Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Predicted Sales of Industrial Homes Exclusive Anugrah Bean Cake Using the Linear Regression Method

Wulan Ayu Ananda Surya[1*], Novriyenni[2], Lina Arliana Nur Kadim[3]

[1,2,3]*STMIK Kaputama*

Wulansurya2000@gmail.com[1*], novriyenni.sikumbang@gmail.com[2], lina_arliana@yahoo.com[3]

Abstract

UD. Anugrah Exclusive is a peanut cake home industry in Binjai City that faces monthly sales fluctuations so that it has an impact on the management of raw material stocks. This study aims to build a sales prediction model using a simple linear regression method with sales data for the period January 2021–June 2025. Independent variables are in the form of time (months) and dependent variables are in the form of sales (pouch). The model is implemented in a web-based system using Python and MySQL and evaluated using Mean Absolute Percentage Error (MAPE). The results of the study resulted in a regression equation **Y = 349.55 + 2.55X** with **a MAPE accuracy rate of 9.03%,** which is in the very good category. The system built can help business owners estimate raw material needs, avoid the risk of overstock or lack of stock, and develop a more appropriate marketing strategy.

Keywords: *Sales Prediction, Linear Regression, MSMEs, Web Dashboard*

1. Introduction

UD. Anugrah Exclusive is a Home Industry business engaged in the sale of peanut cakes. This product has a distinctive taste and is offered at an affordable price, thus attracting many customers, especially ahead of certain holidays and celebrations. However, in the highly competitive culinary business world, UD. Anugrah Exclusive faces challenges in maintaining the stability of its sales from time to time.

Sales instability is a major problem that has a direct impact on the management of raw material stocks. If not managed properly, excess stock can lead to losses due to unused or expired raw materials. Based on peanut cake sales data from UD. Anugrah Exclusive during the period from January 2021 to June 2025, it can be seen that the number of sales per month has fluctuated significantly. For example, in 2021 total sales were recorded at 4,414 pouch, while in 2024 it increased to 5,242 pouch, then in just the first six months of 2025 it reached 3,118 pouch. These fluctuations can be caused by many factors, including competition, seasonal trends, as well as promotional strategies.

It is important for business owners to monitor and analyze sales data on a regular basis. One method that can be used to aid in business decision-making is linear regression, which is able to predict sales based on historical patterns. From the above problems, this research aims to answer the main problem, namely how to maintain sales stability and be able to build a raw material supply strategy so as to prevent losses that will occur. Thus, the author is interested in conducting a final project research with the title "PREDICTION OF SALES OF THE ANUGRAH EXCLUSIVE PEANUT CAKE INDUSTRY HOME USING THE LINEAR REGRESSION METHOD".

2. Literatur Riview

2.1. Data Mining

Data mining is the process of extracting useful information from big data using math, statistics, and artificial intelligence techniques. The information generated from the data mining process can be used to help decision-making and predictions in various fields, such as business, science, technology, health, and others. The definition of data mining has evolved along with the development of technology and research [1]. Data mining is a series of processes that aim to find added value from a set of data by digging up previously unknown knowledge manually. Another definition states that data mining is the process of obtaining valuable information from large databases. In addition, data mining can also be interpreted as an effort to extract new information from large data sets to support decision-making [2].

2.2. Predictions

According to the great dictionary of the Indonesian language (KBBI), prediction is the result of the activity of predicting or estimating values in the future using data from the past. Prediction is the activity of predicting or predicting something that will happen in the future. In terms, the meaning of predictions can vary depending on the context and problem being discussed. Understanding language, prediction is interpreted as a prediction or forecast that has become a general definition [3].
Prediction is a science that is used to gain knowledge systematically based on data. Prediction is also the process of predicting future events by utilizing various relevant information or data from previous periods [4].

2.3. Sales

Sales is an activity in the field of marketing that aims to offer and distribute company products to consumers so that they can be used and utilized as needed. This activity can be seen as a program that includes various marketing efforts to facilitate and simplify the process of distributing goods from the company to buyers.

2.4. Home Industry

Home Industry Law of the Republic of Indonesia No. 20 of 2008 article 1, that "home industry is a productive business owned by an individual or an individual business entity that meets the criteria of micro business.

2.5. Regresi Linear

Simple Linear Regression is a statistical method that functions to test the extent of the causal relationship between the Causative Factor Variable (X) and the Causal Variable (Y). Causal Factors are generally denoted by X or also referred to as (*independent*, predictor) while Consequential Variables denoted by Y are also referred to as (*dependent*, response). The pattern shown by simple regression analysis assumes that the relationship between the two variables can be expressed in a straight line. Simple Linear Regression or often abbreviated as SLR (*Simple Linear Regression*) is also one of the statistical methods used in production to make forecasts or predictions about quality and quantity characteristics [5]. Regression analysis has at least 3 uses, namely for the purpose of describing the data phenomenon or case being studied, for control purposes, and for prediction purposes.
Regression is able to describe data phenomena through the formation of a numerical relationship model. Regression can also be used to control a case or things that are being observed through the use of the obtained regression model. In addition, regression models can also be used to make predictions for bound variables [6].

2.6. Mape

There are several calculations and calculations *of errors* or errors. One of them is called *Mean Absolute Percent Error* (MAPE). *Mean Absolute Percent Error* (MAPE) is the average error between the actual value and the forecast over a given period. From the results of the forecast, information will be provided in the form of error presentation.

3. Analysis And Design

3.1. Research Methodology

The method used in this study is a quantitative method with a *time series analysis* approach. The main focus of this study is to predict UD sales. Anugrah Exclusive uses the Linear Regression model. The design process carried out in this study begins by determining the stages of Linear Regression work to obtain more accurate prediction results. In this design, several stages of Linear Regression are used.

3.2. System Requirement Analysis

a. Data Analysis
Data analysis is the process of organizing, interpreting, and infer the information contained in the data that has been collected. The data analysis stage in this study is very important to ensure the accuracy and state of the research results. All stages must be carried out carefully and carefully to ensure that the results obtained can be relied upon and in accordance with the purpose of the research.

b. Testing and Implementation
The next stage is that the data will be tested using test data that has been prepared beforehand. The results of this test will be how accurate Linear Regression is in predicting UD sales. An Exclusive Gift. After the data is quite accurate and can be used to predict UD sales. Anugrah Exclusive well, the next stage is to implement data into the system that will be used to predict UD sales. Anugrah Exclusive automatically.

3.3. Model Implementation

A flowchart is a rare series or process in the form of a diagram. Rarities in a workflow. To make complex processes easier to understand, both for expert and non-expert users, *flowchart* diagrams can be used. In this study, I used two flowcharts, namely linear regression prediction flowcharts and system flowcharts:

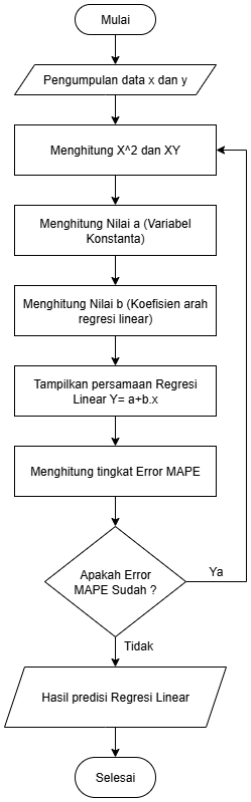

Fig. 1 : Flowchart Prediksi Regresi Linear

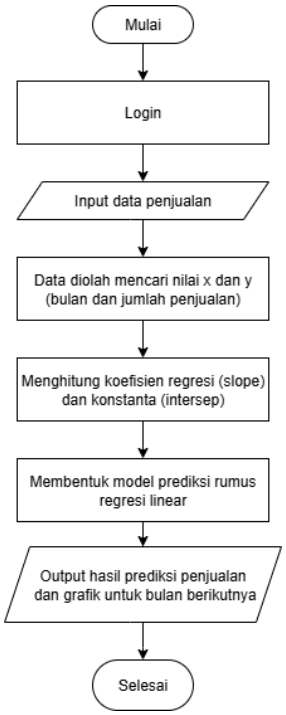

Fig. 2 : Flowchart Sistem

Fig. 1 Description *of the flowchart* :

1. Start: The initial stage of the algorithm. The process starts here.
2. Enter X and Y data: Users are asked to enter data for peanut cakes sold (X) and months (Y).
3. Count X^2 and X: Total X^2 is the sum of each sale that has been multiplied by two, and total XY is the sum of the multiplication of the value of each peanut cake sold by the value according to the time of the month.
4. Calculating the value of a (constant variable): Calculated using a simple linear regression formula.
5. Calculating the value of b (linear regression direction coefficient): Calculated using a simple linear regression formula.
6. Show regression equation: A simple linear regression equation in the form Y = a + b.X is displayed.
7. Calculating MAPE error rate: Calculated using the MAPE formula.
8. Results of the prediction of the Linear Regression algorithm.
9. Done: The final stage of the algorithm. The algorithm process is complete here.

Fig. 2 System Flowchart Description

1. Start: Workflow starts
2. User Accessing Website: User visits website.
3. Login: The user is prompted to log in or regress.
4. User Data Input: The user who successfully logs in enters the seller data on the sales input interface form.
5. Linear Regression Data Processing: Send input data using a linear regression model to predict the sales of peanut cakes sold and the time of the number of them per month.
6. Show Results: Display the results of sales predictions to users through the user interface.
7. Graph View: This view makes it easy for users to see the upward trend of their sales.
8. Done: End of the workflow.

3.3.1. Research Supporting Data

Table 1: UD Sales Research Dataset. Anugrah Exclusive

Year	Moon	Sales/pouch
2021	Januari	319
	Februari	351
	Maret	229
	April	285
	Mei	341
	Juni	425
	Juli	420
	Agustus	337
	September	465
	Oktober	416
	November	389
	Desember	437
2022	Januari	412

	Februari	388
	Maret	416
	April	430
	Mei	404
	Juni	375
	Juli	426
	Agustus	339
	September	354
	Oktober	431
	November	441
	Desember	454
	Januari	439
	Februari	421
	Maret	457
	April	379
	Mei	419
2023	Juni	398
	Juli	429
	Agustus	385
	September	437
	Oktober	423
	November	407
	Desember	428
	Januari	436
	Februari	462
	Maret	419
	April	471
	Mei	439
	Juni	512
2024	Juli	424
	Agustus	431
	September	469
	Oktober	398
	November	412
	Desember	369
	Januari	507
	Februari	495
	Maret	513
2025	April	530
	Mei	522
	Juni	551
Total		**22.666**

The sales data is the result of internal recording from the business and does not include revenue or selling price information per unit. Therefore, this study focuses on the prediction of sales amount based on time (month) as an independent variable using a simple linear regression method.

3.4. Design

A simple design of the Linear Regression Prediction system view where there is a Login view, a Prediction Process Dashboard View, and a Sales Data Input Page.

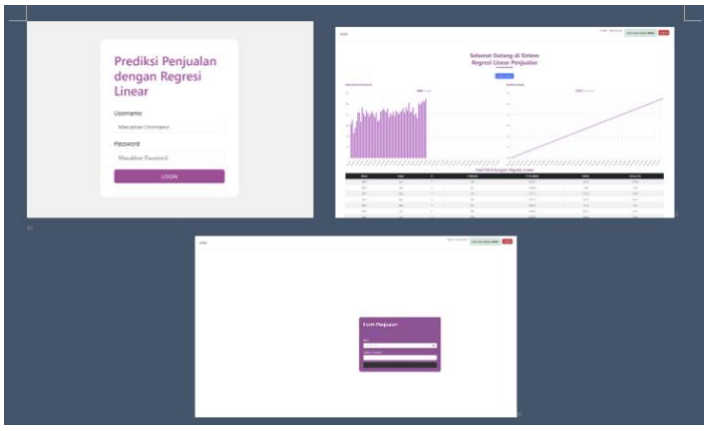

Fig. 3 : Design System

4. Discussion and Implementation

This system is designed to help business owners predict the number of sales every month, so that the process of procuring raw materials and managing stock can be carried out more optimally, efficiently, and with minimal risk of loss. The method used in this study is simple linear regression, which is a statistical method used to model the relationship between one independent variable and one bound variable

(dependent). In the context of this study, the independent variable is time (nth month), while the bound variable is the number of peanut cake sales (in pouch units).
The results of this prediction are very useful in helping business owners:

1. Determine the amount of raw materials that need to be prepared.
2. Avoid overstock or understock.
3. Plan marketing strategies based on sales trends.

With this system, it is hoped that business decision-making can be carried out in a more data-driven and efficient manner.

4.1. Program Test Results

The results of the implementation of the sales prediction system in the UD Home Industry. Exclusive Grace shows that a simple linear regression method can be used effectively to predict the monthly sales amount of peanut cakes.

a. Sales Trend Analysis
From the regression equation obtained:

$$Y=349.55+2.55X$$

It can be seen that a constant value of 349.55 describes the estimated number of sales in the initial period, while the regression coefficient of 2.55 indicates that for every addition of one month (X), sales are predicted to increase by an average of ±9% from the previous month. This indicates that the Home Industry is experiencing a consistent trend of increasing sales in the first half of 2025.

b. Comparison of Actual and Predictive Data
The results of the prediction calculation compared to the actual sales data show a relatively small difference. This can be seen from the average difference between actual data and insignificant predictions, so that the linear regression model is able to capture sales trend patterns well.
Comparisons between actual sales data and predictions using a simple linear regression model show that model performance varies from period to period. At some months, the predicted value is very close to the actual data. For example, in July 2023, actual sales were recorded at 429 pouches, while the predicted result was 428.6 pouches, with an error rate of only 0.09%. This shows that the model is able to capture sales patterns very well in that period.
However, in certain months there is a significant difference. In March 2021 the actual sales were 229 pouch, while the prediction results reached 357.2 pouch, with a relatively high error rate of 55.98%. Likewise in December 2024, where the actual data shows 369 pouches, while the prediction results reach 471.95 pouch, resulting in an error of 27.89%. This difference suggests that a simple linear regression model is not yet able to represent sharp sales fluctuations, which may be influenced by seasonal factors, promotions, or other external conditions.
Although there are some periods with high error rates, in general the model still provides adequate results. This can be seen from the Mean Absolute Percentage Error (MAPE) value obtained at 9.03%, which is included in the very good category (MAPE < 10%). This means that even though there are variations in the error rate in a given month, on average the model is able to provide predictions that are accurate and reliable enough to support decision-making in production planning and marketing strategies.

c. Model Accuracy

The measurement of model accuracy is carried out using Mean Absolute Percentage Error (MAPE), which is an evaluation method that calculates the average percentage of absolute error between the actual value and the predicted value. Based on the results of the calculation of 54 actual sales data and prediction results, a MAPE value of 9.03% was obtained**.**
This value shows that overall, the predictions produced by the simple linear regression model are at an excellent level of accuracy (MAPE < 10%). In other words, the average deviation between the actual value and the predicted results is only about 9% of the actual sales data.
Although the MAPE value is relatively good, there is a significant variation in errors in certain months. For example, in March 2021 the error reached 55.98%, while in July 2023 the error was only 0.09%. This suggests that simple linear regression is superior in capturing long-term trend patterns than predicting monthly fluctuations influenced by external factors such as holidays, seasons, or promotional activities.
The level of accuracy of the linear regression model is sufficient to assist Home Industries in stock and raw material planning, as the prediction errors are still within reasonable limits and do not interfere with strategic decision-making.

d. Implications for the Home Industry

With the results of this prediction, business owners get direct benefits, including:

1. **Raw material planning is more precise** → the amount of materials can be adjusted to the predicted monthly needs.
2. **Stock management is more efficient** → the risk *of overstock* and stock shortages can be minimized.
3. **Planning a marketing strategy** → the owner can organize promotions according to the upward trend of sales.
4. **Data-driven decision-making** → business decisions are no longer speculative, but based on measurable statistical calculations

5. Conclusion and Recommendations

5.1. Conclusion

Based on the results of research and implementation of the sales prediction system in UD Home Industry. Anugrah Exclusive, several conclusions can be drawn as follows:

1. A web-based sales prediction system developed with Python (Flask) and MySQL was successfully implemented to assist business owners in estimating the number of monthly sales of peanut cakes.
2. Linear Regression Method Simple with a free variable (X) in the form of time (nth month) and a bound variable (Y) in the form of sales volume (pouch), it is able to model sales trends from January 2021 to June 2025.

3. The regression equation obtained, namely:

$$Y=349.55+2.55X$$

shows an average increase trend of around 9 pouches per month in the first half of 2025, so it can help in the planning of raw material procurement and sales stock.

4. The evaluation of the model with a Mean Absolute Percentage Error (MAPE) of 9.03% shows an excellent level of accuracy (MAPE < 10%), so that the prediction results can be used as a reference in business decision-making.
5. The system built is able to provide real benefits, namely helping business owners in:
   - Determine the needs of raw materials appropriately,
   - Avoid the risk of *overstock* or lack of stock,
   - Plan marketing strategies based on sales trends.

5.2. Recommendations

In order for this sales prediction system to be developed more optimally, the researcher provides several suggestions:

1. **Development of Prediction Models**
   It can use other methods such as Multiple Linear Regression, ARIMA, or Machine Learning (e.g. Random Forest, LSTM) to get more precise prediction results.
2. **Addition of External Variables**
   Predictions can be improved by adding other variables, such as raw material prices, seasons/celebrations, market demand trends, or promotional strategies that are being carried out.
3. **System Feature Enrichment**
   The system can be added with the visualization of sales trend charts, *auto-reports*, and raw material stock notifications.
4. **Wider Field Tests**
   Further testing is needed on longer data periods and on other Home Industries so that the model can be validated more thoroughly.
5. **System Integration**
   In the future, the system can be integrated with Point of Sales (POS) applications or digital cashier systems, so that sales data can be automatically recorded and predicted without manual input.

References

[1] Muttaqin, Wahyu Wijaya Widiyanto, M. M., Green Ferry Mandias, Stenly Richard Pungus, A. W., Wiranti Kusuma Hapsari, S. A. H., Aslam Fatkhudin,Pasnur, E. F. B., & Mochammad Anshori, Suryani, N. S. (2023). Pengenalan Data Mining (Issue July).

[2] Buaton, R., Anton Sihombing, Fuji Dodo Aritonang, & Clara Rosa Wijaya. (2017). Data Mining Untuk Menentukan Korelasi (Confidence Dan Support)Jurusan Siswa Pada Tingkat Sekolah Menengah Terhadap Indeks Prestasi Kumulatif (Ipk) Di Perguruan Tinggi Sebagai Solusi Tepat Pemilihan Program Studi Di Perguruan Tinggi. Jurnal Sistem Informasi Kaputama (JSIK), 1(2), 1–13. https://doi.org/10.59697/jsik.v1i2.744

[3] Guntur, M., Santony, J., & Yuhandri, Y. (2018). Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 2(1), 354-360. https://doi.org/10.29207/resti.v2i1.276

[4] Anggraini, S. D., & Nurcahyo, G. W. (2021). Prediksi Peningkatan Jumlah Pelanggan dengan Simulasi Monte Carlo. Jurnal Informatika Ekonomi Bisnis, 3, 95–100. https://doi.org/10.37034/infeb.v3i3.92

[5] Suhandi, N., Putri, E. A. K., & Agnisa, S. (2018). Analisis Pengaruh Jumlah Penduduk terhadap Jumlah Kemiskinan Menggunakan Metode Regresi Linear di Kota Palembang. Jurnal Ilmiah Informatika Global, 9(2), 77–82. https://doi.org/10.36982/jiig.v9i2.543

[6] Setyoningrum, N. R., Rahimma, P. J., Teknologi, S. T., Tanjungpinang, I., & Tanjungpinang, K. (2022). Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang. Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK), 4, 13–18. https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/5200