from the root node (root) to the end node (leaf) and then predicting the class that belongs to a certain new data [6].

B.2 Time Series

Time Series are events that occur based on certain times in sequence. Forecasting in time series is a prediction for the occurrence of events that will occur in the future based on previous data [7].

C. Definition of Visualization

Data visualization has three criteria, namely (1) the process is based on qualitative and quantitative data, (2) produces output derived from initial data, and (3) output can be read by use and supports the process of exploration, inspection, and data communication [8]. The main purpose of data visualization is to communicate information clearly and effectively in a graphical way. Not that the visualization of data must look boring so that it works or is very sophisticated in order to look attractive [9].

D. Visualization Methodology

Some methodologies that can be used to create dashboards developed by several vendors include:

1. Visualization of Data Mining

Visual Data Mining is a technique that can be done with repetitive processes [10]. Visualization of data mining has 3 stages, namely: [11]

a. Project Planning Phase

The stage will explain how visualization of this research can help in analyzing business, understanding of data and decision makers and explaining what tools to use.

b. Project Preparation Phase

This stage will explain how data were obtained, selected, processed, or changed. Data must be valid to be used.

c. Project Analyze Phase

The stage will explain how to choose the right tools based on the analysis that has been done and enter the visualization model analysis process.

2. Noetix

Noetix makes a methodology by trying to balance the needs of users with the capabilities of developers [12]. The stages in the noetix method [13]: a. Planning at this stage is the establishment of a project team and the determination of the parties who sponsor and are responsible for the project. b. Identify needs and design prototypes Identification of needs is done to determine the needs and desires of key stakeholders

on the dashboard to be built. c. The design complements the main aspects of the design needed such as smoothing the user interface, ensuring data sources, determining how to get data if data is not available in the database, defining queries, and determining drill-down paths. d. Implementation and validation of front end Implementation, which is the design of the form of a dashboard interface, such as graphs / charts, alerts, and navigation for drill-down paths. Query implementation, which is a query design to retrieve information from the database. Schedule configuration, update, and security. Queries are configured to be able to deliver information regularly, to ensure that the contents of information on the dashboard are always up-to-date. Security rules are also applied to ensure the security of the information presented. Dashboard validation is a testing process that is carried out to ensure that the dashboard produced is in accordance with the requirements specifications in the project plan. e. Deployment and Maintenance Implement the dashboard in its operational environment and subsequently the maintenance process to ensure that the dashboard functions remain in accordance with the needs of the organization.

3. Pureshare

This Pureshare methodology places an emphasis on the balance between business objectives and organizational technology needs as seen from the performance development metrics, the design of this system is done using a top-down and bottom-up approach [12].

The stages contained in the pureshare methodology [13]:

- a. Planning and design stage highlights

This stage is led by a business analyst, focused on designing a top-down approach to business and meeting end user needs, as a metric interface and to identify how these metrics are used.
- b. System and data review highlights

This stage is carried out by planning and designing a bottom-up implementation approach, namely by controlling the system and identifying data sources, by developing the quality of these metrics.
- c. Prototype stage highlights

At this stage, top-down and bottom-up approaches are carried out together. The prototype is built as quickly as possible to provide a final picture that resembles the dashboard system that will be created.
- d. Refinement stage highlights

At this stage, a series of prototypes that have been created will be reviewed together with users to gather feedback to be developed and adapted to the needs of users.

e. Release stage highlights

At this stage, the dashboard that has been implemented is socialized to all users through training. Continuous improvement is the key point. This stage is used to build dashboards in various organizational areas repeatedly.

f. Continuous improvement

At this stage the pure share methodology for proactive performance metrics has long been developed and has been disseminated to industries and companies, it is intended to obtain results directly and extend the life of metric management and project management performance of the project.

E. Prediction Tools

E.1 IBM SPSS Statistics

SPSS Statistics® is a leading statistical software that can be used to solve various business and research problems. SPSS provides a variety of techniques including ad-hoc analysis, hypothesis testing and reporting to make SPSS easier to manage data, select and do analysis, and share the results of calculations. New features include Bayesian statistics, graphs ready for publication and increased integration of third party software. SPSS Statistics offers a basic edition with optional add-ons to expand predictive analytic capabilities [14].

E.2 SAS Enterprise Miner

SAS Enterprise Miner is an integrated software system that makes it possible to create highly accurate predictive and descriptive models based on analysis of large amounts of data from all companies [15].

F. Visualization Tools

F.1 Microsoft Power BI

Power BI is one of the tools owned by Microsoft that can function to do Data mining, cleansing or visualization.

Power BI is a business analysis solution that makes it possible to visualize data and share your views across an organization, or put together in an application or website. And can be connected to hundreds of data sources and can create data in dashboards and reports directly [16].

F.2 Tableau Software

Tableau is the most powerful, safe and flexible end-to-end analysis platform for your data. Attract people with the power of data. Designed for individuals, but enhanced for companies, Tableau is the only business intelligence platform that turns your data into insights that drive action [17].

III. RESEARCH METHODOLOGY

A. Research Object

PD Asia Agung is a trading company engaged in selling food flavorings, PD Asia Agung is the sole distributor of products with the Ajinomoto brand in the West Kalimantan region. The form of the business is individuals. This individual business means the owner is responsible and fully controls the running of the business. In carrying out its business, PD Asia Agung has obtained a business license from the Regional Office of the Ministry of Trade of West Kalimantan with a Trade Business License (SIUP) No. 510.001/92/295/DPP.

B. Data Collection

Data collection is obtained directly from the Head of Admin and Finance of PD. Asia Agung (AJINOMOTO) Pontianak West Kalimantan. In detail the data obtained is as follows:

1. Sales revenue turnover of 5 products AJINOMOTO, MASAKO, Sajiku Tepung Bumbu, Sajiku Tepung Praktis, and SAORI per Regency or city on a monthly basis from 2010-2018.
2. Data Salesmen who work in each district in West Kalimantan.
3. Predictive target data from the company for each year starting from 2011.

C. Prediction Methodology

In this study we will use the Multiple Linear Regression method because the data is a series of data from previous years. Where data also has more than two interconnected variables. To predict the turnover value of PD. ASIA AGUNG (AJINOMOTO) Pontianak can use turnover variables and products which data will be included later using SPSS tools.

The variables used for the formula of multiple linear regression $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i$ are:

1. Total turnover as dependent variable
2. 5 independent variables: Ajinomoto, Masako, Saori, Sajiku Tepung Bumbu, Sajiku Tepung Praktis

D. Visualization Methodology

Data visualization aims to determine what information will be visualized in accordance with the problem statement. In this study the information to be visualized is related to sales data and turnover of each product of the Asia Agung Trading Company (AJINOMOTO) Pontianak, using Tableau Software that will be visualized with graphical forms using the Visualization Data Mining method because VDM has a complete work process, clearly, and structured so that it is easy to understand.

E. Prediction Tools

The SPSS tool will be used to make predictions, because it is already widely used in universities and has a good display for documentation and is easy to use for people who are not experts in statistics.

F. Visualization Tools

Visualization will use Tableau tools, because it is more accessible and simpler to use, but has complex results on the results of visualization. It will use a number of charts to help with the translation of data using line charts and bar charts.

G. Framework of Thinking

The mindset is the flow of the author's thoughts which is used as the rationale to strengthen the indicators underlying this research so that the problems raised in the formulation of the problem can be resolved. The mindset in this study is divided into 2 parts, namely:

1. Prediction

In the prediction framework, it is explained from where the data for predictions is obtained from where and in what form. After that, the data that has been obtained then the data is processed to simplify the process in SPSS then the results from SPSS can be used to predict by entering the results into the Multiple Linear Regression formula.

2. Visualization

The visualization framework is explained directly using the steps of Visualization Data Mining, namely starting from determining the type of project to be used, identifying problems, selecting data to be used for visualization, changing data so that it can be entered into tableau tools, ensuring data is correct, selecting tools that will be used, what will be displayed on the dashboard, and also display the results of the visualization.

IV. ANALYSIS AND RESEARCH RESULT

A. Data Collection for Prediction

PD Asia Agung (Pontianak) data was given by email after conducting an interview to the head of the admin and finance department, the data provided was in the form of excel data. The data contains sales turnover of 11 regions for 5 products, namely Ajinomoto, Masako, Sajiku Tepung Bumbu, Sajiku Tepung Praktis, and Saori in West Kalimantan from 2010-2018 and also predictive data from the company.

B. Change Data

The data is converted into data containing the total number of product turnover per month. Changing the data by adding up the income each month so that it will be easy to use later on SPSS.

C. Prediction

In this study, the level of accuracy was carried out first by comparing the prediction error rate using data every 1 year to 8 years. From research in looking for predictive error rates, Table I is obtained. Analysis of comparison is shown below:

TABLE I. ANALYSIS OF COMPARISON

| Lots of Data Used | Prediction Error Average Results (%) |
|-----------------------------------|--------------------------------------|
| Data every 1 year from 2010-2018 | 9,17% |
| Data every 2 years from 2010-2018 | 8,52% |
| Data every 3 years from 2010-2018 | 8,55% |
| Data every 4 years from 2010-2018 | 8,35% |
| Data every 5 years from 2010-2018 | 7,64% |
| Data every 6 years from 2010-2018 | 8,12% |
| Data every 7 years from 2010-2018 | 6,07% |
| Data every 8 years from 2010-2018 | 4,33% |

From the table above, can be concluded that if predicting using data with a long series of times can make the average error in predictions every year tend to be smaller or predictive, tend to be more accurate. Then from the results of the analysis obtained a comparison of the results of company predictions with the results of the analysis can be seen in Table II.

TABLE II. COMPARISON WITH COMPANY'S DATA

| Lots of Data Used | Prediction Error Average Results | Prediction Error Average Result from |
|-------------------|----------------------------------|--------------------------------------|
| | | |

| | (%) | PD. Asia Agung |
|-----------------------------------|-------|----------------|
| Data every 8 years from 2010-2018 | 4,33% | 5,10% |

In the proses of predicting data for 2019, the data processing will be conducted using data from 2010-2018. The data will be processed using SPSS, and the result are shown below:

a. Table Variable Entered/Removed

The table below describes the variables that are inputted and deleted, in this variable entered / removed table can be seen there are 5 variables that were successfully inputted and 0 variables deleted.

Variables Entered/Removed^a

| Model | Variables Entered | Variables Removed | Method |
|-------|---|-------------------|--------|
| 1 | Saori, STP, Ajinomoto, Masako, STB ^b | | Enter |

a. Dependent Variable: Omset Total

b. All requested variables entered.

Fig. 1. Variable entered/removed

b. Table Model Summary

The Model Summary table below explains the determination coefficients obtained by R-square = 1. This shows that 100 percent of the variation in dependent (Y) can be influenced by 5 independent variables, meaning that the influence of the independent variable is 100%.

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|--------------------|----------|-------------------|----------------------------|
| 1 | 1.000 ^a | 1.000 | 1.000 | 127.207 |

a. Predictors: (Constant), Saori, Ajinomoto, STP, Masako, STB

Fig. 2. Table model summary

c. Table Coefficients

In the coefficients table below, we can see the constant value of -0.3371 which will become β_0 in the multiple linear regression formula and the value of 1 in the ajinomoto, masako, STB, STP, and Saori will be β_1 , β_2 , β_3 , β_4 , β_5 in multiple linear regression formulas.

| Coefficients ^a | | | | | | |
|---------------------------|------------|-----------------------------|------------|---------------------------|-------------|------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -.371 | 72.200 | | -.005 | .996 |
| | Ajinomoto | 1.000 | .000 | .411 | 53311970.99 | .000 |
| | Masako | 1.000 | .000 | .457 | 39621545.35 | .000 |
| | STB | 1.000 | .000 | .114 | 8892685.709 | .000 |
| | STP | 1.000 | .000 | .027 | 2382506.290 | .000 |
| | Saori | 1.000 | .000 | .023 | 2345839.740 | .000 |

a. Dependent Variable: Omset Total

Fig. 3. Table coefficients

After that, a prediction calculation for 2019 is carried out using actual product data 2018.

The multiple linear regression equation is:

$$\hat{Y} = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \dots + \beta_iX_i$$

$$\hat{Y} = -0,371 + (1 * 112227257820) + (1 * 99048536549) + (1 * 21221374611) + (1 * 5035384039) + (1 * 4656462252)$$

Prediction calculation is done for 2019 by using 2018 product data as X Prediction turnover for 2019 after being included in the multiple linear regression equation using data from 2010 to 2018 above is Rp. 242,189,015,270.

d. Visualization

The visualization process will use visual data mining methods using Tableau tools software and data from Microsoft Excel.

1) Justify and Plan The Project

The first step is to determine the type of project that will be used for visualization. The type used is a pilot visual data mining project type which aims to analyze and answer the problem formulation of this research.

2) Identify The Top Business Questions

The second step is to identify the problem faced to get an answer. In this study we will find out the prediction of turnover in 2019 in PD. Asia Agung Pontianak and visualize turnover data for each product from 2010-2018 data.

3) Choose The Data Set

The third step is to select the data to be visualized, data obtained from PD. Asia Agung Pontianak and also the results of predictions that have been made before.

4) Transform The Data Set

The fourth step is the data used for predictions is the monthly data provided by the company but are combined into 1

sheet named "for tableau visualization" and also the data in the annual format.

5) Verify The Data Set

In the fifth step, check the data used for the application tableau software so that there is no error.

6) Choose Visualization Tools

At the sixth step is to choose the tools that will be used. In this study using the Tableau Software application obtained from the official tableau website for free.

7) Analyze The Visualization Model

In the seventh step is to analyze what will be displayed on the dashboard. The data that will be displayed at this stage is the development of each year each month and the annual turnover data growth.

8) Verify and Present The Visualization of Mining Model

After all data is entered and arranged in tableau, it will be used to create the dashboard using the chart available in Tableau, and produce the dashboard as shown below.

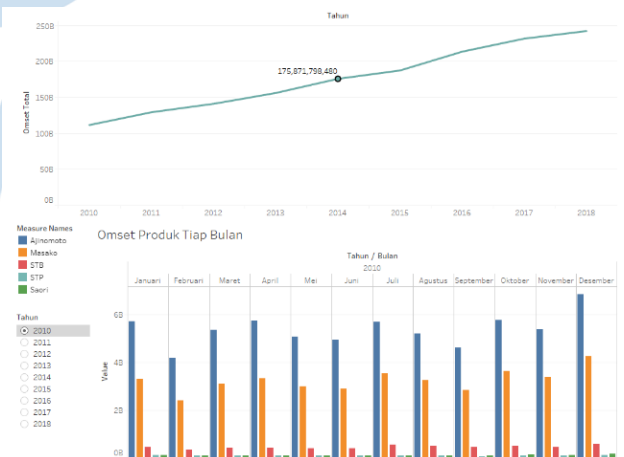


Fig. 4. Results of turnover data visualization dashboard in 2010-2018

V. CONCLUSION

This study aims to predict and visualize the turnover of companies using data from 2010-2018. The data consists of several columns consisting of months, places, total turnover, Ajinomoto, Masako, Sajiku Tepung Bumbu, Sajiku Tepung Praktis and Saori which are then converted to visualize in Tableau.

The results of the modified data are used for predictions using the Multiple Linear Regression method. The results of the research are:

1. Using the longer series of years data will make the results of error predictions decrease.
2. The result of the prediction amount for 2019 is Rp. 242,189,015,270.
3. Visualization is used to display how the development of turnover every month in each company in using real data from the company so that it can help facilitate the company in viewing and analyzing data on the company's turnover development each year.

REFERENCES

- [1] H. Ajinomoto, "Horeka Ajinomoto," Ajinomoto, 2012. [Online]. Available: <https://horeka.ajinomoto.co.id/distributor-penjualan/area-layanan>.
- [2] R. W. Abidatul Izzah, "Kinetik Vol. 2, No. 3," Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier, 2017.
- [3] Y. S. L. Intan Nur Farida, "Cogito Smart Journal/VOL. 1/NO. 1/DESEMBER 2015," Desember 2015. [Online]. Available: <http://cogito.unklab.ac.id/index.php/cogito/article/view/3/3>.
- [4] J. & R. B. Heizer, dalam Operations Management plus MyLab Operations Management, Global Edition, 10e, 10/E, 2010.
- [5] F. A. Minarni, "Jurnal Teknoif : Institut Teknologi Padang," Prediksi Jumlah Produksi Roti Menggunakan Metode Logika Fuzzy, p. Vol.4 No.2, 2016.
- [6] B. As'ad, "e-journal penelitian dan pengembangan komunikasi dan informatika," Prediksi Keputusan Menggunakan Metode Klasifikasi Naive Bayes, One-R, Dan Decision Tree, p. Vol 20 No.1, 2016 .
- [7] F. Hadiansyah, "Indonesian Journal on Computing (Indo-JC)," Prediksi Harga Cabai dengan Menggunakan pemodelan Time Series ARIMA, pp. Vol 2, No.1 , 2017.
- [8] S. E. A. A. G. Tarek Azzam, Data Visualization, Part 1, John Wiley & Sons, 2013, pp. https://dmlcommons.net/wp-content/uploads/2015/12/data_visualization_evaluation.pdf.
- [9] M. S. Hendro Poerbo Prasetya, "Jurnal Teknik Informatika dan Sistem Informasi jutisi.maranatha.edu," Visualisasi Informasi Data Perguruan Tinggi Dengan Data Warehouse Dan Dashboard System, p. Vol 2 no 3, 2016.
- [10] J. S. Taufik Dwi Saputra Tanwir, "ULTIMAInfoSys," Visualisasi Perbandingan Anggaran Pendapatan Dan Belanja Daerah (APBD) Pemerintah Provinsi Kabupaten Dan Kota Di Indonesia Periode 2010-2014, p. Vol 7 No.2, 2016.
- [11] I. D. Tom Soukup, "Visual Data Mining," dalam Techniques and Tools for Data Visualization and Mining, New Jersey: John Wiley & Sons, 2002, p. xxiv introduction.
- [12] M. A. D. A. A. Oleh Soleh, "Open Journal System Semnasteknomedia Online," Metode Peninjauan Dashboard dari Business Intelligent Untuk Membuat Keputusan Lebih Baik, pp. Vol 1, No.1, 2013.
- [13] Kusnawi, "Open Journal System "Jurnal DASI" Vol 12, No. 2," Tinjauan Umum Metode Pendekatan Dashboard Pada Proses Business Inteligence, 2011 .`
- [14] I. S. Statistic, "IBM SPSS Statistic," 2019. [Online]. Available: <https://www.ibm.com/id-en/marketplace/spss-statistics>.
- [15] S. E. Miner, "What Is SAS Enterprise Miner?," 2019. [Online]. Available: <http://support.sas.com/documentation/cdl/en/emgsj/67981/HTML/default/viewer.htm#n1fio8qutibtuxn1j7hyo6rqw7bj.htm>.
- [16] PowerBI, "What is Power BI," 2019. [Online]. Available: <https://powerbi.microsoft.com/en-us/what-is-power-bi/>.
- [17] Tableau, "What is Tableau?," 2019. [Online]. Available: <https://www.tableau.com/products/what-is-tableau>.