

Implementation of Linear Regression Algorithm in a Web-Based Major Prediction System for New Student Applicants at SMK N 1 Percut Sei Tuan

Sabrina Meylani Pulungan^{1}, Mhd. Zulfansyuri Siambaton¹, Heri Santoso²*

¹ Department of Informatics Engineering, Faculty of Engineering, Universitas Islam Sumatera Utara, 20217, North Sumatra, Indonesia

² Department of Information System, Faculty of Computer Science, Universitas Islam Negeri Sumatera Utara, Medan, 20353, North Sumatra, Indonesia

ARTICLE INFORMATION

Received: Jun 21, 2025
Revised: Oct 15, 2025
Available Online: Oct 16, 2025

KEYWORDS

Linear Regression
Major Prediction
Web-Based System
SMK N 1 Percut Sei Tuan
PPDB

CORRESPONDENCE (*)

Phone: +62 851-8683-4701
E-mail: sabrinameylani17@gmail.com

A B S T R A C T

This study aims to develop a web-based major prediction system by applying a linear regression algorithm to enhance transparency and accuracy in the selection process. The system predicts 14 available majors at SMK N 1 Percut Sei Tuan, including: Civil Construction and Housing Engineering, Modeling and Building Information Design, Geomatics Engineering, Electrical Installation Engineering, Electrical Power Network Engineering, Heating, Air Conditioning and Refrigeration Engineering, Audio Video Engineering, Machining Engineering, Welding Engineering, Light Vehicle Engineering, Motorcycle Engineering, Software Engineering, Computer and Network Engineering, and Television Production and Broadcasting. The system uses report card scores from the 5th and 6th semesters of junior high school as predictor variables, including Bahasa Indonesia, Mathematics, Science, and English. The system development method includes data collection through observation, literature study, and interviews, as well as system design using PHP, HTML, JavaScript, MySQL database, and XAMPP. System modeling was carried out using UML (Unified Modeling Language), which includes use case diagrams, sequence diagrams, and activity diagrams. The linear regression algorithm is implemented by calculating subject averages, regression coefficients, and intercepts to predict student acceptance. The results of the study, based on five student data samples, show that M. Dafi and Ahmad Suhendra were not eligible for any major. Adellya Saputri and Alfit Septian were accepted into one major, Television Production and Broadcasting. Meanwhile, Ummi qualified for five majors: Modeling and Building Information Design, Audio Video Engineering, Welding Engineering, Light Vehicle Engineering, and Television Production and Broadcasting.

INTRODUCTION

The New Student Admission (PPDB) process is one of the most crucial stages in education, as it determines the quality of students admitted to a school. SMK N 1 Percut Sei Tuan, as a Vocational High School offering various majors, provides several programs that prospective students can choose through the achievement-based admission track. However, in this major selection process, problems often arise—such as the lack of awareness among prospective students about the opportunities available to them based on their academic report scores.

With advancements in technology and data science, the use of predictive methods in forecasting the most suitable major for new students can enhance the effectiveness and efficiency of the achievement-based PPDB system. One of the methods that can be applied is Linear Regression, a machine learning algorithm that can be used to predict a student's chance of acceptance based on specific variables—in this study, academic report scores are used.

Machine Learning is one of the essential elements in the field of Artificial Intelligence that plays a crucial role in solving various types of problems [1]. This technology represents an implementation of artificial intelligence that focuses on

developing systems capable of learning autonomously without requiring repeated instructions from humans [2]. In general, Machine Learning is a branch of computer science that studies algorithms and statistical models used by computer systems to perform specific tasks without explicit instructions, but rather through the learning of patterns and inference from available data [3]. The main objective of applying Machine Learning is to train machines to process and analyze data more efficiently [4]. Through the process of recognizing and understanding patterns within a dataset, Machine Learning can predict and identify the characteristics of data or objects that were previously unknown [5].

The application of Machine Learning in systematic reviews provides significant advantages, particularly in reducing the amount of manual work typically required in such processes [6]. The algorithms in Machine Learning are capable of learning complex relationships and patterns within data, thus eliminating the dependency on rigid rule-based approaches. As a result, systems are able to produce decisions with a higher degree of accuracy [7].

However, despite its many advantages, the implementation of Machine Learning also presents certain drawbacks. One major issue is its vulnerability to attacks or data manipulation, which can lead to misclassification [8]. In such cases, attackers may exploit weaknesses in the system to cause benign activities to be classified as malicious ones. Furthermore, data imbalance remains a significant challenge, as Machine Learning algorithms tend to achieve higher accuracy for majority classes while often failing to correctly recognize or predict minority classes [9].

The implementation of Linear Regression in a web-based prediction system can assist prospective students in objectively and accurately predicting their chances of being accepted into a desired major according to their report card scores. This system can provide clearer insights into which majors offer better acceptance opportunities, allowing students to make more transparent and data-driven decisions. Additionally, prospective students can estimate their chances before registering, making the PPDB process more informative and efficient.

Based on this issue, this study aims to develop a web-based major prediction system for new student applicants using the Linear Regression algorithm. This system is expected to assist prospective students of SMK N 1 Percut Sei Tuan in improving the effectiveness of the major selection process, reducing errors in decision-making, and enhancing the overall quality of student admissions.

METHOD

Implementation

Implementation refers to the actions taken to achieve the objectives that have been established in a decision. These actions aim to transform those decisions into operational patterns and strive to bring about changes—whether major or minor—as previously determined [10].

Algorithm

Algorithm is an effort consisting of a sequence of operations arranged logically and systematically to solve a problem and produce a specific output [11].

Linear Regression

The Linear Regression method is a type of regression that involves the relationship between one dependent variable and one independent variable, or between the dependent variable (Y) and the independent variable (X) [12].

The relationship between the dependent variable and the independent variable can take several forms of equations, such as linear, exponential, and multiple relationships. The purpose of using regression analysis is to estimate the value of the dependent variable based on the value of the independent variable [13].

The linear regression method is based on patterns of relationships in historical data. In general, the predictable variable, represented by a dependent variable (such as inventory), is influenced by the magnitude of the independent variable. The relationship between the independent variable and the variable to be predicted is expressed as a function [14].

System

System is a collection of interrelated elements that work together to perform activities in order to achieve a specific goal. The definition of a system can be viewed from its inputs and outputs. A system is a series that functions to receive inputs, process those inputs, and produce outputs. A well-functioning system is able to adapt and survive within its environment. When viewed from its procedures or activities, a system is a series of procedures or activities designed to implement a company's program. For example, an accounting system includes procedures involving forms, journals, ledgers,

subsidiary ledgers, worksheets, and financial reports [15]. System is essentially a group of elements that are closely interconnected and work together to achieve a specific goal [16].

Prediction

Prediction is a systematic process of estimating the likelihood of future events based on past and present information. The purpose of prediction is to minimize errors (the difference between what is predicted and what actually occurs). Prediction does not always provide a definite answer about future events; rather, it aims to produce an estimate that is as close as possible to what is likely to happen [17]. Prediction plays an important role in various fields, including economics, healthcare, engineering, and the environment. The application of prediction in certain institutions or organizations enables decision-making or policy development related to what is likely to happen in the future, based on previously available data [18].

PHP

PHP stands for PHP (Hypertext Preprocessor). It is a server-side scripting language commonly used to create dynamic web applications [19]. PHP is a server-side scripting language designed to integrate with HTML in order to create dynamic web pages. The main purpose of using PHP is to ensure that all commands and syntax are executed entirely on the server side, even though they are embedded within the HTML document [20].

RESULTS AND DISCUSSION

Intercept and Coefficient Value Data

The intercept is the constant value in a linear regression equation that represents the point where the regression line intersects the y-axis when all predictor (in dependent) variables are equal to zero. It serves as the initial prediction value before considering the contribution of each subject score. In other words, it is the baseline estimate of graduation if all subject scores are assumed to be zero (although this scenario is rarely encountered in practice). Coefficients in linear regression are values that indicate how much influence each independent variable (e.g., subject scores) has on the dependent variable (e.g., graduation prediction). The coefficients represent the level of contribution each subject makes to the predicted graduation outcome. In this study, the author uses four coefficients: the coefficient of Bahasa Indonesia, the coefficient of Mathematics, the coefficient of Science (IPA), and the coefficient of English, which are used as parameters in determining graduation.

The initial stage in the linear regression algorithm calculation, after inputting the prospective student data, is to calculate the average scores of the odd and even semester subjects. The following are the sample student data:

Table 1. Table of Student Sample Data

Name	Even Semester				Odd Semester			
	B.Indonesia	Matematika	IPA	B. Inggris	B.Indonesia	Matematika	IPA	B. Inggris
M. Dafi	80	80	78	79	87	88	76	76
Ummi	88	89	87	85	88	88	89	86
Ahmad Suhendra	79	75	80	78	80	79	85	80
Adellya Saputri	80	87	85	90	78	88	87	89
Alfit Septian	88	80	86	89	90	81	82	83

Table 2. Average Score Results

Name	B.Indonesia	Matematika	IPA	B.Inggris
M. Dafi	83,5	84	77	77,5
Ummi	88	88,5	88	85,5
Ahmad Suhendra	79,5	77	82,5	79
Adellya Saputri	79	87,5	86	89,5
Alfit Septian	89	80,5	84	86

Table 3. Comparison Result Table of M.Dafi

Major	Prediction	Minimum Passing Score	Result
Teknik Konstruksi dan Perumahan	64,66	72	TIDAK LULUS
Desain Pemodelan dan Informasi Bangunan	67,59	74	TIDAK LULUS
Teknik Geomatika	64,00	75	TIDAK LULUS
Teknik Instalasi Tenaga Listrik	67,70	74	TIDAK LULUS
Teknik Jaringan Tenaga Listrik	64,35	72	TIDAK LULUS
Teknik Pemanasan, Tata Udara dan Pendinginan	63,75	73	TIDAK LULUS
Teknik Audio Video	68,25	73	TIDAK LULUS
Teknik Pemesianan	64,76	76	TIDAK LULUS
Teknik Pengelasan	67,36	70	TIDAK LULUS
Teknik Kendaraan Ringan	67,36	71	TIDAK LULUS
Teknik Sepeda Motor	63,49	70	TIDAK LULUS
Rekayasa Perangkat Lunak	69,41	78	TIDAK LULUS
Teknik Komputer dan Jaringan	67,16	77	TIDAK LULUS
Produksi dan Siaran Program Televisi	69,39	70	TIDAK LULUS

Table 4. Comparison Result Table of Umami

Major	Prediction	Minimum Passing Score	Result
Teknik Konstruksi dan Perumahan	69,31	72	TIDAK LULUS
Desain Pemodelan dan Informasi Bangunan	72,44	74	LULUS
Teknik Geomatika	68,48	75	TIDAK LULUS
Teknik Instalasi Tenaga Listrik	72,74	74	TIDAK LULUS
Teknik Jaringan Tenaga Listrik	69,33	72	TIDAK LULUS
Teknik Pemanasan, Tata Udara dan Pendinginan	68,20	73	TIDAK LULUS
Teknik Audio Video	73,48	73	LULUS
Teknik Pemesianan	69,68	76	TIDAK LULUS
Teknik Pengelasan	72,80	70	LULUS
Teknik Kendaraan Ringan	72,73	71	LULUS
Teknik Sepeda Motor	68,38	70	TIDAK LULUS
Rekayasa Perangkat Lunak	74,48	78	TIDAK LULUS
Teknik Komputer dan Jaringan	72,20	77	TIDAK LULUS
Produksi dan Siaran Program Televisi	74,45	70	LULUS

Table 5. Comparison Result Table of Ahmad Suhendra

Major	Prediction	Minimum Passing Score	Result
Teknik Konstruksi dan Perumahan	63,59	72	TIDAK LULUS
Desain Pemodelan dan Informasi Bangunan	66,63	74	TIDAK LULUS
Teknik Geomatika	62,43	75	TIDAK LULUS
Teknik Instalasi Tenaga Listrik	66,51	74	TIDAK LULUS
Teknik Jaringan Tenaga Listrik	63,68	72	TIDAK LULUS
Teknik Pemanasan, Tata Udara dan Pendinginan	62,49	73	TIDAK LULUS
Teknik Audio Video	67,45	73	TIDAK LULUS
Teknik Pemesianan	63,56	76	TIDAK LULUS

Teknik Pengelasan	66,75	70	TIDAK LULUS
Teknik Kendaraan Ringan	66,73	71	TIDAK LULUS
Teknik Sepeda Motor	62,75	70	TIDAK LULUS
Rekayasa Perangkat Lunak	68,38	78	TIDAK LULUS
Teknik Komputer dan Jaringan	66,44	77	TIDAK LULUS
Produksi dan Siaran	68,50	70	TIDAK LULUS
Program Televisi			

Table 6. Comparison Result Table of Adellya Saputri

Major	Prediction	Minimum Passing Score	Result
Teknik Konstruksi dan Perumahan	67,41	72	TIDAK LULUS
Desain Pemodelan dan Informasi Bangunan	70,44	74	TIDAK LULUS
Teknik Geomatika	66,53	75	TIDAK LULUS
Teknik Instalasi Tenaga Listrik	70,96	74	TIDAK LULUS
Teknik Jaringan Tenaga Listrik	67,98	72	TIDAK LULUS
Teknik Pemanasan, Tata Udara dan Pendinginan	66,60	73	TIDAK LULUS
Teknik Audio Video	72,63	73	TIDAK LULUS
Teknik Pemesianan	67,85	76	TIDAK LULUS
Teknik Pengelasan	71,23	70	TIDAK LULUS
Teknik Kendaraan Ringan	71,10	71	TIDAK LULUS
Teknik Sepeda Motor	66,50	70	TIDAK LULUS
Rekayasa Perangkat Lunak	73,45	78	TIDAK LULUS
Teknik Komputer dan Jaringan	71,53	77	TIDAK LULUS
Produksi dan Siaran	73,03	70	LULUS
Program Televisi			

Table 7. Comparison Result Table of Alfit Septian

Major	Prediction	Minimum Passing Score	Result
Teknik Konstruksi dan Perumahan	67,61	72	TIDAK LULUS
Desain Pemodelan dan Informasi Bangunan	70,98	74	TIDAK LULUS
Teknik Geomatika	66,13	75	TIDAK LULUS
Teknik Instalasi Tenaga Listrik	70,51	74	TIDAK LULUS
Teknik Jaringan Tenaga Listrik	67,15	72	TIDAK LULUS
Teknik Pemanasan, Tata Udara dan Pendinginan	66,04	73	TIDAK LULUS
Teknik Audio Video	71,55	73	TIDAK LULUS
Teknik Pemesianan	67,10	76	TIDAK LULUS
Teknik Pengelasan	70,21	70	TIDAK LULUS
Teknik Kendaraan Ringan	70,48	71	TIDAK LULUS
Teknik Sepeda Motor	66,44	70	TIDAK LULUS
Rekayasa Perangkat Lunak	72,68	78	TIDAK LULUS
Teknik Komputer dan Jaringan	70,59	77	TIDAK LULUS
Produksi dan Siaran	73,10	70	LULUS
Program Televisi			

CONCLUSION

The Linear Regression algorithm has proven effective as a predictive method for determining students' potential majors at SMK N 1 Percut Sei Tuan. By utilizing academic scores in subjects such as Bahasa Indonesia, Mathematics, Science, and English, the system can generate accurate predictions regarding the likelihood of a student's acceptance into specific majors. The implementation results demonstrate the effectiveness of the model in supporting the major selection process. Based on the testing of five student data samples, the system successfully identified each student's eligibility across several majors, with prediction results closely aligned with the school's expected outcomes. The web-based implementation of the system enables a fast, user-friendly, and integrated prediction process for both prospective students and school administrators. Therefore, the system serves as an efficient decision-support tool for the New Student Admission (PPDB) process.

REFERENCES

- [1] J. Homepage, A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper," *IJCIT (Indonesian Journal on Computer and Information Technology)*, 2019.
- [2] C. Chazar and B. E. Widhiaputra, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 12, pp. 67–80, 2020.
- [3] R. R. Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktivitas Manusia," *Jurnal Matrik*, vol. 19, pp. 302–311, May 2020.
- [4] B. Mahesh, "Machine Learning Algorithms — A Review," *International Journal of Science and Research*, 2018, doi: 10.21275/ART20203995.
- [5] M. Ula, A. F. Ulva, and Mauliza, "Implementasi Machine Learning dengan Model Case Based Reasoning dalam Mendiagnosa Gizi Buruk pada Anak," *Jurnal Informatika Kaputama (JIK)*, vol. 5, pp. 333–339, Jul. 2021.
- [6] B. G. Pijls, "Machine Learning Assisted Systematic Reviewing in Orthopaedics," *Journal of Orthopaedics*, vol. 48, pp. 103–106, Feb. 2024, doi: 10.1016/j.jor.2023.11.051.
- [7] S. Jahandideh, G. Ozavci, B. W. Sahle, A. Z. Kouzani, F. Magrabi, and T. Bucknall, "Evaluation of Machine Learning-Based Models for Prediction of Clinical Deterioration: A Systematic Literature Review," *International Journal of Medical Informatics*, vol. 175, Jul. 2023, doi: 10.1016/j.ijmedinf.2023.105084.
- [8] O. Alshaikh, S. Parkinson, and S. Khan, "Exploring Perceptions of Decision-Makers and Specialists in Defensive Machine Learning Cybersecurity Applications: The Need for a Standardised Approach," *Computers & Security*, p. 103694, Dec. 2023, doi: 10.1016/j.cose.2023.103694.
- [9] A. X. Wang, S. S. Chukova, and B. P. Nguyen, "Synthetic Minority Oversampling Using Edited Displacement-Based K-Nearest Neighbors," *Applied Soft Computing*, vol. 148, Nov. 2023, doi: 10.1016/j.asoc.2023.110895.
- [10] Mulyadi, *Implementasi Organisasi*. Yogyakarta: Gadjah Mada University Press, 2015.
- [11] M. Kani, *Algoritma dan Pemrograman*. Tangerang Selatan: Universitas Terbuka, 2020.
- [12] A. Kurniadi and Y. Novianto, "Penerapan Metode Regresi Linier untuk Memprediksi Kebiasaan Pelanggan Studi Kasus: PT Mensa Binasukses," *Jurnal Ilmiah Mahasiswa Teknik Informatika*, vol. 2, no. 2, p. 107, 2020.
- [13] A. Bode, "Perbandingan Metode Prediksi Support Vector Machine dan Linear Regression Menggunakan Backward Elimination pada Produksi Minyak Kelapa," *Jurnal Sistem Informasi dan Teknologi Komputer*, vol. 4, no. 2, pp. 104–107, 2019.
- [14] D. A. Trianggana, "Peramalan Jumlah Siswa-Siswi Melalui Pendekatan Metode Regresi Linear," *Jurnal Media Infotama*, vol. 16, no. 2, pp. 115–120, 2020.
- [15] V. W. Sujarweni, *Sistem Akuntansi*. Yogyakarta: Pustaka Baru Press, 2019.
- [16] Mulyadi, *Sistem Informasi Akuntansi*. Jakarta: Salemba Empat, 2018.
- [17] Srisulistiowati, "Sistem Informasi Prediksi Penjualan Alat Tulis Kantor dengan Metode FP-Growth (Studi Kasus Toko Koperasi Sekolah Bina Mulia)," unpublished.
- [18] C. Mashuri, *Sistem Informasi Prediksi dengan Fuzzy dan RFID pada VMI*. Perkumpulan Rumah Cemerlang Indonesia, 2022.
- [19] J. Winanjar and D. Susanti, "Rancang Bangun Sistem Informasi Administrasi Desa Berbasis Web Menggunakan PHP dan MySQL," *Prosiding Seminar Nasional Sains dan Teknologi (SNAST)*, pp. 97–105, 2021.
- [20] T. Lesmana and M. Silalahi, "Jurnal Comasie," *Comasie*, vol. 3, no. 3, pp. 21–30, 2020.