International Journal of Computer Sciences and Mathematics Engineering Vol. 3 No. 1 (2024)
Published by: CV. HAWARI

International Journal of Computer Sciences and Mathematics Engineering
Journal homepage: www.ijecom.org

Data Mining Analysis in Predicting the Number of New Students at IT&B Indonesia Using Linear Regression Method

Nazlita Febrina Hanum Br. Brahmana[1], Zulham Sitorus [2], Khairul[3] , Rian Farta Wijaya [4], Darmeli Nasution[5]
[1,2,3,4,5]Department of Master of Information Technology, Universitas Pembangunan Panca Budi, Indonesia

Article Info

Article history:

Received May 3, 2024
Revised May , 2024
Accepted May , 2024

Keywords:

Data Mining
New Student
Linear Regression

ABSTRACT

This study explores the application of data mining techniques, particularly Multiple Linear Regression, to predict the number of new students at IT&B Indonesia based on historical admission data from 2010/2011 to 2021/2022 academic years. The findings indicate that the Multiple Linear Regression model achieved a 64% accuracy rate with a mean square error of 8.335, demonstrating its effectiveness in estimating new student admissions. Additionally, the developed system can be utilized to predict student performance based on predefined criteria, providing valuable insights for educational institutions in forecasting future admissions.

Corresponding Author:

Nazlita Febrina Hanum Br. Brahmana
Master Of Information Technology
Universitas Pembangunan Panca Budi
Jl.Gatot Subroto Km.4,5 Sei Sikambing, Kota Medan, Indonesia.

1. INTRODUCTION

This study explores the utilization of data mining, particularly through the linear regression method, to forecast the number of new students at the Institut Teknologi dan Bisnis Indonesia (IT&B). The research findings demonstrate that linear regression can provide fairly accurate predictions, with an accuracy rate of 64%. These findings hold significant potential to assist IT&B in anticipating trends in new student admissions, planning budget allocations, and designing more effective marketing strategies.

It is hoped that the findings of this study will make a meaningful contribution to IT&B's strategic planning in managing new student admissions. The predictive information provided by the linear regression method can serve as a strong foundation for making more informed and timely decisions in addressing the challenges and opportunities related to increasing the number of new students in the future. Thus, this research not only has the potential to enhance operational efficiency and marketing strategy effectiveness at IT&B but also contributes to the institution's vision and mission of providing quality and relevant education to the community.

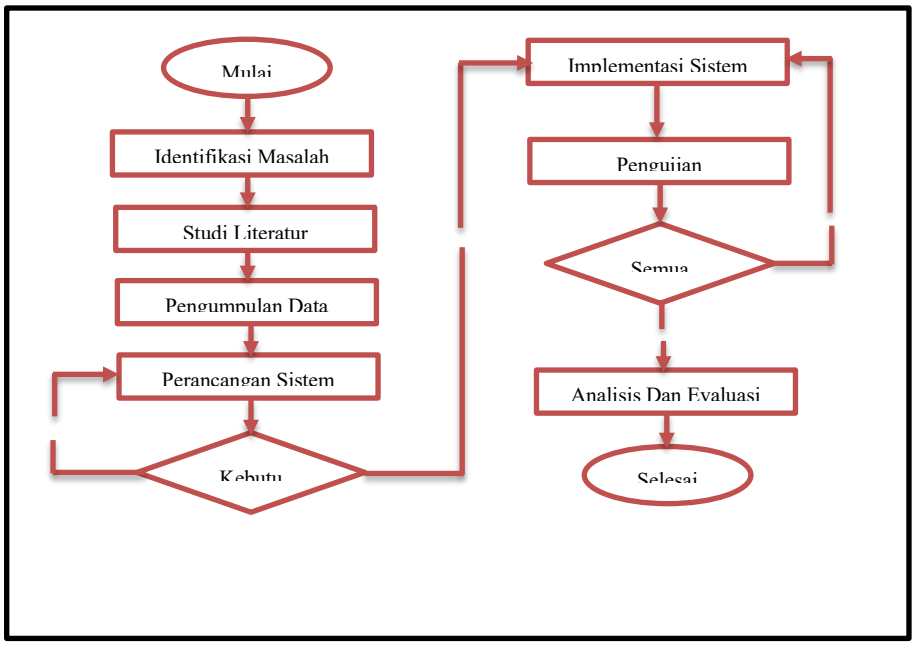
E-ISSN 2962-4274

2. RESEARCH METHOD

This research phase is conducted to address the issues at hand. It aims to gather information related to the problem under investigation. This is done with the purpose of collecting data and understanding the nature or position of the specific problem both theoretically and practically.

Subsequently, the next stage involves data collection, such as interviewing soil science experts and collecting samples in the form of strategic agricultural commodities. This is followed by the next stage, which is system design, where an overview of the system will be created using flowcharts and low-fidelity system interfaces. Implementation and system testing will then follow. If the system passes the testing phase, the final stage is analysis and evaluation, where the method used in the system will be assessed for its effectiveness. The flowchart outlining the steps taken in this research is depicted in Figure.1.

Figure. 1 Flowchart of the Research Process



Based on the research framework above, the following points can be outlined:

1. Problem Identification

The research framework begins with the identification of the problem, which involves determining the object to be studied and formulating the issues faced by the school, particularly to understand how many students will enroll each year. Upon identifying the issues, the researcher seeks solutions or ways out of the said problems.

2. Data Collection

In this stage, the researcher gathers the necessary data for the study. The techniques used for data collection include:

a. Library Research

The method involves studying various literature sources such as journals on data mining, articles, linear regression algorithms, and other scientific books related to the research. Data and information are collected by reading scholarly sources from books or articles as references for the discussion of this issue.

b. Field Research

E-ISSN 2962-4274

To obtain relevant data directly from the object being studied in a real, clear, accurate, and systematic manner, the researcher collects data directly. The data collected from this research are primary data obtained through.

c. Observation

Observation is a systematic and intentional data collection technique conducted through direct observation. In this method, direct observation activities are carried out on the object under study, and the data collected are based on facts or are suitable for the needs of Institut Teknologi dan Bisnis Indonesia.

d. Interviews

To obtain more directed and accurate data, the interview method is used by asking several questions to school administrators and employees of Institut Teknologi dan Bisnis Indonesia.

3. Data Analysis

Once the data has been collected, in relation to the topic of the problem faced, the analysis method used to analyze the issue is:

a. Descriptive Method

The Descriptive Method is a technique carried out by collecting, grouping, formulating, analyzing, and summarizing data so that the data can provide a clear overview of the number of registrants.

b. Deductive Method

The Deductive Method is a method that analyzes data by drawing conclusions based on theories received from a legal reasoning about observed facts. Subsequently, the researcher draws conclusions.

Linear Regression Method Calculation

This stage involves calculating the decisions generated using the linear regression method. After analyzing, the system design is then carried out using the established system design.

Implementation

Implementation is the stage where the previously made designs are turned into applications using PHP programming and MySql databases that can be utilized by the school to predict the number of new students at Institut Teknologi dan Bisnis Indonesia.

System Testing

In this stage of system testing, tests are conducted on the system. Testing is carried out to identify potential errors in the coding process and to ensure that the restricted inputs yield the desired results.

3. RESULTS AND DISCUSSIONS

In this study, the results of field study research will be discussed, starting with descriptive statistics related to the collected data (including an overview of respondents, variable descriptions, and data collection guidelines from the academic years 2010/2011 to 2021/2022); hypothesis testing results and discussions regarding the hypothesis tests conducted using linear regression and Microsoft Visual Studio Program.

1. Linear Regression Method Testing

Based on the problem formulated in the previous chapters above, it can be determined how many variables are needed to address the aforementioned problem. To address the occurring problem, linear regression method is used to predict the number of new students to be enrolled in the subsequent academic years based on related variables.

Addressing the issue of predicting new students involves finding a straight line from the two existing variables, namely the tuition fee amount (x) and the number of new students (y). In solving the issue of predicting the number of new students using two variables and data obtained from Institut Teknologi dan Bisnis Indonesia. The following is the data on the number of new students from the academic years 2010/2011 to 2021/2022.

E-ISSN 2962-4274

Table. 1 Student Data for Academic Years 2010/2011 - 2021/2022

NO	Tahun Ajar	Jumlah Uang Kuliah	Jumlah Pendaftar
1.	2010/2011	460.000	260
2.	2012/2013	500.000	240
3.	2013/2014	540.000	228
4.	2014/2015	580.000	223
5	2015/2016	625.000	235
6.	2016/2017	710.000	233
7.	2017/2018	750.000	220
8.	2018/2019	790.000	210
9.	2019/2020	830.000	200
10.	2020/2021	875.000	183

The method used is linear regression. The calculation process can be seen as follows:

Step 1
The purpose of conducting linear regression analysis is to predict/understand the influence of previous data on the data to be predicted.
Step 2
Identify the independent (X) and dependent (Y) variables:
Independent variable (X): Amount of tuition fee
Dependent variable (Y): Number of applicants
Step 3
Collecting data.

Table. 2 Calculation of Independent and Dependent Variable Values

No	Tahun Ajaran	Jumlah Uang Kuliah (X)	Jumlah Pendaftar (Y)	X^2	Y^2	X.Y
1	2010/2011	460,000	260	211,600,000,000	67600	119,600,000
2	2011/2012	500,000	240	250,000,000,000	57600	120,000,000
3	2012/2013	540,000	228	291,600,000,000	51984	123,120,000
4	2013/2014	580,000	223	336,400,000,000	49729	129,340,000
5	2014/2015	625,000	235	390,625,000,000	55225	146,875,000
6	2015/2016	710,000	233	504,100,000,000	54289	165,430,000
7	2016/2017	750,000	220	562,500,000,000	48400	165,000,000
8	2017/2018	790,000	210	624,100,000,000	44100	165,900,000
9	2018/2019	830,000	200	688,900,000,000	40000	166,000,000
10	2019/2020	875,000	183	765,625,000,000	33489	160,125,000
12	2021/2022					
TOTAL		**6,660,000**	**2,232**	**4,625,450,000,000**	**502,416**	**1,461,390,000**

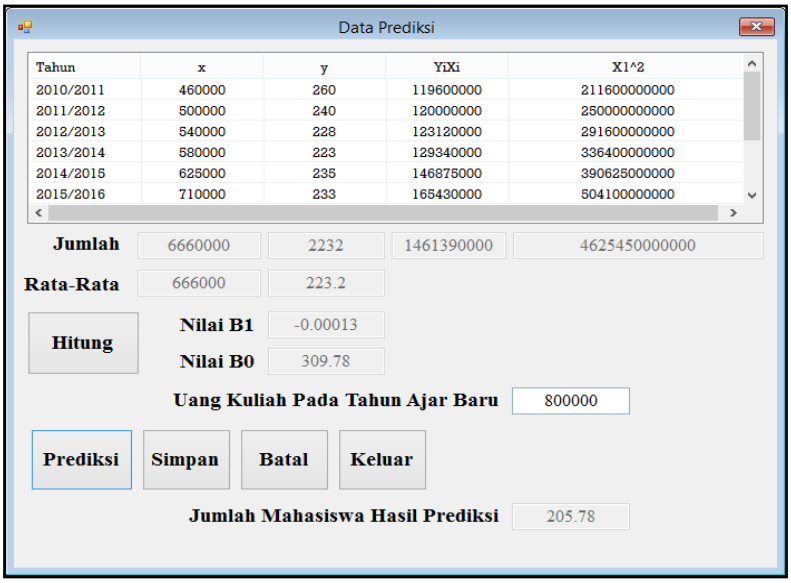
E-ISSN 2962-4274

Step 4
Calculating the constant values *a* and *b*
b.The calculation is done using the following formulas:
a. Calculating the constant value *a*
a = 311.31
b = - 0.00013

Step 5 The Linear Regression Equation Model, The Linear Regression Equation Model is as follows:
Y = a + bX
Y = 311.31+(-0.00013) (X)

Step 6
Performing forecasting for the causal factor or outcome variable. Forecast the number of new student registrations taken from the last 10 years (Variable X) as follows:
Y = 311.31+(-0.00013) (11)
Y = 311.308
From the above result, it can be concluded that if the number of registrations in the following year based on the above calculation is as much as 311.308, rounded to 311 people.

Figure. 2 Prediction Form Display



4. CONCLUSION

Based on the research conducted, the analysis and discussion regarding data mining to predict the number of new students based on the enrollment data of new students from the academic year 2010/2011 to 2021/2022 using the Multiple Linear Regression method conducted at IT&B Indonesia, the following conclusions can be drawn:

The prediction results obtained through data mining using the multiple linear regression method to determine the level of enrollment of new students in this system can be implemented for estimating the enrollment of new students using the multiple linear regression method fairly

E-ISSN 2962-4274

accurately, with a 64% accuracy rate using Multiple Linear Regression being quite accurate because it has a mean square error of 8.335.

It can be used to predict student performance based on predetermined criteria, and the system built using desktop-based programming languages can assist IT&B Indonesia in predicting the enrollment of new students based on research findings.

The results obtained show that using the Multiple Linear Regression algorithm has a fairly accurate level of accuracy.

REFERENCES

[1] Afuan, L. (2010). Pullmanfaatan Framullwork Codullignitullr dalam Pullngullmbangan Sistullm Informasi Pullndataan Laporan Kullrja Praktullk Mahasiswa Program Studi Tullknik Informatika Unsoulld. Juita, I(2), 39–44.Ayuni, G. N., & Fitrianah, D. (2019). Pullnullrapan Mulltodull Rullgrullsi Linullar Untuk Prulldiksi Pullnjualan Propullrti pada PT XYZ. Jurnal Tulllullmatika, 14(2), 79–86. https://journal.ithb.ac.id/tulllullmatika/articlull/viullw/321

[2] Cahyodi, S. C., & Arifin, R. W. (2017). Sistullm Informasi Point Of Salulls Bullrbasis Wullb Pada Colony Amaranta Bullkasi. Mulli Rullviullw Mulli, 1(12), 189–204. http://download.portalgaruda.org/articlull.php?articlull=535077&val=10928&titlull= Sistullm Informasi Point Of Salulls Bullrbasis Wullb Pada Colony Amaranta Bullkasi.

[3] Dullwi, K. R., Mauladi, K. F., Informatika, T., Tullknik, F., & Lamongan, U. I. (2020). Analisa Algoritma C4. 5 untuk Prulldiksi Pullnjualan Obat Pullrtanian di Toko Dullwi Sri. 109–114.

[4] Hijriani, A., Muludi, K., & Andini, Ull. A. (2016). Implullmullntasi Mulltodull Rullgrullsi Liniullr Sulldullrhana Pada Pullnyajian Hasil Prulldiksi Pullmakaian Air Bullrsih Pdam Way Rilau Kota Bandar Lampung Dullngan Sistullm Informasi Gullofrafis. Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputullr, 11(2), 37. https://doi.org/10.30872/jim.v11i2.212.

[5] Khotimah, T., & Nindyasari, R. (2017). Forullcasting Dullngan Mulltodull Rullgrullsi Liniullr Pada Sistullm Pullnunjang Kullputusan Untuk Mullmprulldiksi Jumlah Pullnjualan Batik (Studi Kasus Kub Sarwo Ullndah Batik Tulis Lasullm). Jurnal Mantik Pullnusa, 1(1), 71–92. http://ulljurnal.pulllitanusantara.ac.id/indullx.php/mantik/articlull/viullw/541.

[6] Marisa, F., & Yuarita, T. G. (2017). Pullrancangan Aplikasi Point of Salulls (Pos) Bullrbasis Wullb Mullnggunakan Mulltodull Siklus Hidup Pullngullmbangan Sistullm. Jurnal Tullknologi Dan Manajullmulln Informatika, 3(2), 167–171. https://doi.org/10.26905/jtmi.v3i2.1514.