**JASa (Jurnal Akuntansi, Audit dan Sistem Informasi Akuntansi)**
ISSN 2550-0732 print / ISSN 2655-8319 online
Vol. 10 No. 1 / April 2026
DOI;10.36555/jasa.v10i1.2992

# Integration of Artificial Intelligence and Data Mining in Accounting Information Systems for Analysis of New Student Uptake Patterns: An Open University Case Study

**Diky Paramitha[1], Etik Ipda Riyani[1], Kan Wen Huey[2]**
Universitas Terbuka, Indonesia[1], Wawasan Open University, Malaysia[2]
*Coresponding Email : dikyparamitha@ecampus.ut.ac.id

## ABSTRACT

Increased enrollment of new students is an important measure of the university's ability to meet the needs of Education. To remain relevant and competitive, institutions are increasingly expected to incorporate artificial intelligence (AI) technologies that improve data-driven decision-making. The combination of data mining methods and Accounting Information Systems, which utilize computational techniques to analyze vast data sets can reveal hidden patterns. This study investigates the application of artificial intelligence in creating a direct data mining tool that aims to identify trends in new student enrollment at the Open University. Using a research and development approach, the study examined admissions data from 2022 to 2024 to uncover underlying patterns and significant factors influencing enrollment dynamics. The results show that AI-integrated applications significantly improve the efficiency of data analysis while improving the reliability and usability of institutional data for strategic purposes. The study contributes to a larger conversation about how aligning AI implementation with institutional goals can drive innovation, accountability, and responsiveness in the context of open and distance learning. In summary, the findings imply that incorporating AI into data management and development processes provides greater clarity regarding future trends, making AI a powerful forecasting and decision-making resource for higher education institutions.

## ARTICLE INFO

**Article history:**
Submitted: 06 January 2026
Revised: 22 April 2026
Accepted: 24 April 2026
Published: 30 April 2026

**Keyword:**
Artificial Intelligence,
Accounting Information Systems,
Data Mining.

To cite this article (APA Style): Paramitha, D., Riyani, E. I., & Huey, K. W. (2026). Integration of Artificial Intelligence and Data Mining in Accounting Information Systems for Analysis of New Student Uptake Patterns: An Open University Case Study. JASa : Jurnal Akuntansi, Audit dan Sistem Informasi Akuntansi. Vol 10 (1), p.026-036. https://doi.org/10.36555/jasa.v10i1.2992

## INTRODUCTION

In the era of increasingly advanced technology, especially the development of Artificial Intelligence, it is easier to complete work, especially those related to data. The level of accuracy of a data is needed in daily life. The accuracy of data is important in everyday life. Every piece of information matters because it helps make decisions in different situations. This means that providing information needs to be carefully analyzed and turned into useful knowledge when making a decision. Data mining is the process of extracting important information that is hidden and not known before from a dataset (Witten, 2011). Data mining is a process where people gather and use past data to discover patterns, order, and connections within big sets of information (Santosa et al., 2007). Data mining is the process of finding interesting and useful patterns in large amounts of data. These patterns are not obvious, are not already known, and can provide valuable insights (Han et al., 2006). Using data mining helps each data set or warehouse share important knowledge that turns into valuable information for an organization, like a school or educational institution.

Currently, the distance education system is an education trend because with increasingly advanced technology, it can make it easier for all educational institutions to

implement a distance learning system. Accuracy in market segmentation in getting prospective new students is very necessary where market segmentation is an effort that is needed to improve marketing accuracy in an organization or company so that it can make it easier for businesses to run marketing programs or approach customers (Kotler, 2012). To deal with technological developments and the increasing development of each higher education institution in implementing a distance learning system, the Universitas Terbuka must be quick in responding to these changes, one of which is by utilizing its big data which is an important asset for the university. Big data must be managed properly in order to provide important information and knowledge, especially in the main sectors, namely the recruitment of prospective new students.

The recruitment process is one of the essential functions of the human resources department and is the first step towards the creation of a competitive force, while campus recruitment is the main recruitment mode for new talented graduates (Sivaram et al., 2010). In fact, a lot of data is Big data on a system in an organization that has not been utilized optimally and is converted into more useful knowledge, especially to predict the future. Most of them are just statistical data to see the development of a data. For this reason, further research related to data mining on big data is needed. Data mining is a tool to strengthen the perception of data.

The increase in the number of new students is one of the indicators of the success of universities in reaching and serving the educational needs of the community. Universitas Terbuka, which has the unique characteristics of being a distance learning-based university, an in-depth analysis of new student intake data is essential. Understanding the patterns and factors that affect the number of new students can help institutions to formulate more effective strategies to increase the number of enrollees and improve overall distance education services.

Data mining, as a technology-based method of data analysis, allows universities to unearth hidden patterns from student absorption data. However, this method can become more sophisticated and accurate when combined with artificial intelligence (AI). AI can improve data mining capabilities to recognize complex patterns in big data with "speed and accuracy that conventional techniques can't achieve. With AI, Universitas Terbuka can identify the demographic, geographical, and preference factors of prospective students that influence the decision to enroll, so that it can design a more personalized and targeted strategy.

The use of AI and data mining is also in line with technological developments in higher education, where institutions are required to "leverage technology for better decision-making, which is based on real-time data and predictive analytics. Therefore, the development of AI-based data mining applications can help Universitas Terbuka optimize its new student recruitment strategy, increase competitiveness, and better understand the needs of prospective students.

## Ai Concept

AI can be defined as "the science and techniques for creating intelligent machines, specifically computer programs that are capable of understanding, learning, and acting according to their environment (Russell and Norvig, 2020). The development of AI has gone through various stages, from the creation of simple rules-based algorithms to the use of machine learning and deep learning that enable big data analysis with a high level of complexity. Deep learning is one of the biggest breakthroughs in AI, allowing models to learn from unstructured data such as images, text, and sound (LeCun et al., 2015). In Artificial Intelligence, there are several main components in supporting the system process. Machine Learning in AI is a subfield of AI that allows systems to learn from data without being explicitly programmed. Commonly used algorithms include decision trees, support vector machines (SVMs), and neural networks (Goodfellow et al., 2016). Natural Language Processing (NLP) is a technique used to understand and generate human language, such

as in chatbot applications or sentiment analysis. The application of AI in education can be used for learning personalization, student performance analysis, and graduation prediction (Zawacki-Richter et al., 2019).

**Data Mining Concept**

This data mining is also known as pattern recognition (Santosa et al., 2007). Data mining is a large-scale data processing method, therefore data mining has an important role in the fields of industry, finance, weather, science and technology. In general, data mining studies discuss methods such as clustering, classification, regression, variable selection, and market basket analysis (Santosa et al., 2007). Output dari data mining ini dapat digunakan untuk menentukan kebijakan atau keputusan di masa yang akan dating (Santosa et al., 2007). The actual task of data mining is to analyze large amounts of data to extract previously unknown patterns such as data record clusters (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining) (Han et al., 2002).

**Growth Ratio and Classification Data Concept**

Growth ratio is a metric used to measure the growth rate of a variable in a certain period, such as revenue, profit, number of customers, or population growth. In the business context, the growth ratio is often an important indicator to evaluate the success of an organization in achieving its targets. growth ratios reflect past performance and future growth potential (Gitman and Zutter, 2012). Classification is the most commonly applied data mining technique, which uses a set of pre-classified attributes to develop a model that can classify record populations in general (Singh et al., n.d., 2013).

**Accounting Information Systems and the Naive Bayes Method**

Accounting Information Systems play an important role in improving the quality of accounting information, the effectiveness of internal controls, and the operational efficiency of organizations. Along with the development of information technology, Accounting Information Systems no longer only function as a transaction recording tool, but also as a means of analyzing financial and non-financial data in an intelligent and integrated manner. In the context of data-based decision-making, the application of data mining techniques in Accounting Information Systems is becoming increasingly relevant. Data mining allows organizations to extract hidden patterns and knowledge from large and complex data sets. Naive Bayes is a simple probabilistic classifier that calculates a set of probabilities by summing the combination of frequencies and values of a given dataset. This algorithm uses Bayes' theorem and assumes all the independent or non-interdependent attributes assigned by the value of a class variable (Patil, T. R., 2013). Another definition says Naive Bayes is a classifier with probability and statistical methods brought by the British scientist Thomas Bayes, predicting future opportunities based on previous experience (Bustami., 2013). The Naïve Bayes algorithm is a probabilistic method used to classify data classes. The Outline Naïve Bayes method is a statistical analysis in which the initial probability (previous probability) is estimated from traning data (Jantawan et al., 2014).

## METHODS

**Data**

This research uses the Research and Development method to develop AI-based applications and data mining. The data used are big data Universitas Terbuka students and data on graduates of the National Higher Secondary School of the Ministry of Primary and Secondary Education. The type of data used in this study uses primary and secondary data. Primary data in this study was obtained directly through observation through the student admission information system at the Universitas Terbuka. Meanwhile, secondary

data is obtained through coordination with the data team of the Ministry of Primary and Secondary Education and historical data of new students. The data is processed using the growth ratio formula to see the growth of the data. The data will be analyzed using data mining techniques such as K-Means clustering, classification, and Decision tree to identify patterns and trends. Non-probability sampling is suitable for exploratory research intended to generate new ideas to be tested systematically, non-probability techniques make it possible to sample from such populations whose elements are unlimited in number (Alvi, 2016). The steps can be seen in the image below:

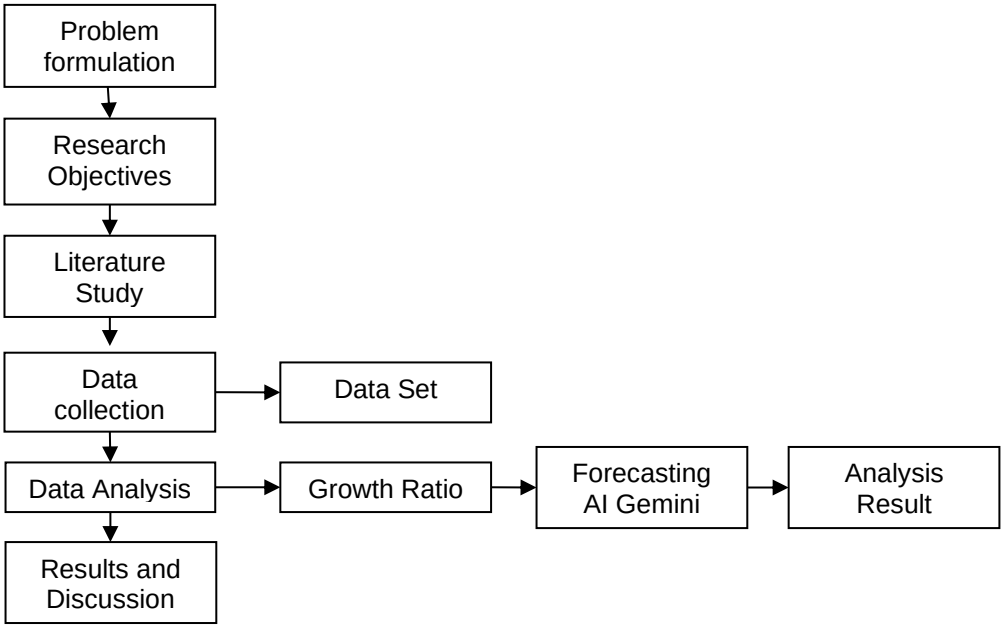


**Figure 1. Research Stage Flow**
Source: Data Process (2026)

## METHODS

The study adopts a quantitative approach within the framework of research and development (R&D). The main goal is to develop and implement artificial intelligence (AI)-based data mining tools to identify new student enrollment patterns and trends. The methodological process begins with the identification of the problem, followed by data collection and preparation of the data set. Furthermore, an AI-assisted analytics model was developed to support pattern recognition and forecasting. In this study, Gemini AI is used to improve the analysis process by enabling efficient data processing, pattern detection, and insight generation. In addition, semantic layer integration is implemented to improve the contextual understanding of the data, allowing AI systems to generate more accurate and relevant outputs that align with institutional needs.

### Validation

To ensure the validity and reliability of the dataset, several data validation procedures are performed. First, data cleansing is done to address missing values, eliminate duplicate records, and fix inconsistencies in the dataset. Second, consistency checks are applied across different time periods to ensure uniformity in data reporting and measurement. In addition, outlier detection techniques are used to identify and manage extreme values that can favor the analysis. The use of standardized and institutionally verified data sources further strengthens the reliability and reproducibility of research findings. These validation

steps ensure that the dataset is accurate, complete, and suitable for subsequent analysis.

**Analysis Technique**

The analysis is conducted using growth ratio techniques to examine trends in new student enrollment and generate forecasts of future patterns. The growth ratio is calculated to measure the rate of change in enrollment between two consecutive periods, defined as follow:

$$GR_t = \frac{E_t - E_{t-1}}{E_{t-1}} \text{X } 100\ \%$$

Where $GR_t$ represents the growth ratio at time $t$ , $E_t$ denotes the number of enrollments in the current period, and $E_{t-1}$ represents the number of enrollments in the previous period. This formulation allows for the quantification of relative changes in enrollment over time to estimate future trends, the study applies a simple projection approach based on the average growth rate, expressed as:

$$GR = \frac{1}{n}\sum_{t-1}^{n} GR_t$$

Future enrollment is then forecasted using:

$$E_{t+1} = E_t \text{ x } (1 + GR)$$

This approach provides a basic predictive model for identifying potential enrollment trajectories. To improve analytical performance, the study integrates artificial intelligence, specifically Gemini AI, as a data mining tool. AI systems facilitate the identification of hidden patterns and relationships in large data sets, thereby increasing the depth and efficiency of analysis. Additionally, the use of semantic layer integration allows for more context-aware data processing, resulting in more accurate and meaningful insights. For data visualization and user interaction, AI-assisted features in Looker Studio are used. This allows for the presentation of analytical results in an intuitive and accessible format, supporting faster interpretation and more effective decision-making. Overall, the combination of growth rate analysis and AI-based data mining provides a robust, technical-based framework for forecasting and strategic planning in higher education enrollment management.

## RESULTS AND DISCUSSION

The data collection method in this study is carried out through dataset processing that contains information about the number of new students at the Universitas Terbuka as well as national data related to the number of high school graduates in the 2022–2024 period. The data used includes quantitative and numerical information nationally. The projection of student absorption is based on the assumption of the percentage growth of absorption per year. In addition, this study also utilizes Gemini AI as a tool in the process of data calculation and coding so that it can improve the accuracy and efficiency of the analysis carried out.

**JASa (Jurnal Akuntansi, Audit dan Sistem Informasi Akuntansi)** p-ISSN : 2550-0732
Volume 10, No. 1 / April 2026, p. 026-036 e-ISSN : 2655-8319

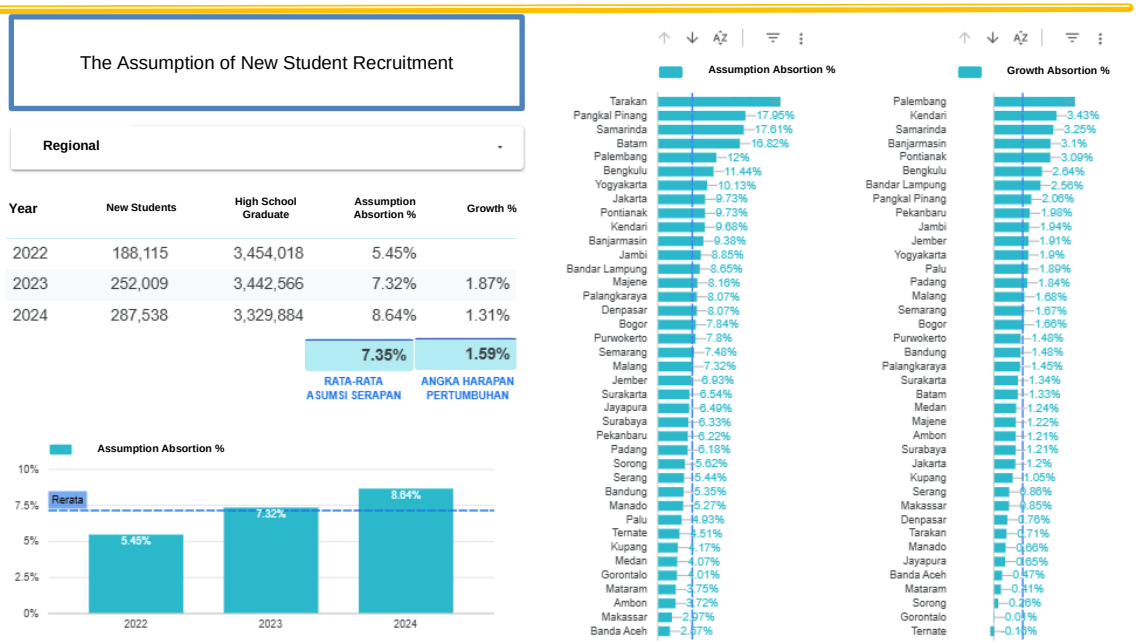


**Figure 2. The Assumption of New Student Recruitment**
Source: Data Process (2026)

Based on the graph image above, it can be shown that the trend of the number of new students has increased, this shows that the public's interest in continuing higher education through UT continues to grow, although the percentage increase is relatively moderate. This is in line with the trend of the need for distance learning-based flexible education. The absorption percentage of high school graduates by UT is still in the range of 5-8%. In 2022 it will be in the range of 5.64%, in 2023 it will be around 7.30% and in 2024 it will be at 8.15%. Although the absolute number of new students has increased, the proportion of absorption of the total national high school graduates is still relatively small. This is a challenge as well as an opportunity for UT to expand its reach and increase promotional campaigns among new graduates.

When viewed from annual growth, the Growth Percentage in 2022–2023 is 1.87% and the Growth in 2023-2024 decreases to 1.35%. This indicates that the growth of new students is starting to slow down even though the number is increasing. Some of the possibilities of slowdown can be caused by competition factors with other universities, limited digital infrastructure in some regions, or socio-economic factors. Based on Assumption of Absorption and Growth per Region On the right graph, it can be seen that the ranking of the regions with the highest absorption is in the eastern region of Indonesia, namely West Papua, Mountainous Papua, Central Papua, South Papua, North Maluku, Papua, NTT, West Sulawesi, Maluku, and Gorontalo in the top 10 of the highest absorption. On the other hand, the areas with relatively low absorption are large provinces such as East Java, West Java, Central Java, DKI Jakarta, and Banten. The graph shows that the Universitas Terbuka has experienced a growth in the number of new students from year to year with an average absorption of 7.63% of high school graduates, although growth is starting to slow down. The highest absorption actually comes from provinces in the Eastern region of Indonesia, which shows the importance of UT as an educational solution in areas with limited access.

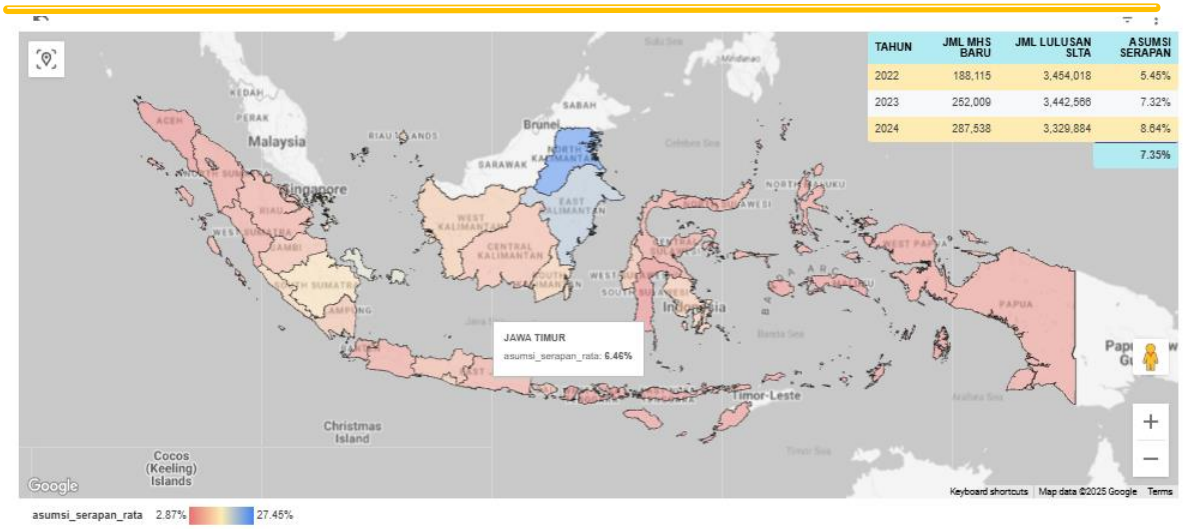


**Figure 3. Map of the assumption of new student absorption - province**
Source: Data Process (2026)

The graph shows that the Universitas Terbuka has experienced an increasing trend in the number and proportion of new students in the last three years, with an average absorption of 7.35% of high school graduates. The distribution map shows that provinces in the eastern region of Indonesia tend to have a higher absorption rate, while provinces in Java and Sumatra are relatively low despite the large number of high school graduates. This condition provides a strategic opportunity for UT to strengthen branding in densely populated areas while improving education services in areas with high absorption so that growth remains sustainable. The distribution on the map of Indonesia shows variations in the absorption of new students between provinces or regional UTs. The provinces with high absorption (dark blue color) are in the North Sulawesi, Gorontalo, North Maluku, and parts of West Papua. This shows that UT is the main choice of higher education in areas with limited access to face-to-face universities. Provinces with medium absorption (light yellow orange) such as Central Kalimantan, Central Sulawesi, and East Nusa Tenggara, still show significant contributions. Provinces with low absorption (pink to dark red color) are generally in Java, Sumatra, and Papua. The low absorption in Java and Sumatra can be caused by the large number of conventional university options, while in Papua it can be related to geographical factors, internet access, and limited learning infrastructure.

In the application, users can select each geographical location of the map to see the percentage of absorption in each area. For example, if the user chooses the province of East Java, a percentage of the absorption assumption value will appear of 6.46%. The application visually makes it easy for users to find out the percentage of absorption of the number of new students in each region in Indonesia.

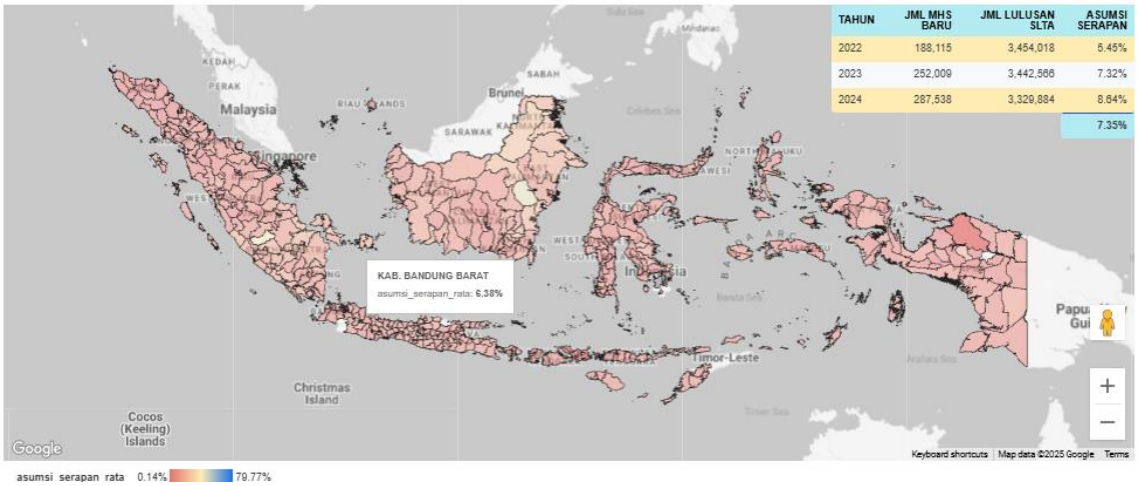


**Figure 4. Map of the assumption of new student absorption – UT Region**
Source: Data Process (2026)

This application also makes it easier for users to see the average absorption assumptions per pokjar or salut. The data is more condensed to see how much the absorption percentage is for each salut or per city district. For example, if the user chooses pokjar or salut in West Bandung district, an average absorption assumption value percentage of 6.38% will appear.

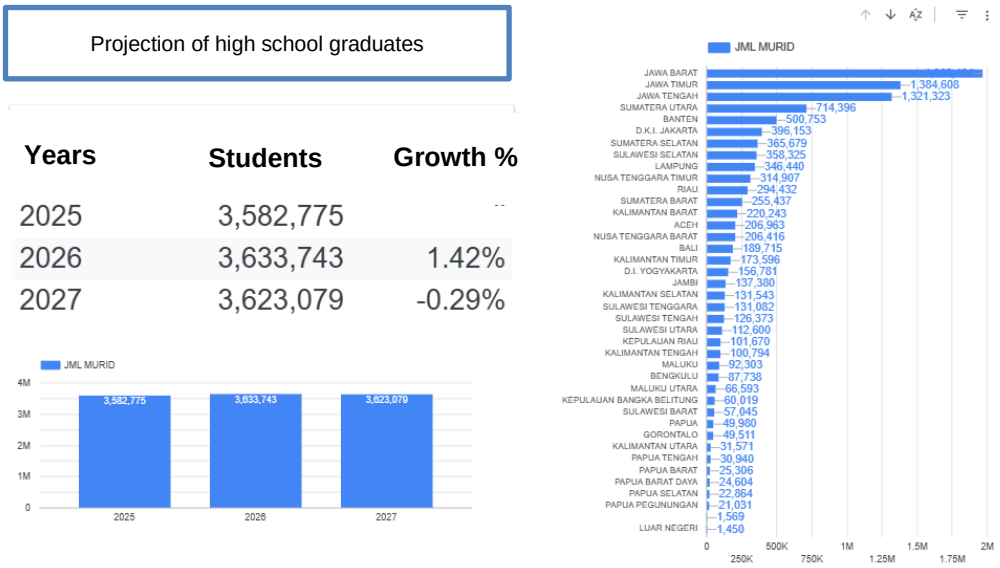


**Figure 5. Projection of high school graduates**
Source: Data Process (2026)

Based on the projected trend of the number of graduates in the next 3 years. The number of high school graduates was relatively stable in the 2025-2027 period, with a slight increase in 2026 and then a slight decrease in 2027. In 2025 it is projected that there will be 3,582,775 high school graduates, in 2026 it will increase to 3,633,743 graduates (growth +1.42%) and in 2027 it will decrease slightly to 3,623,079 graduates (growth -0.29%). This shows the saturation of the number of graduates, so that universities, including the Universitas Terbuka (UT), can no longer rely only on the natural growth of high school graduates, but must strengthen recruitment strategies.

New Student Projections

UT DAERAH -

| RATA2 ASUMSI SERAPAN 2022-2024 | RATA2 PERTUMBUHAN 2022-2024 | ASUMSI SERAPAN 2025 | ASUMSI SERAPAN 2026 | ASUMSI SERAPAN 2027 |
|---|---|---|---|---|
| 7.35% | 1.59% | 8.94% | 10.54% | 12.13% |
| (1) | (2) | (3) = (1) + (2) | (4) = (3) + (2) | (5) = (4) + (2) |

| JML LULUS SLTA 2025 | JML LULUS SLTA 2026 | JML LULUS SLTA 2027 |
|---|---|---|
| 3,581,374 | 3,632,651 | 3,622,553 |

| PROYEKSI MHS BARU 2025 | PROYEKSI MHS BARU 2026 | PROYEKSI MHS BARU 2027 |
|---|---|---|
| 320,314 | 382,819 | 439,513 |

**Figure 6. New Student Projections**
Source: Data Process (2026)

Based on the projection table for new students of the Universitas Terbuka (UT) in 2025–2027, it was compiled using historical data on the average absorption of new students for the 2022-2024 period which is 7.35% and an average growth of 1.59% per year. By combining these two components, the assumption of new student absorption is obtained of 8.94% for 2025, which then increases gradually to 10.54% in 2026 and 12.13% in 2027. This increase shows a positive trend in UT's ability to absorb high school graduates as new students. The calculation is made in an application with the help of Artificial Intelligence Gemini in the calculation process to improve the output of the results.

In addition, this table also presents projected data on the number of national high school graduates that are the basis for the calculation, namely 3,581,374 graduates in 2025, 3,632,651 graduates in 2026, and 3,622,553 graduates in 2027. By multiplying the number of graduates by the percentage of absorption assumptions, an estimate of the number of new students who will enter UT is obtained. The projection results show a consistent increase, from 320,314 students in 2025 to 382,819 in 2026, and reaching 439,513 students in 2027.

Overall, this projection illustrates the potential for significant growth in the number of new students at UT in the next three years. This upward trend can be the basis for UT's strategic planning, such as adjusting service capacity, developing online learning infrastructure, as well as more aggressive promotional strategies to reach more high school graduates. If the realization of absorption is in accordance with the projection, UT will experience a large surge in the number of new students and needs to ensure academic readiness and services so that this growth can be managed properly.

## CONCLUSION

Based on the projected results of the absorption of new students at the Universitas Terbuka (UT) for the 2025–2027 period, an overview of a significant upward trend in the number of new students, from 320,314 students in 2025 to 439,513 students in 2027. This projection is produced by utilizing Artificial Intelligence (AI) technology used to process historical data on average student absorption and average growth for 2022–2024, resulting

in more accurate and data-driven estimates. With AI support, the analysis process can be done faster, with fewer manual errors, and allows for automatic testing of various absorption assumption scenarios. The use of AI in the creation of these projection applications not only helps with the calculation of absorption figures, but also provides an easy-to-understand and interactive visualization of the data. This supports policymakers at UT to make more appropriate strategic decisions, such as service capacity planning, tutor recruitment, and determination of promotion targets in various regions. Thus, UT can optimally utilize big data to increase the effectiveness of new student growth strategies.

Overall, the application of AI in the calculation process and projection application is proof that UT has utilized cutting-edge technology to support evidence-based planning. With this approach, UT has a greater chance to achieve the target of increasing the number of new students while maintaining the quality of the educational services provided. AI integration is expected to be the foundation for UT's digital transformation in the future, creating an adaptive and sustainable planning system.

## REFERENCES

Alvi, M. (2016). *A manual for selecting sampling techniques in research*. https://mpra.ub.uni-muenchen.de/id/eprint/70218

Bustami, B. (2013) Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, TECHSI : *Jurnal Penelitian Teknik Informatika,* Vol. 3, No.2, Hal. 127-146. https://doi.org/10.29103/techsi.v5i2.154

Gitman, L. J., & Zutter, C. J. (2012). *Principles of Managerial Finance*. Pearson

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Han, J., Kamber, M., Berzal, F., & Marín, N. (2002). Data Mining: Concepts and Techniques. *SIGMOD Record*, *31*(2), 66–68. https://doi.org/10.1145/565117.565130

Han, J., Kamber, M., kaufmann, D. (2006). Concepts and techniques. *Liacs.Leidenuniv.Nl*.

Jantawan, B., & Tsai, C. F. (2014). A comparison of filter and wrapper approaches with data mining techniques for categorical variables selection. International Journal of Innovative Research in Computer and Communication Engineering, 2(6), 4501-4508.

Kotler, P. (2012). *Kotler on marketing*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Patil, T. R., & Sherekar, M. S., (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, International Journal of Computer Science and Applications, Vol. 6, No. 2, Hal 256- 261.

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Prentice Hall

Santosa, B., Conway, T., & Trafalis, T. (2007). A hybrid knowledge based-clustering multi-class svm approach for genes expression analysis. *Springer Optimization and Its Applications*, *7*, 231–274. https://doi.org/10.1007/978-0-387-69319-4_15

Singh, S. (2013). Performance analysis of engineering students for recruitment using classification data mining techniques. *Ijcset.Net*. Retrieved November 29, 2023, from http://ijcset.net/docs/Volumes/volume3issue2/ijcset2013030202.pdf

Sivaram, N., & Ramar, K. (2010). Applicability of clustering and classification algorithms for recruitment data mining. International Journal of Computer Applications, 4(5), 23-28. https://ui.adsabs.harvard.edu/abs/2010IJCA....4e..23S/abstract

Witten, D. M. (2011). Classification and clustering of sequencing data using a Poisson model. *The Annals of Applied Statistics*, *5*(4), 2493–2518. https://doi.org/10.1214/11-AOAS493

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of

research on artificial intelligence applications in higher education–where are the educators?. International journal of educational technology in higher education, 16(1), 39. https://link.springer.com/content/pdf/10.1186/s41239-019-0171-0%E2%80%8C.pdf