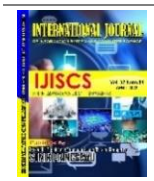


IJISCS (International Journal of Information System and Computer Science)
IJISCS, Volume 7, Number 3, December 2023
E ISSN: 2598-246X; P ISSN: 2598-0793, pp. 278-285
Accredited SINTA 4 Number 79/E/KPT/2023
https://jurnal.ftikomibn.ac.id/index.php/ijiscs

DATA MINING TO PREDICT INVENTORY AT DAPUR PINTAR STORE USING MULTIPLE LINEAR REGRESSION METHODS

Yuli Syafitri[1], Putri Mayang Sari[2], Rustam[3], Supriyanto[4]
[1,2,3,4]Informations System, Institut Teknologi Bisnis Dan Bahasa Dian Cipta Cendikia, Lampung, Indonesia
[1,2,3,4]Negara Street No.3 Candimas Kotabumi, North Lampung, Lampung, Indonesia

Email :
putrimayang616@gmail.com[1]
rustamdcc89@gmail.com[2]
Supriyanto@dcc.ac.id

Article history:
Received November 23, 2023
Revised December 12, 2023
Accepted December 31, 2023

Keywords:
Data Mining;
Multiple Linear Regression;
Inventory.

Abstract
Dapur pintar store is a shop that sells household furniture. However, there are several obstacles faced by Dapur Pintar Store, such as the shop still having difficulty predicting future inventory, and the absence of the application of data mining in predicting inventory. This is interesting to solve in order to make it easier for the store to determine inventory predictions at Dapur Pintar Store. Utilizing data mining, especially using linear regression, can provide valuable insight for predicting sales of household products at Smart Kitchen Stores. Identify and collect relevant data for sales analysis. This includes historical sales data, customer demographic data, promotions or discounts applied, and other factors that may influence household product sales. Perform data preprocessing to clean incomplete or inaccurate data. In addition, perform data transformations such as normalization to ensure consistency and accuracy in analysis. Apply a linear regression model to understand the linear relationship between the independent variable (for example, time, promotion, or demographics) and the dependent variable (household product sales). This model can be used to make predictions based on historical patterns. Use historical data to train a linear regression model. The training process involves adjusting the model parameters to fit the training data, so that it can provide accurate predictions. Validate the model using data that was not used in training to ensure that the model can provide good predictions on new data. This helps avoid overfitting and ensures generalization of the model. In making inventory predictions for goods, we will use the Multiple Linear Regression method. Software that will be used to support data processing is RapidMiner.

1.0 INTRODUCTION

The development of information technology resulted in a greatly increased need for information. To get information, data is needed and the right way to process existing data. The amount of data that can be processed. cause difficulties in data grouping. However, with the

development of Information Technology there is a solution to overcome these difficulties, one of which is using Data Mining Technology. Data is a compound term which means facts or parts of facts that contain meaning combined with facts, symbols, pictures, numbers, letters, or symbols that indicate an idea, object, condition, or situations and others. Data mining is a process of extracting data or filtering data by utilizing a large enough data set through a series of processes to obtain valuable information from the data[1]. Data mining can be used by large companies to mine data to obtain information that can support and improve the company's business processes. Prediction is a process of systematically estimating something that is most likely to happen in the future based on past and present information that is owned, so that the error (the difference between something that happened and the forecast result) can be minimized[2].

[3]Several studies that have been carried out focus on sales predictions, such as those carried out by association rules, which show that if a buyer buys a broom, they will tend to buy a mop, so it can be ascertained that Hasana Mart must provide a stock of brooms and mops, as can be seen from the transaction data. done very often. Efforts are made to place brooms and mops side by side to make it easier for buyers to search for the items they need. The above efforts are a sales strategy to increase consumer satisfaction by paying attention to associative patterns and consumer tendencies in buying and selling household furniture. [4] The C4.5 algorithm can be used to facilitate decision making by projecting existing data into a decision tree, based on the entropy and gain values of each data attribute. For more accurate prediction results, large amounts of data are needed, meaning that the greater the amount of data used, the more accurate the prediction results produced. [5] Based on the results of this research, Data Mining K-Means Clustering is able to provide information to the owner of the Kasih Ibu Restaurant regarding the sales grouping of existing menus, in the form of 10 types of menus that are most popular, 19 menus that are moderately popular, and 4 menus that are least popular. So it is hoped that it can be useful as a reference in making decisions regarding stocking menus at the Kasih Ibu Restaurant. [6] The results from the data mining prototype in a large medical distribution company provided the rationale for the strategy to reduce the total level of inventory by 50% (from a billion dollars to half a billion dollars) in the particular organization, while maintaining the same level of probability that a particular customer's demand will be satisfied. The second case study highlights the use of neural network-based data mining techniques for forecasting hot metal temperatures in a steel mill blast furnace.

Dapur Pintar store is a shop that sells household furniture. However, there are several obstacles faced by Dapur Pintar Store, such as the shop still having difficulty predicting future inventory, and the absence of the application of data mining in predicting inventory. This is interesting to solve in order to make it easier for the store to determine inventory predictions at Dapur Pintar Store. In making inventory predictions for goods, we will use the Multiple Linear Regression method. Software that will be used to support data processing is RapidMiner.

2.0 METHODOLOGY

2.1 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) is the process of determining useful information and patterns in data. This information is contained in a large database that was previously unknown and potentially useful. The following is the process of the KDD stages which can be seen in Figure 1.

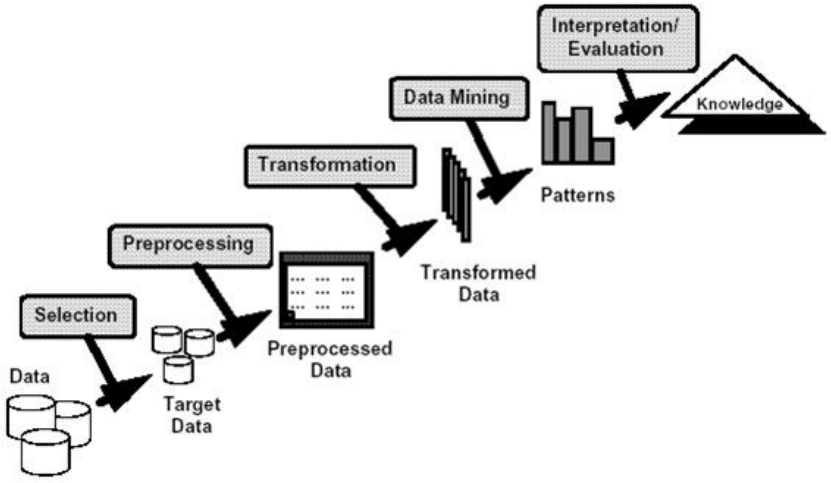


Figure 1. Stages in KDD[2], [3], [6][2], [7]–[9]

The stages of the KDD process consist of:

1. Data Selection
   Data selection from a set of operational data needs to be done before the information mining stage in KDD begins. Selected data used for data mining processes, stored in a file, separate from the operational database.
2. Pre-processing
   Before the data mining process can be carried out, it is necessary to carry out a cleaning process on the data that is the focus of KDD. The cleaning process includes removing duplicate data, checking inconsistent data, and correcting errors in data.
3. Transformation
   Coding is the process of transforming the data that has been selected, so that the data is suitable for the data mining process. The coding process in KDD is a creative process and really depends on the type or pattern of information to be searched for in the database.
4. Data Mining
   Data mining is the process of looking for patterns or interesting information in selected data using certain techniques or methods. Techniques, methods, or algorithms in data mining vary widely. The choice of the right method or algorithm really depends on the goals and the overall KDD process.
5. Interpretation
   Patterns of information resulting from the data mining process need to be displayed in a form that is easily understood by interested parties.

3.2 Multiple Linear Regression

[10], [11] Multiple Linear Regression is a linear relationship between two or more independent variables (X1, X2,..Xn) and variable (Y). Where to determine the actual value of the relationship between Y and x1 and x2. In general, the regression equation for k variables is as follows Y = a + b1 X1 + b2 X2 + ... = bn Xn

The regression equation for the two Independent Variables is Y = a + b1 X1 + b2 X2

3.0 RESULANTS

3.1 Data Analysis

Data analysis method is a method used to process research results to obtain conclusions. The data analysis used in this study is a quantitative analysis using multiple linear regression methods. At this stage the authors group the data obtained based on requests every month

and year as well as supplies from the store every month to simplify the process of calculating predictions.

Table 1. Stock Summary Data for Dapur Pintar Stores in 2020-2022

Number	Month	2020	2021	2021
1	January	3813	3847	3839
2	February	3834	3871	3861
3	March	3837	3864	3849
4	April	3804	3789	3797
5	May	3801	3796	3789
6	June	3851	3846	3846
7	July	3790	3798	3793
8	August	3971	3976	3981
9	September	4171	4161	4161
10	October	4334	4329	4319
11	November	4351	4361	4356
12	December	4480	4475	4455
	Total	**48037**	**48ll3**	**48046**

3.2 Data Mining Implementation Results

The process of calculating inventory predictions at the Smart Kitchen Shop uses the Multiple Linear Regression method using Microsoft Excel 2010. With information obtained from the 2020-2022 goods data at the Smart Kitchen Shop, the information is as follows:

- ✓ Inventory For 2020 = 3.154
- ✓ Inventory For 2021 = 3.169
- ✓ Inventory For 2022 = 3.164

From the data above, the author will perform prediction calculations using the RapidMiner 7.1 application with the following steps. Data training Which used in this implementation using data inventory in 2020.

Table 2. Inventory Training Data for 2020

	First stock (X1)	Total Sales (X2)	Price (QG)	Last stock (Y)
January	3813	2118	16525500	1695
February	3834	1955	16525500	1871
March	3837	2015	16525500	1817
April	3804	1973	16528000	1821
May	3801	2054	16537000	1737
June	3851	2167	16528000	1684
July	3790	2154	16529000	1636
August	3971	2327	16531500	1634
September	4171	2223	16533000	1938
October	4334	2351	16532000	1983
November	4351	2398	16534000	1953
December	44SO	2597	16535000	1878
Amount	**48037**	**26332**	**198364000**	**21647**

The test data used for implementation on RapidMiner this time is December data from 2020, the results of which can later be used to predict supplies for the following month, January 2021.

Table 3. Test Data for 2021 Inventory Prediction

	First stock (X1)	Total Sales (X2)	Price (QG)	Last stock (Y)
December	4480	2597	16535000	

The following is an implementation of predictions in Rapid Miner using training data and test data above.

1. The first stage is to find the read excel operator to then drug it to main process

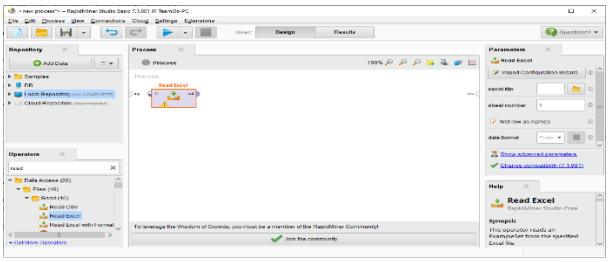


Figure 2. Entering Read Excel to Main Process

2. Next, enter the training data that has been prepared by importing data through the import configuration wizard as shown below.

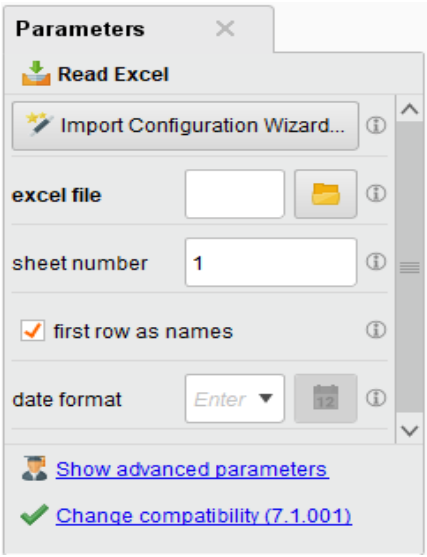
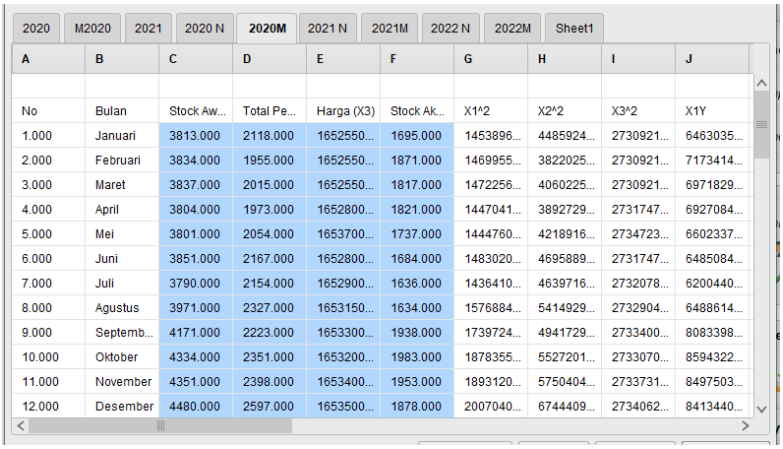


Figure 3. Display Entering Files

3. Block of data that will be used as training data



2020 | M2020 | 2021 | 2020 N | **2020M** | 2021 N | 2021M | 2022 N | 2022M | Sheet1

A	B	C	D	E	F	G	H	I	J
No	Bulan	Stock Aw...	Total Pe...	Harga (X3)	Stock Ak...	X1^2	X2^2	X3^2	X1Y
1.000	Januari	3813.000	2118.000	1652550...	1695.000	1453896...	4485924...	2730921...	6463035...
2.000	Februari	3834.000	1955.000	1652550...	1871.000	1469955...	3822025...	2730921...	7173414...
3.000	Maret	3837.000	2015.000	1652550...	1817.000	1472256...	4060225...	2730921...	6971829...
4.000	April	3804.000	1973.000	1652800...	1821.000	1447041...	3892729...	2731747...	6927084...
5.000	Mei	3801.000	2054.000	1653700...	1737.000	1444760...	4218916...	2734723...	6602337...
6.000	Juni	3851.000	2167.000	1652800...	1684.000	1483020...	4695889...	2731747...	6485084...
7.000	Juli	3790.000	2154.000	1652900...	1636.000	1436410...	4639716...	2732078...	6200440...
8.000	Agustus	3971.000	2327.000	1653150...	1634.000	1576884...	5414929...	2732904...	6488614...
9.000	Septemb...	4171.000	2223.000	1653300...	1938.000	1739724...	4941729...	2733400...	8083398...
10.000	Oktober	4334.000	2351.000	1653200...	1983.000	1878355...	5527201...	2733070...	8594322...
11.000	November	4351.000	2398.000	1653400...	1953.000	1893120...	5750404...	2733731...	8497503...
12.000	Desember	4480.000	2597.000	1653500...	1878.000	2007040...	6744409...	2734062...	8413440...

Figure 4. Display of training data

4. Next, change the attribute data type in the month and inventory column. The attribute month column is changed to id while for the attribute inventory label.

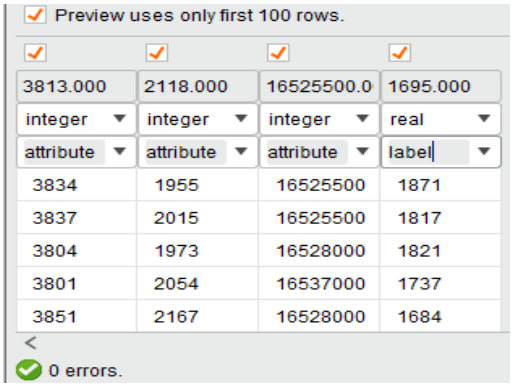


Figure 5. Display Change Data Type

5. After importing the training data, the next step is to find operators Linear Regression to then drag to the main process.

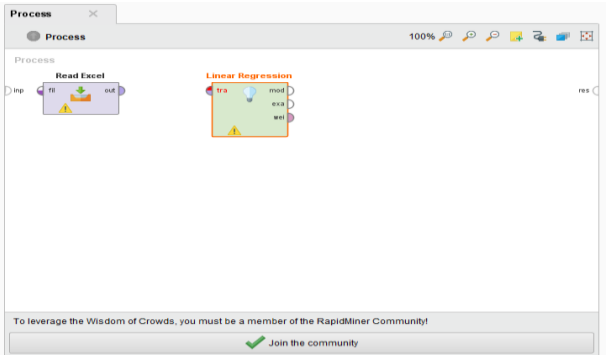


Figure 6. Linear Regression Operator on Main Process

6. Look for the Apply Model operator then drag it to the main process.

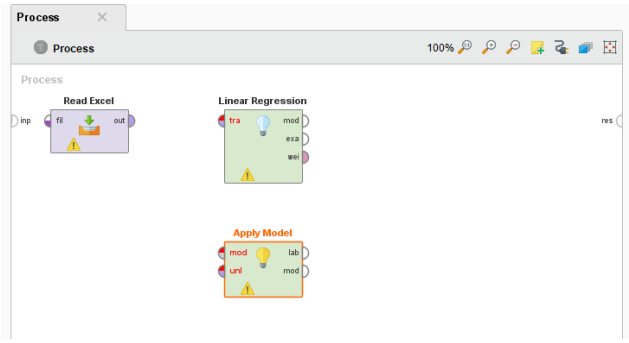


Figure 7. Appearance of the Apply Model in Main Process

7. Then look for read excel to import the tested data prepared.

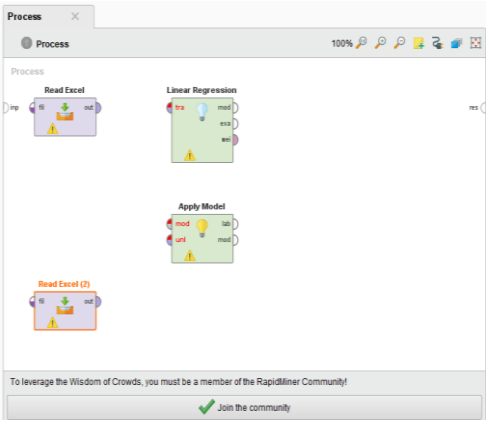


Figure 8. Display of Read Excel Test Data

7. Then import the test data and the test data block that will be used for prediction

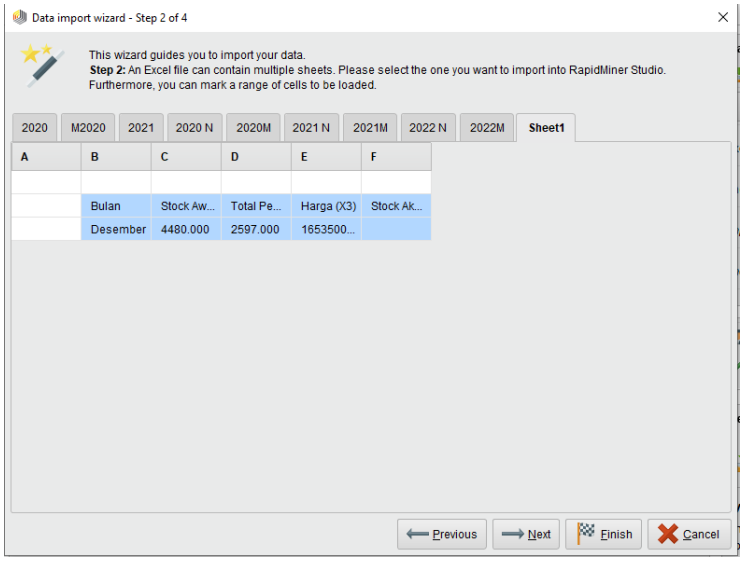


Figure 9. Test Data

8. To get the predicted results, connect the training data to the linear regression operator then connect it to the apply model, on the test data connect it to the apply model operator then connect it to the result.

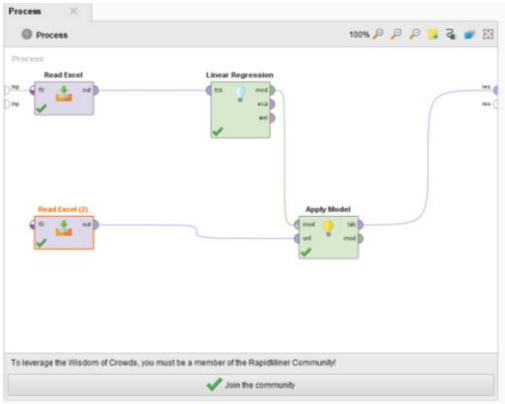


Figure 10. Linear Regression Implementation Model Arrangement

Using data mining techniques such as regression or time series analysis, stores can forecast demand for certain products at certain times. This helps in production planning and ordering raw materials. Data mining can also be used to analyze operational efficiency, such as service time, customer satisfaction levels, and employee performance. This understanding can help restaurants to improve services and optimize their operations. With data mining, you can identify trends quickly and respond quickly by adjusting products or creating special promotions. This is important to remain competitive and meet changing customer expectations.

5.0 CONCLUSION

In this study, modeling was carried out using multiple linear regression. Based on the results of data mining calculations using the multiple linear regression method, the prediction results for 2021 with the RapidMiner software are obtained, namely 1880,841 with an error value of 18.77. This prediction can help companies in making decisions and reduce the occurrence of inventory shortages. Through data mining and analysis of historical sales data, Dapur Pintar Store can understand customer demand patterns. This includes seasonal trends, product preferences, and changes in consumer behavior. With this understanding, companies can

make more accurate predictions about future demand. Implementing a prediction model using machine learning techniques can help companies predict the level of sales of certain products in a certain period. This helps in planning inventory and ordering goods more efficiently.

REFERENCES

[1] T. A. Muhammad Husni Rifqo, "Implementasi Algoritma C4.5 Untuk Menentukan Calon Debitur Dengan Mengukur Tingkat Risiko Kredit Pada Bank Bri Cabang Curup," *J. Pseudocode*, vol. 3, no. 2, pp. 83–90, 2016.

[2] D. A. C, D. A. Baskoro, L. Ambarwati, and I. W. S. Wicaksana, *Belajar Data Mining dengan RapidMiner*. 2013.

[3] N. Silalahi, "Penerapan Data Mining Dalam Prediksi Penjualan Prabot Rumah Tangga Menggunakan Metode Apriori Pada Toko Hasanah Mart," *Build. Informatics, Technol. Sci.*, vol. 2, no. 1, pp. 33–38, 2020.

[4] A. Zakir, Y. Ndruru, and E. Hadinata, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Makanan Terlaris Dengan Algoritma C45," *JIFTI - J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, no. 2, pp. 7–12, 2020.

[5] H. A. Siregar, N. Yanti, and L. Gaol, "Penerapan Data Mining Pada Penjualan Rumah Makan Kasih Ibu Menggunakan Metode K-Means Clustering," *J. Sist. Inf. TGD*, vol. 2, no. 5, pp. 750–760, 2023.

[6] A. Dhond, "Data Mining Techniques for Optimizing Inventories for Electronic Commerce," in *Knowledge discovery and data mining*, 2020, pp. 480–486.

[7] S. Ipnuwati, "Sistem Pendukung Keputusan Perencanaan Promosi Kampus Berbasis Data Mining Dengan Metode Klasifikasi Pada Stmik Pringsewu Lampung," 2013.

[8] F. S. Fauzi, Rita Irviani, Andino Maseleno, "Revolutionizing Education through Technology : Big Data and Online Learning," in *CICCSE*, 2017, vol. 1, no. 1, p. 44.

[9] K. Rismayanti, Fera Damayanti, "Penerapan Data Mining Algoritma C4.55 Dalam Menentukan Rekam Jejak Kinerja Dosen STT Harapan Medan," *J. Penelit. Tek. Inform.*, vol. 3, no. 1, pp. 99–104, 2018.

[10] L. S. Viana Ilma Edyan, "Penerapan Model Sekuensial Linier Pada Rancang Bangun Aplikasi Siskah Sma Rimba Madya Bogor," vol. 4, no. November, pp. 71–84, 2019.

[11] A. Faradina and B. Satrio, "Pengaruh Promosi dan Kualitas Pelayanan Terhadap Kepuasan Pelanggan pada Rumah Cantik Almanda," *J. Ilmu Dan Ris. Manaj.*, vol. 5, no. 7, pp. 1–18, 2016.