Literasi Nusantara
vol.6 no. 1, February 2026
https://journal.citradharma.org/index.php/literasinusantara/index

ISSN 2746 8208 (Online)
ISSN 2746 3575 (Print)

Analysis of Sales Prediction Using Linear Regression as a Data Analytics Learning Media for Students

Agus Susilo Nugroho[1]*, Tyasmiarni Citrawati[2] , Agung Setyawan[3]

[1] Bisnis Digital, UIN KH. Abdurrahman Wahid, Pekalongan, Indonesia
[2,3] Universitas Trunodjoyo Madura,Jawa Timur, Indonesia

* agus.susilo.nugroho@uingusdur.ac.id

DOI : https://doi.org/10.56480/jln.v6i1.34

Received: September 24, 2025
Revised: November 16, 2025
Accepted: January 20, 2026

Abstract

This research aims to analyze the application of the linear regression method in sales prediction and evaluate its use as a data analytics learning media for students. The method used is a quantitative descriptive approach with two main stages: sales data analysis using linear regression and collecting student perception data through Likert scale-based questionnaires. The dataset used consists of simple sales data that is easy for students to process and understand. The analysis results show that the linear regression method can be used to predict sales effectively and provides a clear overview of the relationships between variables. Furthermore, the questionnaire results indicate that students responded positively to the use of real datasets in learning, particularly in improving material comprehension, engagement in the learning process, and ease of data analysis. Thus, the use of linear regression with real datasets can serve as an effective alternative learning media to enhance student data literacy. This research is expected to be a reference for developing data-based learning methods in the field of data analytics.

Keywords– *Linear Regression, Sales Prediction, Data Analytics, Learning, Real Dataset.*

Literasi Nusantara. vol.6 no. 1, February 2026
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

1. Introduction

The development of information technology has increased the demand for data processing and analysis skills across various fields, including higher education. Data analytics has become a vital competency for students to effectively understand, process, and draw conclusions from data. Several studies indicate that data literacy plays a crucial role in improving the quality of learning and enhancing students' critical thinking skills when facing data-driven problems.

One of the most commonly used methods in data analysis is linear regression. This method is utilized to determine the relationship between independent and dependent variables and to predict specific values. Sholeh et al. (2023) demonstrated that linear regression can be effectively used to predict student exam scores. This aligns with research by Retnowati (2026), which states that regression models are capable of providing fairly accurate predictions in an educational context. Furthermore, Qoiriah and Yamasari (2021) emphasize that linear regression is an easily understood method for students during their initial stages of learning data analytics.

However, the process of teaching data analytics in higher education still faces various obstacles. Sa'adah (2024) revealed that the use of non-contextual datasets causes students to experience difficulty in understanding the real-world application of concepts. This is reinforced by Muthahharah and Fatwa (2022), who state that irrelevant learning media can decrease student engagement in the learning process. Additionally, Hartianto and Erikson (2022) found that low student engagement results in suboptimal comprehension of the learning material.

The use of real-world datasets in education serves as a solution that can enhance the quality of the learning process. Research shows that utilizing real data provides a more contextual and applicable learning experience, making it easier for students to understand the connection between data and analysis

results. Consequently, students can become more actively involved in the learning process and gain a better grasp of data analytics concepts.

Utilizing simple methods like linear regression combined with real datasets presents a potential approach to data analytics education. Yordan et al. (2019) showed that linear regression can be successfully applied to simple, easy-to-understand data. Moreover, various studies in data-based learning indicate that a hands-on practical approach can increase learning effectiveness and students' analytical abilities.

Based on these issues, this research aims to apply the linear regression method for sales prediction and evaluate its utilization as a data analytics learning media for students. Furthermore, this study analyzes student perceptions regarding the use of real datasets in the learning process through the distribution of questionnaires. It is hoped that the results of this research will contribute to the development of more effective data-based learning methods, particularly in enhancing data literacy, engagement, and student comprehension in the field of data analytics.

2. Method

This research employs a quantitative descriptive approach to analyze the application of the linear regression method in sales prediction and to evaluate its utilization as a data analytics learning media for students. This approach was selected because it is capable of describing data analysis results numerically and measuring student perceptions systematically (Aryani, 2020).

The data used in this study consists of secondary and primary data. The secondary data is a sales dataset obtained from open sources, which is used for analysis and prediction through the simple linear regression method. The variables involved include the independent variable (time period) and the dependent variable (sales volume). Linear regression analysis is used to determine the relationship between variables and to generate a prediction model.

Literasi Nusantara. vol.6 no. 1, February 2026
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

Meanwhile, primary data was obtained by distributing questionnaires to students who had participated in data analytics learning using the linear regression approach and real-world datasets. The questionnaire was designed using a Likert scale ranging from 1 to 5 to measure several aspects: learning interest, material comprehension, learning engagement, ease of data analysis, and learning experience. The use of the Likert scale in quantitative research is widely adopted to measure respondent perceptions in a structured manner (Rahmadani & Tanti, 2025).

The questionnaire data is analyzed by calculating the mean value of each question to determine the level of student perception. The average results are then interpreted into assessment categories: low, fair, good, and very good. This analysis aims to determine the effectiveness of using linear regression and real datasets as learning media. Such a descriptive analysis approach is commonly used in quantitative research to numerically describe the distribution of respondent perceptions (Sipahutar et al., 2022).

Furthermore, to ensure the quality of the analysis results, a validation process was conducted on the linear regression model. Validation was performed by observing the Mean Absolute Error (MAE), which is an evaluation metric used to measure the prediction error rate in regression models. MAE calculates the average absolute difference between actual and predicted values regardless of the error direction (positive or negative). Thus, MAE provides an overview of the average prediction deviation from the actual values in the same units as the original data. Subsequent evaluation was conducted by calculating the coefficient of determination (R2) to determine the model's ability to explain variations in sales data. A higher R2 value indicates a better representation of the relationship between the independent and dependent variables. This process ensures that the model used in learning has adequate accuracy and serves as a relevant example for students in understanding data prediction concepts.

On the other hand, the questionnaire instrument used in this study also considers aspects of clarity and ease of understanding for respondents. Before distribution, the questionnaire underwent a simple review to ensure that each question could be well understood by students. This is crucial so that the obtained data truly reflects student perceptions toward the use of real datasets and linear regression methods in learning. Consequently, the research results are expected to have a high level of reliability and provide a more accurate picture of the effectiveness of the applied learning approach.

The research stages were conducted systematically, starting from sales dataset collection, data processing using linear regression, analysis of prediction results, preparation and distribution of questionnaires, to questionnaire data processing and drawing conclusions. The flow of these research stages can be seen in Figure 1 below.

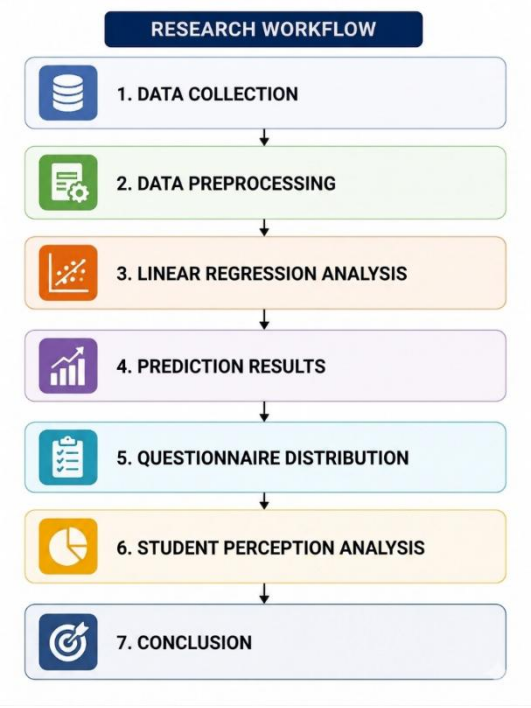

Figure 1. Research Workflow

With this approach, the research is expected to provide a clear overview of the application of the linear regression method within the context of data analytics learning.

Literasi Nusantara. vol.6 no. 1, February 2026
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

3. Result and Discussion

Result

The data analysis was conducted by applying a simple linear regression method to a sales dataset collected from open sources. Based on the data processing results, a linear regression model was obtained, showing a relationship between the time variable as the independent variable and the sales volume as the dependent variable. The resulting model is capable of describing general sales trend patterns, where an increase in the time variable value is followed by changes in the sales value. The visualization of the regression line between time and sales can be seen in figure 2 below.

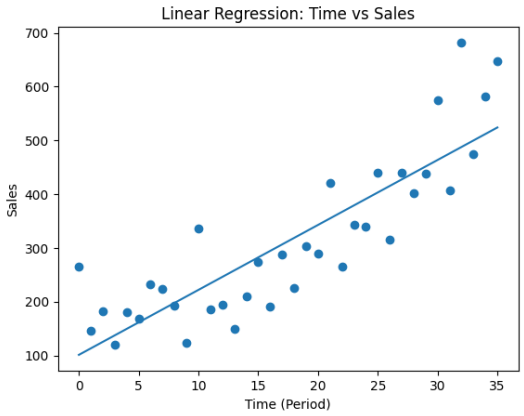

Figure 2. Regression line

Furthermore, the prediction results obtained from the regression model show values that are not significantly different from the actual data; thus, it can be concluded that the model has an adequate level of accuracy for use as a learning media. A comparison between the actual data and the predicted data can be seen in Figure 3.

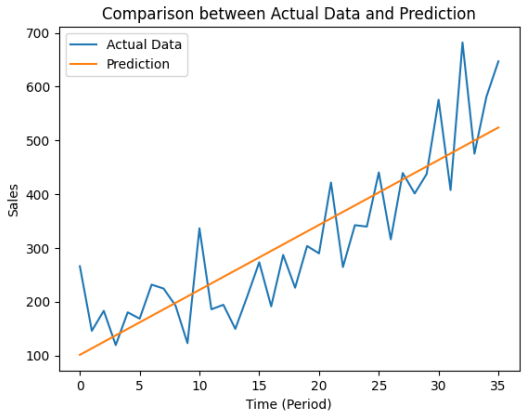

Figure 3. Comparison between the actual data and the predicted data

The comparison between actual data and prediction results indicates that the linear regression model is able to follow the sales data pattern quite well, although there are some discrepancies at certain points.

Based on the evaluation of the linear regression model, a coefficient of determination R2 of 0.73 was obtained. This value indicates that the model can explain approximately 73% of the variation in sales data based on the time variable. This suggests a reasonably strong relationship between time progression and sales volume. Meanwhile, the Mean Absolute Error (MAE) value of 62.13 shows that the average difference between the predicted and actual values is 62.13 sales units. This value is still categorized as quite good, considering the model used is a simple linear regression with a single independent variable. The results of this model evaluation can be seen in Figure 4.

```
Result of R²: 0.7301028704508052
Result of MAE: 62.13656799656798
```

Figure 4. Results of evaluation model

Based on the questionnaire results, the average student rating for learning interest was 4.3 (Good category). Material comprehension obtained a score of 4.3 (Good), learning engagement was 4.4 (Good), ease of data analysis was 4.3 (Good), and learning experience was 4.6 (Very Good). The average results of student perceptions are also presented in the form of a visualization diagram in Figure 5.

Literasi Nusantara. vol.6 no. 1, February 2026
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

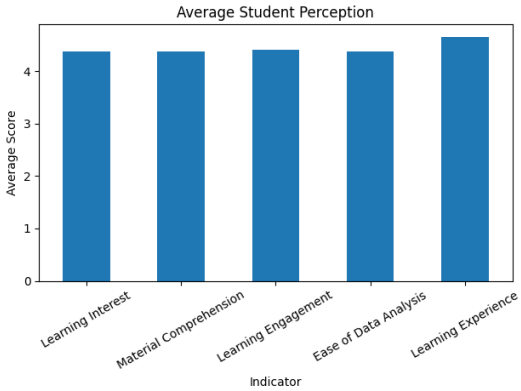

Figure 5. The questionnaire

Discussion

From the perspective of learning evaluation, the questionnaire results show that the majority of students provided positive responses toward the use of linear regression and real datasets in the learning process. The average score for each indicator falls within the "Good" to "Very Good" categories. Students assessed that the use of real data helped them understand data analysis concepts more concretely compared to using only theoretical examples.

The aspect of student engagement in learning also showed positive results. Students tended to be more active and interested when using datasets derived from real-world conditions, as they could directly see the relationship between variables and the resulting prediction outcomes. This indicates that a real-data-based learning approach is capable of increasing student engagement in the learning process. In terms of analysis ease, students also gave a good assessment of the dataset used in this study. The simple and structured dataset made it easier for students to perform the analysis process, from preprocessing to the interpretation of results. This is crucial in the early stages of data analytics learning so that students are not overwhelmed by excessive data complexity.

The average perception results indicate that using real datasets in data analytics learning can enhance student comprehension and engagement. Students also evaluated that the data analysis process became easier and provided a positive learning experience.

Additionally, the linear regression model evaluation was conducted using the coefficient of determination R2 and Mean Absolute Error (MAE). The R2 value shows that the model has a fairly good ability to relate the progression of time (period) and the sales volume, which indicates a specific trend. The higher the R2 value, the greater the proportion of data variation that can be explained by the model. Meanwhile, MAE was used to measure the average absolute error between predicted and actual values. The calculation results show that the MAE value is relatively small, indicating that the difference between prediction results and actual data is not too large. Therefore, the linear regression model used in this study can be considered to have a sufficiently good level of accuracy for use as a basic learning media in data analytics.

The results of this study show that the application of simple linear regression using real datasets is not only effective in producing reasonably accurate prediction models but also has a positive impact on the data analytics learning process. This approach is able to enhance student comprehension, engagement, and learning experiences in a more applicable and contextual manner.

4. Conclusion

This research demonstrates that the application of the simple linear regression method on time-based sales datasets is capable of producing a reasonably good prediction model, as evidenced by a relatively high coefficient of determination R2 and a low Mean Absolute Error (MAE). These results indicate that the developed model possesses adequate capability in representing the relationship between the time variable and sales volume.

The use of real-world datasets in the learning process provides a more contextual experience for students. Based on the questionnaire results, students showed a positive response toward this approach in terms of conceptual understanding, learning engagement, and ease of data analysis. This suggests that integrating the linear regression method with real datasets can enhance the effectiveness of data analytics learning.

Literasi Nusantara. vol.6 no. 1, February 2026
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

In conclusion, this study contributes to the development of data-based learning methods that are simple yet applicable. For future research, it is recommended to utilize more complex datasets and compare them with other machine learning methods to achieve more optimal prediction results and further enrich the student learning experience.

References

Aryani, Y. (2020). Sistem Informasi Penjualan Barang dengan Metode Regresi LInear Berganda dalam Prediksi Pendapatan Perusahaan. *Jurnal Riset Sistem Informasi dan Teknologi Informasi (JURSISTEKNI), 2*(2), 39-51. https://doi.org/10.52005/jursistekni.v2i2.47

Hartianto, D., & Erikson, H. (2022). Optimasi Prediksi Kelulusan Mahasiswa dengan Menggunakan C5.0 dan Regresi Linear. *Jurnal Sistem Informasi, Teknik Informatika dan Teknologi Pendidikan, 2*(1), 22-32. https://doi.org/10.55338/justikpen.v2i1.36

Muthahharah, I., & Fatwa, I. (2022). Analisis Regresi Linear Berganda untuk Media Pembelajaran Daring Terhadap Prestasi Belajar Mahasiswa di STKIP Pembangunan. *Jurnal Matematika dan Statistika serta Aplikasinya, 10*(1), 53-60. https://doi.org/10.24252/msa.v10i1.25145

Qoiriah, A., & Yamasari, Y. (2021). Prediksi Nilai Akhir Mahasiswa dengan Metode Regresi. *Journal of Information Engineering and Educational Technology (JIEET), 5*(1), 40-43. https://doi.org/10.26740/jieet.v5n1.p40-43

Rahmadani, A., & Tanti, L. (2025). Estimasi Penjualan Sembako Menggunakan Teknik Data Mining Regresi Linear Berganda. *Jurnal Sistem Informasi dan Komunikasi (JSON), 7*(2), 645-656. https://doi.org/10.30865/json.v7i2.9171

Retnowati, S. (2026). Model Regresi untuk Prediksi Nilai Akhir Mahasiswa pada Mata Kuliah Introduction to Computer and ICT. *Stardia Journal, 2*(1), 13-24.

Sa'adah, N. (2024). Penerapan Metode Regresi Linear pada Data Pembelajaran Online, 1(4), 559-569. https://doi.org/10.61722/jmia.v1i4.2297

Sholeh, M., Nurnawati, E. , & Lestari, U. (2023). Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner. *JISKA (Jurnal Informatika Sunan Kalijaga), 8*(1), 10-21. https://doi.org/10.14421/jiska.2023.8.1.10-21

Sipahutar, A., Munthe, I., & Juledi, A. (2024). Sales Trend Analysis with Machine Learning Linear Regression Algorithm Method. *Jurnal dan Penelitian Teknik Informatika, 8*(3), 1724-1728. https://doi.org/10.33395/sinkron.v8i3.13809

Yordan, A., Putri, T. N., & Lamkaruna, D. H. (2019). Peramalan Penerimaan Mahasiswa Baru Universitas Samudra Menggunakan Metode Regresi Linear Sederhana. *Jurnal Teknik Informatika (J-Tifa), 2*(1), 21–27. https://doi.org/10.52046/j-tifa.v2i1.237