



Implementation of data mining to estimate the need for toast bread supply at junction cafe using the multiple linear regression method

Sutrisno Situmorang¹, Sartika Dewi Purba², Leliana Harahap³, Kamson Sirait⁴, Jonas Franky Rudianto Panggabean⁵

^{1,2,3,4,5}Akademi Informatika dan Komputer Medicom

Article Info	ABSTRACT
<p>Article history:</p> <p>Received Jun 27, 2024 Revised Jul 05, 2024 Accepted Jul 15, 2024</p> <hr/> <p>Keywords:</p> <p>Data Mining Estimate Multiple Linear Regression Supply</p>	<p>Junction Cafe Medan, which is managed by individuals, has a large supply of bread services with different specifications. The inventory system for toasted bread at Junction Cafe Medan still uses a manual system in data processing. Handling data with this system has several obstacles: including causing shortages and excesses of bread stock, impacting guests due to running out of bread stock, excessive costs incurred for stocking bread, and lack of accuracy in recording incoming and outgoing bread stock resulting in shortages and errors in ending stock inventory. Based on this research problem, a data mining application is needed that is capable of estimating bread supplies at Junction Café Medan, where each Bread inventory data at Junction Café Medan will be calculated using one of the data mining methods that is capable of estimating bread supplies based on bread usage by applying the Regression method. The result of this research is a data mining application that uses the Multiple Linear Regression method which is able to solve the bread stock inventory problem at Junction Cafe Medan by estimating bread stock inventory more quickly and accurately.</p>

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

eliana Harahap,
eknik Komputer,
kademi Informatika dan Komputer Medicom,
.Darat No.74 Medan, Sumatera utara, 20153, Indonesia,
mail : leliharahap05@gmail.com

Introduction

Junction Cafe Medan, which is managed by individuals, has a large supply of bread services with different specifications. The toast inventory system at Junction Cafe Medan still uses a manual system in data processing. Handling data with this system has several problems, including causing shortages and excesses of bread stock, impacting guests due to running out of bread stock, costs incurred too much to stock bread, and less accurate recording of incoming and outgoing bread stock, resulting in shortages and errors in ending stock inventory. Inventory has an important (even essential) role in every organization. Without supplies most activities cannot be carried out (H.D. Khanh, 2020).

The characteristics of the inside of good bread include: Bright inside color (crumb) (Rosa.A.S, 2019). This is what makes bread highly recommended for those of you who want to lose weight (Wati, 2019). Whole bread contains more nutrients such as vitamin B and fiber than white bread, this is due to the loss of layers in the formation of white wheat flour (Hakim, 2021). The advantages of whole wheat bread make it the right choice for those of you who are losing weight (D. S. Purnia, 2019). One way to maintain ideal body weight is to maintain food intake (Muhamad Muslih, 2019).

Various studies have proven that regularly consuming whole wheat can reduce the risk of heart disease by up to 22% (Tesa Nur Padilah, 2019). One of the important elements in maintaining and improving the quality of restaurant service is the availability of sufficient food stock, to be able to provide the best service at the cafe, so that the budget for providing bread is mostly used to organize service efforts. and the costs used (Harsiti, 2022). Various studies have proven that regularly consuming whole wheat can reduce the risk of heart disease by up to 22% (Tesa Nur Padilah, 2019). One of the important elements in maintaining and improving the quality of restaurant service is the availability of sufficient food stock, to be able to provide the best service at the cafe, so that the budget for providing bread is mostly used to organize service efforts. and the costs used (Harsiti, 2022). Data Mining is the process of looking for interesting patterns or information in selected data using certain techniques or methods (D. Hartama, 2019).

The solution that can be applied to solve the bread supply problem at Junction Cafe Medan is to use the multiple linear regression method. Multiple linear regression is a regression model that involves more than one independent variable. Multiple linear regression analysis was carried out to determine the direction and how much influence the independent variable has on the dependent variable. In this research example, a multiple linear regression test was carried out to get an idea of how the independent variables which include CSR, liquidity, capital intensity and inventory intensity affect the dependent variable, namely tax aggressiveness, with a level of significance. To process data, data mining is needed. Data Mining is the process of looking for interesting patterns or information in selected data using certain techniques or methods (D. Hartama, 2019). Data Mining itself has several methods, one of which is the multiple linear regression method. The multiple linear regression method is widely used in the title Researching production forecasting using the multiple linear regression method (Mohammed J. Zaki, 2020). Simple linear regression only requires 1 variable, while multiple linear regression uses more than 1 estimated variable (Amrin, 2021). And also in research on optimizing multiple linear regression modeling in predicting the number of motorbike accidents using genetic algorithms (Sulistiyowati, 2020) and so on. However, among these algorithms there is one algorithm that is often used in data mining to analyze borrowing and association patterns (Alkhairi, 2019).

Method

Research Method is a stage or a scientific method used to obtain data that will be used to solve problems by conducting direct field studies to collect data. The framework is the steps in solving the problems discussed.

The stages carried out in the research in order to solve the problems described in the previous chapter, including the background to the problem, include:

Data Collection Method

In general, research data collection methods can be divided into questionnaire methods, observation, interviews, focus group discussions, laboratory and field experimentation. Data collection in this research was carried out at the Junction Cafe using 2 methods. The following is the description used:

a. Observation

Observation is a data method that is widely used in experimental designs (laboratory and field) and qualitative studies (ethnography, exploration and participant observation).

b. Interview

Interviews are a data collection method that is widely used in exploratory research and field studies. In principle, interview techniques are no different from questionnaires with an open question/question format. Interviews aim to obtain information or confirm some information directly to respondents or participants. It will be used in determining the features to be built. The interview stage was carried out by interviewing Junction Cafe staff about the estimated bread supply needs

Data Mining

Data Mining is an analysis step towards the knowledge discovery process in a database or Knowledge Discovery Database (KDD) (Moore, 2020). In particular, data mining has its own methods based on the objectives of using data sets, namely estimation, prediction, classification, clustering and association (Dewayanti, 2021). Data mining functions can be combined in certain cases to answer existing problems, the following are general data mining functions (Elmayati, 2019): (a) Classification (b) Clustering (Clusterring), (c) Regression (Regression), (d) Anomaly Detection, (e) Summarization

Some of the goals of Data Mining are (Alican Dogan, 2022): (a) Know the loss of customers to competitors. (b) Know product or consumer items that have similar characteristics. (d) Identify products that are sold together with other products. (e) Predict sales levels and goods production targets. (f) Assess the level of risk in determining the production quantity of an item.

Results and Discussions

Data Collection Method

In general, research data collection methods can be divided into questionnaire methods, observation, interviews, focus group discussions, laboratory and field experimentation (S. Y. Fraticasari, 2020). Data collection in this research was carried out at the Junction Cafe using 2 methods, the following is the description used (Reavindo, 2021): (a) Observation (b) Interview

System Design Methodology

System Design Methodology is a stage that must be carried out after analyzing a problem, it is at this stage that the design of a system is planned. One way to design or build a system is to use the Waterfall Method (Rahman, 2020). (a) Problem and Need Analysis, (b) System Design and Modeling, (c) Coding, (d) Initial Trial, (e) Final Trial, (f) System Implementation

System Algorithm

System Algorithms are the steps taken by a system in processing and solving a problem. The following is a flowchart or flow of problem solving using the Multiple Linear Regression method. The waterfall method or what is often called the waterfall method is often called the classic life cycle. The name of this model is actually "Linear Sequential Model" which describes a systematic approach (Hamid Kurniawan, 2020)

Multiple Linear Regression Method

The Multiple Linear Regression Method is a regression that predicts the relationship between one dependent variable (Y) and two or more independent variables (x_1, x_2, \dots, x_n) (Wahid, 2020). This analysis was carried out in order to determine the relationship between the independent variable and the dependent variable. To predict Y, if all the values of the independent variables are known, the multiple linear regression equation is used. The actual relationship between Y and x_1, x_2, \dots, x_n , is as follows:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad \dots\dots\dots (1)$$

Information:

Y = Dependent variable

a = Constant

b_1, b_2, b_3 = Regression coefficient (value of increase or decrease)

x_1, x_2, x_3 = Independent variables

To obtain the regression coefficients a, b_1 and b_2 , they can be obtained simultaneously from the three equations as follows:

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2$$

To balance the numbers for each variable, the variable Y will be divided by 10. The following is the data initialization process where each data sheet will be labeled with a variable name so that it is easy to use the Linear Regression algorithm:

Table 1. Data Simplification

No	Supplies	Usage	Request
1	5.6	1.2	17.5
2	3.5	1.4	16.3
3	4.5	1.5	16
4	6.5	1.6	15
5	8.6	1.2	16.6
6	2.3	1.5	12.3
7	7.5	0.8	13
8	7.6	0.9	14.5
9	8.5	1.6	11.5
10	9.5	1.2	10.4

Calculating Regression Coefficients

Regression coefficients are important in analyzing regression. The benefit of regression coefficients is to form a regression equation model for the problem being studied.

$$\begin{aligned} \sum Y &= na + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y &= a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y &= a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{aligned}$$

Table 2. Regression Coefficients

No	X ₁	X ₂	Y	X ₁ Y	X ₂ Y	X ₁ X ₂	X ₁ ²	X ₂ ²
1	5.6	1.2	17.5	98	21	6.72	31.36	1.44
2	3.5	1.4	16.3	57.05	22.82	4.9	12.25	1.96
3	4.5	1.5	16	72	24	6.75	20.25	2.25
4	6.5	1.6	15	97.5	24	10.4	42.25	2.56
5	8.6	1.2	16.6	142.76	19.92	10.32	73.96	1.44
6	2.3	1.5	12.3	28.29	18.45	3.45	5.29	2.25
7	7.5	0.8	13	97.5	10.4	6	56.25	0.64
8	7.6	0.9	14.5	110.2	13.05	6.84	57.76	0.81
9	8.5	1.6	11.5	97.75	18.4	13.6	72.25	2.56
10	9.5	1.2	10.4	98.8	12.48	11.4	90.25	1.44

Dari tabel diatas maka diketahui nilai :

- $\sum X_1 = 209.6$
- $\sum X_2 = 34.6$
- $\sum Y = 428.1$
- $\sum X_1^2 = 1709.66$
- $\sum X_2^2 = 44.3$
- $\sum X_1 X_2 = 256.6$
- $\sum X_1 Y = 3031.67$
- $\sum X_2 Y = 509.03$

Simplifying the Linear Regression Equation

After the values $\sum x_1^2$, $\sum x_2^2$, $\sum y^2$, $\sum x_1 x_2$, $\sum x_1 y$, and $\sum x_2 y$ are obtained, a linear equation will be formed. The following is the linear equation formed:

$$\sum Y = a n + b_1 \sum X_1 + b_2 \sum X_2 \dots\dots\dots (2)$$

$$428.1 = 29a + 209.6b_1 + 34.6b_2 \dots \text{Equation I}$$

$$\begin{aligned}\sum X_1Y &= a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1X_2 \\ 3031.67 &= 2096a + 1709.66b_1 + 256.6b_2 \dots \text{Equation II} \\ \sum X_2Y &= a \sum X_2 + b_1 \sum X_1X_2 + b_2 \sum X_2^2 \\ 5090.30 &= 346a_0 + 256.6b_1 + 44.3b_2 \dots \text{Equation III}\end{aligned}$$

The values, b1 and b2 are obtained by substitution or elimination based on the 3 equations obtained in the previous process elimination of Equations I and II

$$\begin{array}{r} 428.1 = 29a + 209.6b_1 + 34.6b_2 \quad | \times 209.6 \\ 30316.7 = 209.6a + 1709.66b_1 + 256.6b_2 \quad | \times 29 \\ \hline 897297.6 = 60784a + 4393216b_1 + 725216b_2 \\ 879184.3 = 60784a + 4958014b_1 + 744140b_2 \\ \hline 18113.3 = -564798b_1 - 18924b_2 \end{array}$$

$$18113.3 = -564798b_1 - 18924b_2 \dots \text{Equation IV}$$

Eliminate Equations I and III

$$\begin{array}{r} 428.1 = 29a + 209.6b_1 + 34.6b_2 \quad | \times 34.6 \\ 5090.30 = 34.6a + 256.6b_1 + 44.3b_2 \quad | \times 29 \\ \hline 148122.60 = 10034a + 725216b_1 + 119716b_2 \\ 147618.7 = 10034a + 744140b_1 + 128470b_2 \\ \hline 503.90 = -18924b_1 - 8754b_2 \end{array}$$

$$503.90 = -18924b_1 - 8754b_2 \dots \text{Equation V}$$

Eliminate Equations IV dan V

$$\begin{array}{r} 18113.3 = -564798b_1 - 18924b_2 \quad | \times -18924 \\ 503.90 = -18924b_1 - 8754b_2 \quad | \times -564798 \\ \hline -342776089.2 = 10688237352b_1 + 358117776b_2 \\ -284601712.2 = 10688237352b_1 + 49442416926b_2 \\ \hline -58174377 = -4586123916b_2 \\ b_2 = 7.621498511 \end{array}$$

$$\text{Substitute } b_2 \text{ into the V equation } 503.90 = -18924b_1 - 8754b_2$$

$$503.90 = -18924b_1 - 8754(0.012684868)$$

$$503.90 = -18924b_1 - 8754(0.012684868)$$

$$503.90 = -18924b_1$$

$$614.94 = -18924b_1 - 111.0433354$$

$$b_1 = 0.242871773$$

Substitute b_1 and b_2 To Equation I

$$428.1 = 29a + 209.6b_1 + 34.6b_2$$

$$428.1 = 29a + 209.6(0.242871773) + 34.6(7.621498511)$$

$$428.1 = 29a + -68.11040113 + 4.388964365$$

$$428.1 = 29a - 63.72$$

$$491.82 = 29a$$

$$a = 3.1525056333$$

From the results of the calculations a, b1, b2 above, if the results are entered into the following equation:

$$Y = a + b_1X_1 + b_2X_2 \dots\dots\dots (3)$$

This will produce the equation below :
 $Y = 3.152506333 + 0.03249542 X_1 + 0.012684868 X_2$

Case testing based on equations: Junction Cafe wants to estimate the supply of bread based on data on the number of requests/day and the number of uses (sold), for example, let's say the number of requests is 50 and the number of uses is 12 boxes of bread, then the formula is as follows: $Y = a + b_1X_1 + b_2X_2$
 $Y = 3.152506333 - 0.03249542 X_1 + 0.012684868 X_2$
 $Y = 3.152506333 - 0.03249542 (50) + 0.012684868 (12)$
 $Y = 106.7540771$
 $Y = 106.7540771 \times 10$ (because during normalization the variable y is divided by 10)
 $= 106.7540771$

Based on the calculations above, an estimate of bread inventory can be obtained based on data on the number of requests/day and the number of uses (sold), so the inventory value obtained is 106,754 or 107 Boxes of Toasted Bread.

Interface Display Results are the stage where the system or application is ready to be operated in its actual state according to the results of the analysis and design carried out, so that it will be known whether the system or application designed can actually achieve the goals achieved.

This Data Mining application is equipped with a display that aims to make it easier for users. The function of this interface is to provide input and display output from the application. This application has an interface consisting of a Login Form, Main Menu Form, Inventory Data Form, Regression Form and Report Form.

Login Form

The Login Form is used to secure the system from irresponsible users before entering the Main Menu. The following is the appearance of the Login Form:

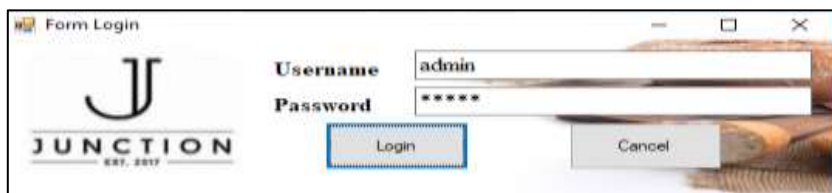


Figure. 1. Form Login

The following is the information in Figure 5.1 Login Form: (a) The Login button is used to validate the username and password that we have filled in in the text box provided. (b) The Cancel button is used to close the login form.

Main Menu Form

The Main Menu Form is used as a link for the Inventory Data Form, Regression Form and Report Form. The following is a display of the main menu form.



Figure 2. Form Main Menu

Inventory Data Form

Inventory Data Form is a form used to manage Inventory Data in the System. The following is the Inventory Data form display:



Figure 3. Inventory Data Form

Regression Form

Regression Form is a form used to find Sales Estimates. The following is the appearance of the Regression Form:



Figure 4. Form Regresi

Report Form

The Report Form is a form used to display the results of the Linear Regression algorithm regarding estimating bread supplies. The following is a display of the Report form:

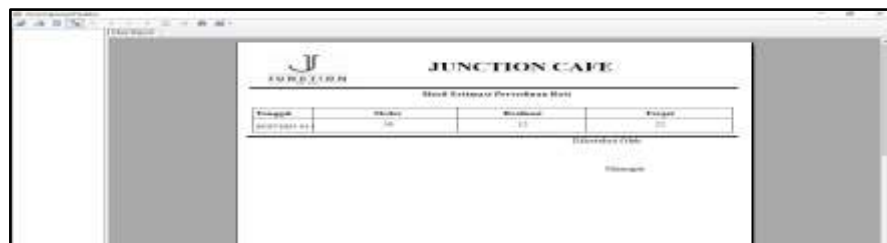


Figure 5. Report Form

System Testing

After implementation is carried out, the next step is to carry out system testing of the linear regression method calculation process. This system testing is intended to find out how accurate and precise the application has been designed and to find out what bugs are found. The following is the data that will be processed.

The data is then input into the system on the Inventory Data Form, and you can also use the Import data function.



Figure 6. Inventory Data Form

After the data matches what was input into the system on the Inventory Data Form, then calculate the data value based on the following data:

Demand = 50 => $x_1 = 50$

Usage = 12 => $x_2 = 12$



Figure 7. Form Regresi

Conclusions

The conclusion of this system is as follows: This application can perform calculations to estimate inventory using linear regression quickly, thereby saving time and being more efficient compared to the estimation results carried out by Junction Cafe management. The system was created with a simple appearance and is easy to use by Junction Cafe. The estimation results issued by the system have a high level of accuracy, so that the estimated value obtained can be used as a sales reference.

References

- Amrin2021Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi*Jurnal Techno Nusa Mandiri*
- Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi2020*Jurnal Ilmu-ilmu Informatika dan Manajemen STMik*
- Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang2019*Jurnal Pendidikan Matematika dan Matematika*
- Data Mining and Machine Learning Fundamental Concepts and Algorithms*2020Cambridge University Press
- Elmayati2019Data Mining Dengan Metode Clustering Untuk Pengolahan Informasi Persediaan Obat Pada Klinik Srikandi Medika Berbasis Web*Jurnal Pelita Informatika*
- Machine learning and data mining in manufacturing2022*Expert Systems with Applications*
- Memprediksi Pengaruh Limbah Pada Biaya Proyek Menggunakan Regresi Linier Berganda2020*Prosiding IEEEJurnal Informatika Upgris*
- Memprediksi Pengaruh Limbah Pada Biaya Proyek Menggunakan Regresi Linier Berganda2020*Jurnal Informatika Upgris*

- Optimasi Pemodelan Regresi Linier Berganda Pada Prediksi Jumlah Kecelakaan Sepeda Motor Dengan Algoritme Genetika2020*Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*
- Penerapan Data Mining Pada Data Nilai Siswa Dengan Menggunakan Algoritma Asosiasi Rule Metode Apriori2021*Seminar Nasional*
- Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara2019*Seminar Nasional Teknologi Komputer & Sains*762-767
- Penerapan Metode Waterfall dalam Perancangan Sistem Informasi Aplikasi Bantuan Sosial Berbasis Android2019*Semnastek*
- Penerapan Metode Waterfall Dalam Perancangan Sistem Informasi Penggajian Pada SMK Bina Karya Karawang 2020*Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*
- Pengaruh Luas Panen Dan Harga Produksi Tanaman Jagung Kabupaten Karo2021*Agrica (Jurnal Agribisnis Sumatera Utara)*74-79
- Peramalan Produksi dengan Metode Regresi Linier Berganda2020*Productivity, Optimization and Manufacturing System Engineering*
- Perbandingan Metode Fuzzy Dengan Regresi Linier Berganda Dalam Peramalan Total Produksi2019*Garuda*273-284
- Prediksi Persediaan Obat Dengan Metode Regresi Linier2022*Jurnal Sistem Informasi*
- Rosa.A.S2019*Pemodelan sistem Rekayasa perangkat lunak*JakartaPT. Elex Media Komputindo
- Strategi Pemasaran Dengan Regresi Linier Untuk Kepentingan Siswa Baru, Kuala Lumpur2019*Kuala Lumpur*
- The Application of Data Mining in Determining Patterns of Interest2019*Journal of Physics*
- Visual Basic 2008 for Pemula Banget*2021JakartaPT. Elex Media Komputindo
- X-means: Extending k-means with efficient estimation of the number of clusters2020*IOP Conf. Series: Materials Science and Engineering*School Of computer Science