Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

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

Student Admission Prediction System Using Multiple Linear Regression

M. Irsyad Ibadillah[1]*, Purnomo Hadi Susilo[2], Musta'in[3]

[1,2,3] Universitas Islam Lamongan
mirsyadibadillah57@gmail.com[1*], purnomo@unisla.ac.id[2] , mustain@unisla.ac.id[3]

Abstract
New student admission is an important aspect of educational institution operations. MI Hidayatussibyan Glugu has experienced difficulties in predicting the number of applicants, which has led to inaccurate planning of facilities, teachers, and infrastructure. This study aims to develop a prediction system for new student admissions using the Multiple Linear Regression method, analyzing the relationship between the number of previous students (X1), the number of surrounding residents (X2), and school facilities (X3) on the number of applicants (Y).
This research applied a quantitative approach with a case study at MI Hidayatussibyan Glugu. Ten years of historical data were analyzed using multiple linear regression. The system was developed using PHP and MySQL with a web based interface. The model accuracy was tested using the Mean Absolute Percentage Error (MAPE), and the results showed the value was in the "good" category (below 20%), ensuring the system's feasibility in predicting student admissions.
The results indicate that the developed system can consistently estimate the number of new students based on historical data. The evaluation using MAPE values below 20% confirmed that the prediction accuracy was in the good category. This shows that the system can be utilized to support schools in planning student admissions more accurately and effectively.

***Keywords**: New Student Prediction, Multiple Linear Regression, MAPE*

1. Introduction

The admission of new students is one of the crucial activities that influences the sustainability of educational institutions, including MI Hidayatussibyan Glugu, located in Dlanggu Village, Deket District, Lamongan Regency. In recent years, the school has faced challenges in predicting the number of applicants and newly admitted students. The large number of applicants each year often complicates the planning process, particularly in preparing educational quality improvement programs in a more systematic and well-structured manner. The main problem faced is the absence of a system that can accurately and data-drivenly predict the number of new students. As a result, school planning tends to be reactive and unstructured, often leading to shortages or surpluses in classrooms, facilities, and teaching staff. This condition directly affects the quality of educational services and the overall operational effectiveness of the school. To address this challenge, the proposed solution in this study is to design and develop a prediction system for new student admissions using the Linear Regression algorithm. This algorithm is selected because it can analyze the relationship between variables that influence student enrollment, such as historical admission data, school facilities, and local demographic data.

Research has shown that the number of newly admitted students each year can significantly impact academic operations in educational institutions. One important factor in planning is predicting the number of students for the following year, as it is closely related to lecturer availability, classroom allocation, and supporting facilities [1].

In this regard, the objective of this study is to design a prediction system for student admissions at MI Hidayatussibyan Glugu using the Linear Regression algorithm. This system is intended to serve as a reliable decision-support tool for planning and management, improving operational efficiency, maintaining educational quality, and providing greater benefits to the community. Furthermore, several studies have demonstrated the effectiveness of this method. For example, one study applied data mining with multiple linear regression to predict the number of new students in 2023, yielding a forecast of 38 students. The prediction was tested using RapidMiner 5.0, and the results confirmed the same figure [2].

This research is also expected to serve as a starting point for the application of data mining technology in the education sector, particularly in madrasahs. By adopting a technology-driven approach, MI Hidayatussibyan Glugu can demonstrate its commitment to continuous adaptation and innovation in improving educational services. The system's implementation is also expected to inspire other madrasahs to adopt similar technologies as a means of addressing various challenges in educational management in the modern era. Previous studies supporting this work include research on forecasting new student admissions at Universitas Samudra using simple linear regression. That study aimed to predict new admissions by modeling the dependent variable (Y), representing the number of students, against the independent variable (X), based on data from 2014 to 2018 [3].

Previous studies that serve as references include research entitled *"New Student Admission Forecasting System Using a Simple Linear Regression Method Based on a Website."* This study aimed to build a web-based forecasting system for predicting new student admissions using the simple linear regression method. Historical admission data was utilized to estimate the number of applicants in the future. A case study was conducted at SD Negeri 3 Sembung, with testing performed using admission data from several previous years. The findings showed that the simple linear regression method was able to provide reasonably accurate estimates, thereby supporting school capacity planning and promotional strategies efficiently [4].

Another relevant study is *"Prediction of the Number of New Students Using the Simple Linear Regression Method."* This research focused on predicting the number of new students at a university in Surabaya using simple linear regression. The independent variable in this study was the academic year (X), while the dependent variable was the number of new students (Y). Data from the Faculty of Science and Technology for the period 2016–2020 was used. Predictions were then evaluated using the Mean Absolute Percentage Error (MAPE) to measure accuracy. The results indicated a relatively high prediction accuracy for each program, with the trend graph showing a linear decline in subsequent periods [5].

Previous studies employing different methods have demonstrated that the K-NN algorithm is effective in classifying complaint services at SMK NU 2 Kedungpring Lamongan, achieving a high level of accuracy and thereby supporting data-driven decision-making processes. This provides evidence that data-based algorithms can enhance information management efficiency and support predictive systems [6]. Another study, using a different approach to predict passenger numbers at Juanda Airport through the Double Exponential Smoothing method, indicated that a prediction system based on historical data can assist in planning aircraft allocation and schedules with high accuracy. This confirms that the use of data-driven prediction methods can improve operational planning, aligning with the objectives of this study in designing a student admission prediction system [7].

The present study focuses on the application of the Linear Regression algorithm to predict the number of new student admissions at MI Hidayatussibyan Glugu. The main issues addressed are how to apply linear regression for prediction and the extent of error rates generated based on historical data and supporting variables. To maintain research focus, the data used is limited to new student admissions over the past ten years, with independent variables consisting of the local population, school facilities, and previous year's student data. Testing is limited to the school's internal data, excluding other institutions, and the developed system is specifically designed to generate predictions of new student admissions without additional management features.

The objectives of this research are to apply the linear regression algorithm to produce more accurate predictions of new student admissions and to analyze the accuracy of the method by considering historical data and relevant supporting variables. The expected benefits include providing deeper insights into the application of linear regression in education, offering a technology-based solution to assist schools in planning for new student admissions, and contributing scientifically for future researchers interested in developing data mining-based prediction systems in similar educational contexts.

2. Research Methodology

The research method employed in this study is a quantitative approach with a case study focused on MI Hidayatussibyan Glugu. The data analyzed includes ten years of historical student admission records, demographic data from the surrounding community, and school facility data. The analysis applies the Linear Regression method to produce an accurate prediction model.

Quantitative research is one of the most widely used scientific methods, designed to systematically, objectively, and rigorously uncover knowledge while minimizing errors through testing. This approach emphasizes causal relationships, variable interactions, hypotheses, and specific research questions. It involves measurement, observation, and theory testing, and is commonly carried out through surveys or experiments. The results are then analyzed using statistical methods to derive meaningful conclusions.

2.1 Research Flow

Before conducting the study, a research flow is required to facilitate the research process. The research flow represents systematic stages or steps carried out to achieve the predetermined objectives. This flow provides a structured guideline for conducting the study, starting from the initial stage of problem identification to the final stage of implementation. The research flow used in this study is illustrated in Figure 1:

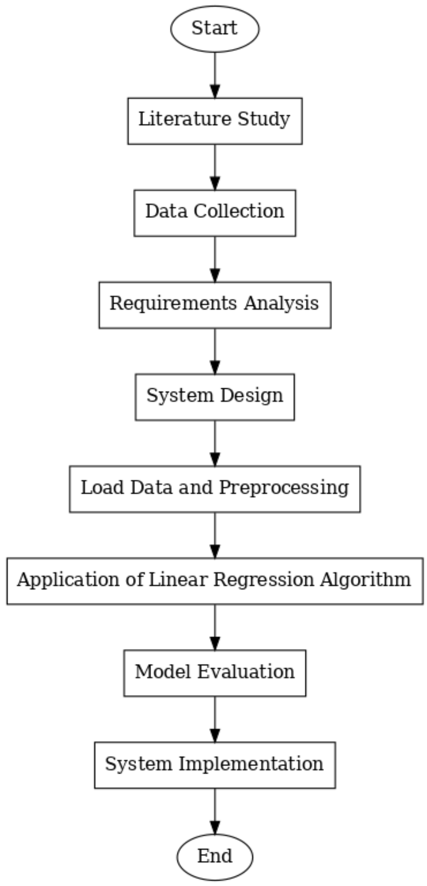


Fig. 1 : Research Flow

2.2 Data Collection

At this stage, the data required for the study is collected from historical records of new student admissions at MI Hidayatussibyan Glugu. The data includes the number of registrants, population data, school facilities, and student records from the past ten years.

2.3 System Design

At the system design stage, the architecture of the application is planned. This includes designing the database, user interface, and other required components that will support the prediction system.

2.4 Application of Linear Regression Algorithm

In this phase, the multiple linear regression algorithm is applied to the preprocessed dataset. The algorithm is used to build a predictive model that estimates the number of new student admissions based on the independent variables.

2.5 System Implementation

In the final stage, the predictive model is implemented into the system. The system allows users, particularly school administrators, to input relevant variables and obtain predictions of new student admissions, making the decision-making process more efficient.

3. Result and Discussion

3.1 Data Collection

The dataset used in this study is secondary data consisting of student admission records from the 2015/2016 to 2024/2025 academic years, total student data over the last ten years, school facility data, and population data of Dlanggu Village based on official reports. The summary of total students for the last ten years can be seen in Table 1:

Table. 1: Total Number of Grade 1–6 Students (Last 10 Years)

Tahun	Total
2015	234
2016	246
2017	237
2018	248
2019	252
2020	245
2021	226
2022	224
2023	213
2024	217

This table presents the total number of students from grades 1 to 6 over the past ten years, which is used as variable X1 in the multiple linear regression model. The data illustrates the trend of student growth at MI Hidayatussibyan Glugu and serves as an indicator of school stability and popularity. The fluctuations observed each year may reflect the influence of external factors such as demographics or competition with other schools. The summary of new student admissions over the last ten years is shown in Table 2:

Table. 2: Historical New Student Enrollment (Last 10 Years)

Tahun	Junlah Siswa Baru
2015	52
2016	45
2017	39
2018	42
2019	42
2020	33
2021	32
2022	45
2023	34
2024	45

This table presents the number of new student applicants per academic year over the past ten years, which is used as the dependent variable (Y) in the prediction model. The purpose is to forecast future values of Y based on the independent variables (X1, X2, X3). The population data is presented in Table 3:

Table. 3: Dlanggu Village Population Data

Tahun	Total
2015	4.815
2016	4.775
2017	4.756
2018	4.758
2019	4.705
2020	4.713
2021	4.713
2022	4.678
2023	4.541
2024	4.355

This table presents the population of Dlanggu Village from 2015 to 2024, which is used as variable X2. The data is significant because population size reflects the potential number of new students. Trends in population growth or decline also provide insights for designing more effective school promotion strategies.

3.2 System Design

The system design uses a Use Case diagram as a modeling tool to describe the planned software behavior. This diagram illustrates the interactive relationship between the actors and the system. The main actor is the administrator, who has five primary use cases: managing student data, community data, facilities, performing predictions, and evaluating the model. These use cases represent the core functions of the system in supporting decision-making. The overall system process is illustrated in the Use Case diagram shown in Figure 2.

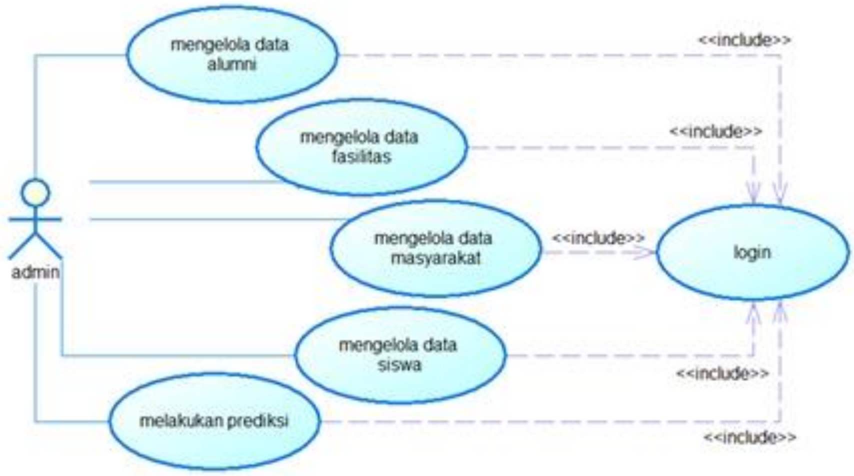


Fig. 2: Use Case

3.3 Application of Linear Regression Algorithm

Several calculations used to build the multiple linear regression model are based on the research data. The dataset consists of three independent variables and one dependent variable. The analysis follows the multiple linear regression method, where each variable is examined to determine its contribution to the prediction results, both simultaneously and partially. The outcomes of these calculations are then used to form a regression equation that serves as the foundation for the prediction system. This equation is expected to accurately represent the relationship between variables and support better decision-making. The variables X and Y are presented in Table 4.

Table. 4: Independent Variables (X) and Dependent Variable (Y)

Tahun	X_1 (jumlah siswa)	X_2 (jumlah masyarakat)	X_3 (fasilitas)	Y (siswa baru)
2015	234	4815	81	52
2016	246	4775	81	45
2017	237	4756	82	39
2018	248	4758	82	42
2019	252	4705	82	42
2020	245	4713	82	33
2021	226	4713	95	32
2022	224	4678	95	45
2023	213	4541	95	34
2024	217	4355	40	45
jumlah	2342	46809	815	409

The first step is to calculate the sum of each independent variable (X) and the dependent variable (Y). This summation provides the total values needed for subsequent calculations, such as averages, regression coefficients, and the formation of the multiple linear regression matrix. Next, an auxiliary table is created by multiplying variables (X1*Y, X2*Y, X3*Y, X1*X2, X1*X3, X2*X3), followed by summing the results. This table serves as the basis for manually computing the main components of the regression coefficients. The results of this step are shown in Table 5

Table. 5: Supporting Calculation Table

No	X1Y	X2Y	X3Y	X1X2	X1X3	X2X3
1	12168	250380	4212	1126710	18954	390015
2	11070	214875	3645	1174650	19926	386775
3	9243	185484	3189	1127172	19434	389992
4	10416	199836	3444	1179984	20336	390156
5	10584	197610	3444	1185660	20664	385810
6	8085	155529	2706	1154685	20090	386466
7	7232	150816	3040	1065138	21470	447735
8	10080	210510	4275	1047872	21280	444410
9	7242	154394	3230	967233	20235	431395
10	9765	195975	1800	945035	8680	174200
∑	95885	1915409	32994	10974139	191069	3826954

The next step is squaring each variable ($X1^2$, $X2^2$, $X3^2$) along with all related data in the table. This process generates squared values that are essential for constructing the regression matrix and calculating regression coefficients more accurately. Squaring also helps measure the contribution of each independent variable to the dependent variable in the statistical analysis. The results are presented in Table 6:

Table. 6: Tabel Auxiliary Table of Squared Values

No	X_1^2	X_2^2	X_3^2
1	54756	23184225	6561
2	60516	22800625	6561
3	56169	22619536	6724
4	61504	22638564	6724
5	63504	22137075	6724
6	60025	22212369	6724
7	51076	22212369	9025
8	50176	21883684	9025
9	45369	20620681	9025
10	47089	18966025	1600
∑	550184	219275103	68693

Next, all calculated values from the auxiliary table are inserted into matrices A, H, A1, A2, A3, and A4. Matrix A contains the basic computation results of variable combinations, while matrix H integrates independent variables (X1, X2, X3) and the dependent variable (Y) in a square form for determinant calculation. Matrices A1 to A4 are formed by replacing specific columns in matrix H with column Y to compute regression coefficients. This process isolates the influence of each independent variable on Y and serves as the foundation for constructing an accurate multiple linear regression equation.

10	2342	46809	815
2342	550184	10974139	191069
46809	10974139	219275103	3826954
815	191069	3826954	68693

Matrix H is the main matrix in the determinant method (Cramer's rule), constructed based on the summation (∑) of each variable and the product of the variables.

∑Y	∑X1*Y	∑X2*Y	∑X3*Y
409	95885	1915409	32994

Matrix A is the primary matrix in the multiple linear regression method, containing total values (summations) and the products of independent variables, such as the totals of X1, X2, X3, the squares of each variable, and the cross-products between two variables.

3.4 Design Interface

Interface design is the process of creating a user interface (UI) in a system or application to ensure that users can interact with it effectively, efficiently, and pleasantly. The interface is designed to be simple yet informative, making it easy to use for users at various levels. The design also emphasizes consistency in colors, icons, and navigation to enhance the overall user experience. The interface design of the system being developed is presented in Figure 3 below.

1. Login Page

Fig. 3: Login Page

Users must first log in by entering their registered username and password. The login page serves as the entry point to the prediction system, ensuring that only authorized users can access its features. Once successfully logged in, users are directed to the main page to perform the prediction process.

2. Register Page

The registration page is used by new users to create an account before accessing the system. This form serves as the initial step that allows users to log in and access the prediction features. The registration process ensures that each account is securely recorded in the database, allowing only verified users to utilize the system. This design is illustrated in Figure 4.

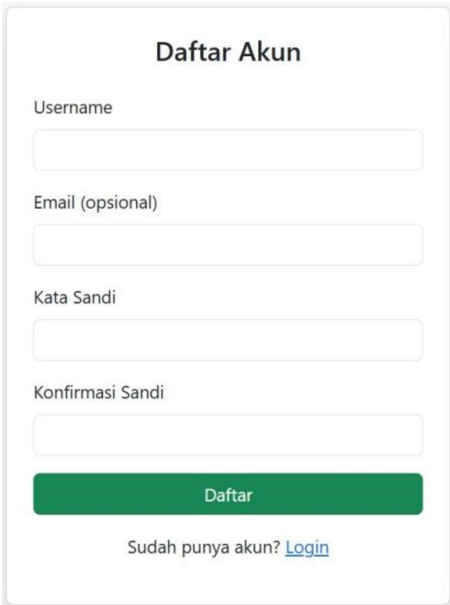


Fig. 4: Register Page

The registration page allows new users to create an account before accessing the prediction system. Users are required to provide a username, email, password, and password confirmation. The system validates the data and stores the account in the database if it is valid. Once successfully registered, users can log in to access the prediction features.

3. Home Page

The homepage displays a summary of the main features, including access to the prediction menu and training data, as shown in Figure 5.

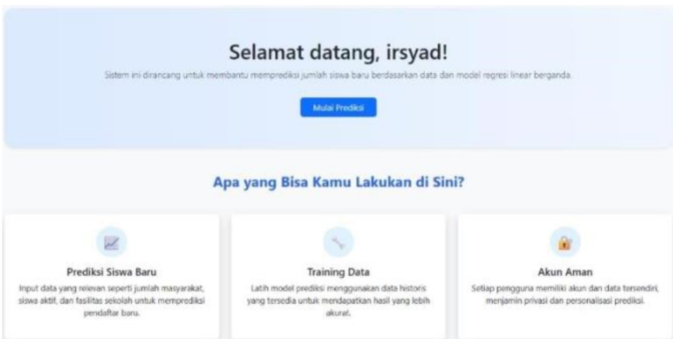


Fig. 5: Home Page

The homepage is displayed after the user successfully logs in and serves as the main control center of the system. This page greets the user, displays the active account name, and provides access to key features such as prediction and training data. The interface is designed to be simple and responsive, ensuring ease of use across different devices. Navigation buttons are provided to facilitate movement between features, while the page also emphasizes the active user's identity to maintain session security.

4. Prediction Page

The prediction page is used to input variable data and display the calculation results of the number of new students based on the multiple linear regression model. This can be seen in Figure 6.

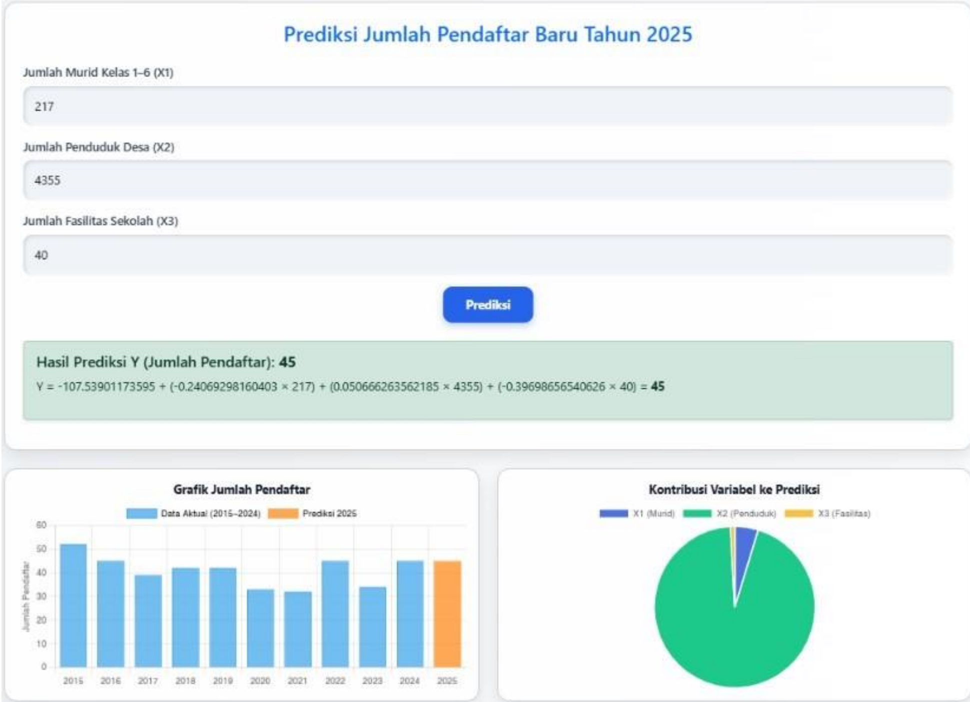
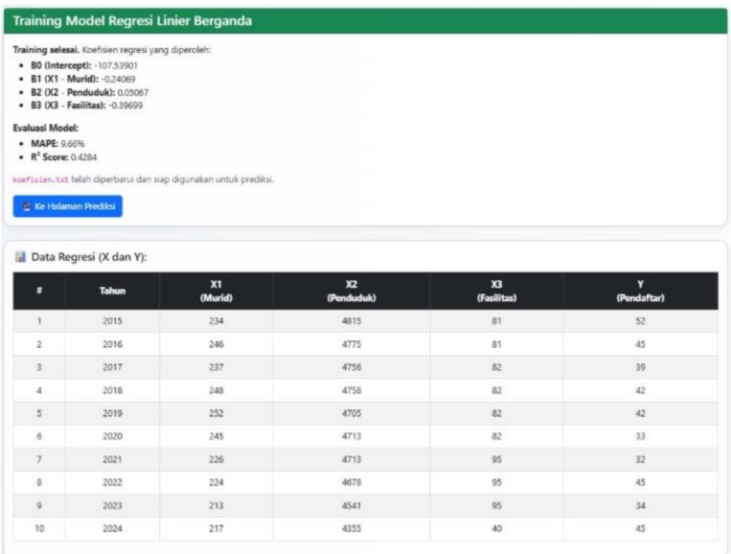


Fig. 6: Prediction Page

Users can input the values of variables X1 (number of students in grades 1–6), X2 (village population), and X3 (number of school facilities) into the system to automatically obtain predictions of new student enrollment based on the established multiple linear regression model. The system processes the input data using mathematical calculations, and the prediction results are displayed immediately after the form is submitted. The calculations are performed using regression coefficients previously obtained from the training process. These predictions can serve as a reference for planning future student admissions.

5. Training Data Page

The training data page is used to load and manage historical data that will be utilized in building the prediction model, as shown in Figure 7.



#	Tahun	X1 (Murid)	X2 (Penduduk)	X3 (Fasilitas)	Y (Pendaftar)
1	2015	234	4815	81	52
2	2016	246	4775	81	45
3	2017	237	4756	82	39
4	2018	248	4758	82	42
5	2019	252	4705	82	42
6	2020	245	4713	82	33
7	2021	226	4713	95	32
8	2022	224	4678	95	45
9	2023	213	4541	95	34
10	2024	217	4355	40	45

Fig. 7: Training Data Page

The training data page is used to train the multiple linear regression model based on the available historical data. The system calculates the regression coefficients (B0, B1, B2, B3) through mathematical processes and stores them in a file for future predictions. Once the training process is completed, the system displays the regression coefficients, model accuracy metrics such as MAPE and R^2 Score, as well as the dataset used. This page helps users understand that the system has successfully learned from the data and is ready for prediction. The process is essential to ensure that the model is built on reliable calculations and can be trusted. The interface is designed to be simple and informative, making it easily accessible for users at any time.

4. Conclusion

Based on the results of the study and the implementation of the new student admission prediction system at MI Hidayatussibyan Glugu using the multiple linear regression algorithm, it can be concluded that the system was successfully developed as a web-based application built with PHP and MySQL. The system includes features such as login, registration, and a user-friendly dashboard, which allow the school to easily access and manage prediction data. By utilizing historical data, the system automatically generates a multiple linear regression model to estimate the number of new students based on three independent variables: the number of students in grades 1–6 (X1), the village population (X2), and the number of school facilities (X3). The prediction results are directly displayed through the system interface, making it a practical and reliable tool to support planning for future student admissions.

References

[1] S. P Sipayung, S. Novriadi Antonius, and G. Anirma Kandida Br., "Analisis Prediksi Jumlah Penerimaan Mahasiswa Baru Dengan Metode Regresi Linier Sederhana," *SNISTIK Semin. Nas. Inov. Sains Teknol. Inf. Komput.*, vol. 1, no. 2, pp. 376–382, 2024.

[2] Hendra Di Kesuma, D. Apriadi, H. Juliansa, and E. Etriyanti, "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda," *J. Ilm. Bin. STMIK Bina Nusant. Jaya Lubuklinggau*, vol. 4, no. 2, pp. 62–66, 2022, doi: 10.52303/jb.v4i2.74.

[3] A. Yordan, T. N. Putri, and D. H. Lamkaruna, "Peramalan Penerimaan Mahasiswa Baru Universitas Samudra Menggunakan Metode Regresi Linear Sederhana," *J. Tek. Inform.*, vol. 2, no. 1, pp. 21–27, 2019, doi: 10.52046/j-tifa.v2i1.237.

[4] K. A. Mahendra, K. Gus, O. Ciptahadi, and N. M. Astiti, "Sistem Peramalan Penerimaan Siswa Baru Menggunakan Metode Regresi Linier Sederhana Berbasis Website," vol. 1, no. 3, pp. 334–339, 2024.

[5] N. Almumtazah, N. Azizah, Y. L. Putri, and D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana," *J. Ilm. Mat. DAN Terap.*, vol. 18, no. 1, pp. 31–40, 2021, doi: 10.22487/2540766X.2021.V18.I1.15465.

[6] A. D. Feriawan, M. Musta'in, M. Munif, and D. B. Reknadi, "Klasifikasi Layanan Pengaduan Di Smk Nu 2 Kedungpring Lamongan Dengan Menggunakan Metode Algoritma K-Nn," *J. Khatulistiwa Inform.*, vol. 12, no. 2, pp. 98–106, 2024, doi: 10.31294/jki.v12i2.24696.

[7] A. Yaqin, A. S. Budi, and P. H. Susilo, "Sistem Prediksi Jumlah Penumpang Di Bandar Udara Juanda Surabaya Dengan Metode Double Exponential Smooting," *Joutica*, vol. 7, no. 1, p. 546, 2022, doi: 10.30736/jti.v7i1.801.