JASa (Jurnal Akuntansi, Audit dan Sistem Informasi Akuntansi)
ISSN 2550-0732 print / ISSN 2655-8319 online
Vol. 9 No. 1 / April 2025
DOI;10.36555/jasa.v9i1.2799

Educational Data Mining in Accounting: Market Segmentation Strategy for Financial Management of New Student Admissions

Diky Paramitha*[1], Novita Nugraheni[1]
Universitas Terbuka, Indonesia[1]
*Coresponding Email : dikyparamitha@ecampus.ut.ac.id

ABSTRACT

In higher education, effective financial management is a key factor in ensuring institutional sustainability, including at Universitas Terbuka. One of the main challenges in financial management is the ability to accurately analyze and predict new student enrollment. This study applies an Educational Data Mining (EDM) approach in accounting to develop market segmentation strategies that enhance financial management efficiency at Universitas Terbuka. The methods used include Growth Ratio and Naïve Bayes Classifier (NBC), utilizing data obtained from the national Senior High School database managed by the Indonesian Ministry of Education, Culture, Research, and Technology. The analysis results indicate that with the application of Growth Ratio calculations, the national SLTA graduate absorption rate is projected to increase by 1.2% annually. Based on this trend, the projected absorption rate of SLTA graduates into Universitas Terbuka is expected to grow to 7.2% in 2024, 8.4% in 2025, and 9.6% in 2026. These findings reflect a positive growth trend in new student enrollment, providing a strategic basis for Universitas Terbuka in making data-driven financial and resource management decisions. Thus, the implementation of Educational Data Mining techniques has the potential to be an innovative solution for supporting data-driven financial planning in higher education.

ARTICLE INFO

Article history:
Submitted: 19 April 2025
Revised: 21 April 2025
Accepted: 29 April 2025
Published: 30 April 2025

Keyword:
Educational Data Mining,
Financial Management,
Market Segmentation,
Growth Ratio.

To cite this article (APA Style): Paramitha, D., and Nugraheni, N., (2025). Educational Data Mining in Accounting: Market Segmentation Strategy for Financial Management of New Student Admissions. JASa : Jurnal Akuntansi, Audit dan Sistem Informasi Akuntansi. Vol 9 (1), p.117-125. https://doi.org/10.36555/jasa.v9i1.2799

INTRODUCTION

In the era of increasingly advanced technology, especially the development of Artificial Intelligence, it makes it easier to complete work, especially those related to data. In the digital era, effective financial management is a key factor for the sustainability of higher education institutions. The level of accuracy of a data is needed in daily life. Every information that exists is important to determine every decision in a certain situation. One of the main challenges faced by universities is the uncertainty in new student enrollment, which impacts budget planning and resource allocation (Altbach et al., 2019).This causes the provision of information to be a means to be analyzed and summarized into knowledge from data that is useful when making a decision. Data mining is the extraction to obtain important information that is implicit and previously unknown from a data (Witten, 2011). Data mining is an analytical step for Knowledge Discovery in Databases (Zhao, 2020). Data mining aims to uncover patterns and other valuable information by sorting through large data sets (Lu et al., 2016). By using data mining, each data collection or warehouse can provide important knowledge that becomes very valuable information for an organization, such as in an educational organization. The recruitment process is one of the important functions of the human resources department and is the first step towards the creation of competitive strength, while in campus recruitment it is the main recruitment mode for new talented graduates (Sivaram et al., 2010).

Currently, many universities are implementing education by implementing a distance learning system because after the COVID-19 pandemic and the community's attitude towards education that implements a distance learning system has become better. 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 needed to improve the accuracy of marketing in an organization or company so that it can make it easier for businesses to run marketing programs or approach customers (Kotler, 2012). Many educational institutions possess *big data* related to student admissions but often lack data analysis capabilities that support strategic decision-making (Siemens et al., 2011). In managing and developing organizational performance strategies, in-depth analysis is required to encourage better decision making in designing and implementing strategies to improve organizational performance (Larasati et al., 2024).

Traditional methods of predicting new student enrollment still rely on simple historical trends or intuition-based approaches. Currently, Universitas Terbuka (UT) has begun implementing a data mining-based approach to improve efficiency and accuracy in predicting new student enrollment. One such initiative is a study conducted at UPBJJ-UT Central Sulawesi, which uses the Neural Network (NN) algorithm with the backpropagation method to model predictions of new student enrollment in basic education programs. The results of the study show that this algorithm is able to provide predictions with a low error rate (Syahrullah et al., 2016). However, with the increasing availability of data, *Educational Data Mining* (EDM) has emerged as an innovative solution to analyze enrollment patterns and assist in university financial management (Baker et al., 2009). EDM enables universities to identify potential market segments, analyze prospective students' behavior, and predict enrollment rates based on historical patterns and external factors (Romero et al., 2020).

This study aims to apply the *Growth Ratio* and *Naïve Bayes Classifier* (NBC) methods to classify prospective Universitas Terbuka students based on national Senior High School (SLTA) data managed by the Indonesian Ministry of Education, Culture, Research, and Technology (KemendikbudRistek). By utilizing these techniques, the university is expected to improve the accuracy of new student enrollment projections, thereby strengthening financial strategies and optimizing resource allocation (Dutt et al., 2017). The study used the Naïve Bayes algorithm to analyze new student registration data, with the aim of identifying prospective students who have the potential not to re-register (Saputri et al., 2021). Naïve Bayes is used to determine a more effective and efficient new student admissions promotion strategy (Mirza, 2019). The application of Naïve Bayes to predict student graduation based on academic data, shows that this method is effective in identifying students at risk of not graduating on time (Cahyo, 2024). Although various studies have applied the Naïve Bayes algorithm in the context of education, there are several research gaps, including: Research on the application of data mining in the context of open universities or distance education is still limited, even though its characteristics are different from conventional education and the use of methods such as growth ratio is still dominant in predicting new student registration, but is less able to capture the complex variables that influence prospective student decisions.

Educational Data Mining (EDM) has become a crucial approach for analyzing large datasets in education to improve decision-making processes. EDM applies data mining techniques to identify patterns, predict student behaviors, and optimize academic and administrative strategies (Romero et al., 2020). The implementation of EDM in higher education institutions has been widely researched, particularly in enrollment prediction and financial management. Emphasized the importance of using analytics to penetrate the complexity of student data and enhance institutional sustainability (Siemens et al., 2011). Data Mining is often referred to as Knowledge Discovery in Database (KDD), which is an

activity that includes the collection, use of historical data to find regularity, relationship patterns in very large data sets. The output of this data mining can be used to determine policies or decisions in the future (Santosa et al., 2007).

Predicting student enrollment is essential for financial planning and resource allocation in universities. Various models have been developed to forecast student admissions, including regression analysis, machine learning algorithms, and Bayesian classification. Machine learning techniques, such as Naïve Bayes Classifier (NBC), can significantly improve enrollment prediction accuracy (Dutt et al., 2017). Using Growth Ratio models can provide insights into enrollment trends based on historical data. These methods allow institutions to anticipate student intake and plan their financial strategies accordingly (Peña-Ayala, 2014).

Financial sustainability in universities is closely linked to market segmentation and enrollment forecasting. A well-structured segmentation strategy enables institutions to target potential students effectively and allocate resources efficiently (Altbach et al., 2019). Studies have shown that analyzing student demographics, geographical distribution, and economic factors helps universities refine their recruitment strategies. Integrating EDM with financial planning could optimize budget allocation and reduce financial risks associated with fluctuating enrollment rates (Romero et al., 2020).

Growth Ratio and NBC are widely used techniques in data mining for classification and prediction. The Growth Ratio method helps in understanding enrollment trends and projecting future student intake based on historical patterns (Baker et al., 2009). On the other hand, NBC is an effective probabilistic model for classification tasks, allowing institutions to segment student markets and predict their likelihood of enrolling (Peña-Ayala, 2014). Combining these techniques enhances the accuracy of financial projections and ensures a more data-driven approach in educational institutions. Naive Bayes is a simple probabilistic classifier that calculates a set of probabilities by summing up combinations of frequencies and values from a given dataset. The algorithm uses Bayes' theorem and assumes all independent or non-interdependent attributes given 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 British scientist Thomas Bayes, predicting future odds based on previous experience (Bustami., 2013).

Despite extensive research on EDM and predictive analytics in education, most studies focus on academic performance analysis rather than financial planning and market segmentation. This study aims to bridge this gap by applying EDM specifically to financial forecasting and strategic planning for new student enrollment. By leveraging Growth Ratio and NBC techniques, this research provides a comprehensive model for predicting enrollment rates and optimizing financial management at Universitas Terbuka.

METHODS

The research uses a quantitative approach that aims to achieve an understanding of how it is properly constructed and built and how it works (Berndtsson et al., 2008). Educational Data Mining (EDM) method to analyze the segmentation pattern of prospective new students of Universitas Terbuka based on national high school data. The framework for this research is shown in Figure 1.

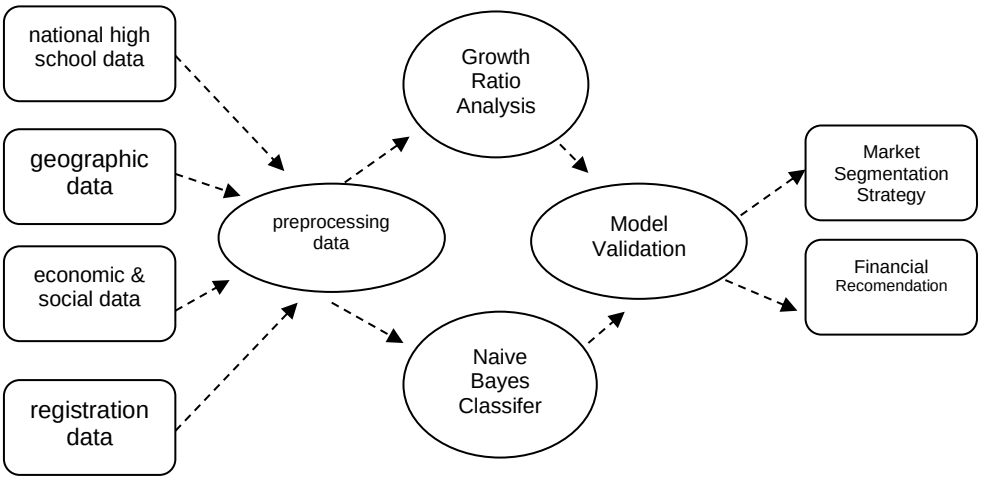

Figure 1. Conseptual Framework
Source: Processing Data (2021-2024)

This study consists of several main stages, namely data collection, data processing, data analysis using Growth Ratio and Naïve Bayes Classifier (NBC), and evaluation of model results. The type of data used in this study uses primary and secondary data. Primary data in this study were obtained directly through the process of interviews, observations, triangulation, surveys and questionnaires. As for secondary data, it was obtained through coordination with the data team of the Ministry of Education, Culture, Research and Technology. Meanwhile, secondary data was obtained through coordination with the data team of the Ministry of Education, Culture, Research and Technology. The data uses basic data of high school educators related to student graduation data. The data was processed using the growth ratio formula to see the growth of the data. In this study, the data used is data from the Ministry of Education and Culture's National High School graduates, and is time series data from 2021 to 2023 and projection data for the next 3 years, namely 2024 to 2026. In looking at the growth of data in each series, the growth ratio calculation is used to see the percentage growth of data variables in a period of time. The growth ratio equation is as follows:

Growth Ratio = ((Present – Past)/Past) x 100%

(Garrison et al., 2018)

Present is the value of a current data or present value, while Past is the value of a data in the past. The result of the growth ratio is a reflection of the growth of a data. This study uses a data mining approach to explore patterns and trends from historical data of high school graduates. One of the methods used is the Naïve Bayes Classifier, a classification algorithm based on Bayes' Theorem. This method works by calculating the probability of data belonging to a certain class based on the values of the attributes it has.

The target population refers to all the members who meet certain criteria specified for the investigation of the research whereas the Sample can be defined as a group of people selected from a comparatively smaller number of the population for the purpose of the investigation and the Members of the sample are referred to as participants (Alvi, 2016). The population in this study is high school nationally. Non-probability Sampling Methods are sample selection that is carried out on the basis of subjective assessment from the researcher, Non-probability Sampling is very suitable for exploratory research that is intended to produce new ideas that will be tested systematically, non-probability techniques allow sampling from the population whose elements are unlimited in number (Alvi, 2016).

The sampling used is non-probability sampling methods with research samples in schools that are included in the Universitas Terbuka market segment.

In data analysis, the steps used in this study use analysis with data collection at the beginning of the study, reduction of collected data, data processing of data presentation of reduction results, and finally drawing conclusions or called verification. Conduct study literature in support of research as a reference that will be used to obtain information related to research supporting theories (Suwayudhi et al., 2022). observation through data search that is carried out openly or closed in the first information retrieval by observing the object to be researched (Senika et al., 2022).

RESULTS AND DISCUSSION

The time series data processed is data on National High School graduates and data on new students of open universities for the last 3 years with the period 2021 to 2024. The data is divided into components to be used as projections for the future. The time series is based on data that has the same components in time, namely semesters and years, so it is expected to produce forecasts for the next year. The following are the results of data cleansing trends for the absorption of high school graduates per Regional Universitas Terbuka

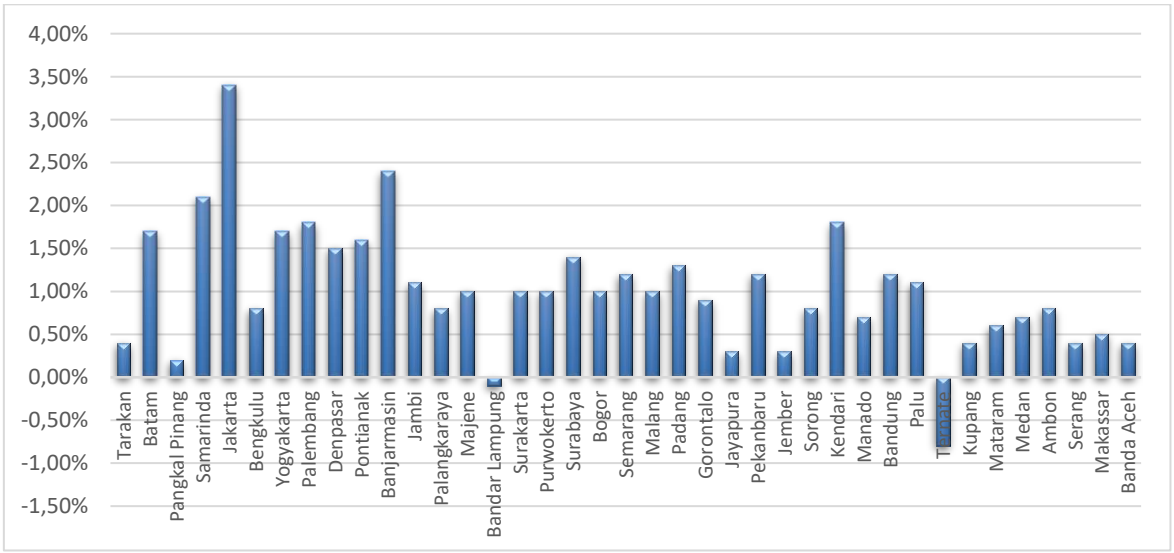

Figure 2. Average Absorption Data
Source: Processing Data (2021-2024)

Based on the results of data processing shown in the average development graph of the absorption rate, the highest absorption rate of high school graduates is in the UT Jakarta area with an average absorption rate of 3.40% during the 2021-2024 period, and the second and third places are in the UT Banjarmasin and UT Samarinda areas with an average absorption rate of 2.4% and 2.1%. But not all regional UTs experienced an increase in that period, there were several regions that experienced a decrease in the absorption rate of high school graduation during that period such as what happened at UT Ternate which experienced a decrease with an average absorption rate of high school graduation of -0.8% and UT Bandar Lampung by -0.1%. It can be seen that a large number of UT Regions have a positive absorption rate, but there are several cities that show low or even negative absorption. Absorption between each Regional UT varies greatly with significant differences between the regions with the highest and lowest absorption. If viewed from the entire regional UT, the average high school graduation absorption rate nationally is 1.2% with details of the period 2021-2024 as shown in figure 2.

Table 1. Table. Comparison of High School Graduates and UT New Students

Graduation year SLTA	Amount	Number of New Students	% Absorption of High School Graduates	Average Absorption Rate
2021	3.366.588	123.078	3.7%	
2022	3.463.453	160.955	4.6%	1.2%
2023	3.449.918	207.187	6.0%	

Source*:* data that has been processed by the author (2021-2024)

Based on the data shown in the table above, UT's absorption rate nationally during the 2021-2023 period has increased with details in 2021 the number of graduates was 3.36 million students with the number of new students in that year of 123,078 or 3.7% of the absorption of high school graduates nationally. In 2022, the number of high school graduates increased to 3.46 million students, when viewed from the absorption trend in UT new students from the previous year there was an increase of 0.9% so that the % absorption became 4.6% in 2022. In 2023, high school graduates experienced a downward trend to 3.44 million students, when viewed from the absorption trend in UT new students, there was an increase in absorption with the number of new students at 207,187 or an increase of 1.4% from 2022 to 6.0%.

There are several analyses that can be concluded from the data Increase in the Number of New Students: There is an increase in the number of new students entering higher education every year, from 123,078 in 2021 to 207,187 in 2023. This shows an increase in the acceptance capacity or interest of high school graduates to continue to higher education. Increase in Absorption Percentage: There was also a significant increase in absorption percentage from 3.7% in 2021 to 6.0% in 2023. This may reflect better policies in the absorption of high school graduates into college. Number of High School Graduates: Although the number of high school graduates has been fairly consistent from year to year, with a slight decrease from 2022 to 2023, the percentage of absorption has continued to increase, indicating efficiency in the university admissions system. Overall, the data illustrates a positive trend, with the increasing number of high school graduates, the percentage of absorption into universities, and the number of new students projected to enter the Open University.

Table 2. Projection of UT High School Graduates and New Students

Graduation Year	high school graduate	Projection of High School Graduate Absorption	UT New Student Projections
2024	3,331,960	7.2%	239,901
2025	3,579,941	8.4%	300,715
2026	3,626,150	9.6%	348,110

Source*:* data that has been processed by the author (2021-2024)

Based on the calculation of the projected trend of absorption of high school graduates for the next 3 years, it is obtained that the projected absorption value of high school graduates which will later be absorbed by UT is 7.2% in 2024 or as many as 239,901 students, the projected absorption value has increased to 8.4% in 2025 or as many as 300,715 students and 9.6% in 2026 or as many as 348,110 students. The average annual increase trend during the 2024-2026 projection period is 1.2%. There are several analyses that can be concluded from the data Increase in the Number of High School Graduates: There is an increase in the number of high school graduates from year to year, with relatively stable growth. From 3,331,960 graduates in 2024, it is expected to increase to 3,626,150 in 2026. Increase in Absorption Percentage: The percentage of absorption of

high school graduates who enter higher education has also increased significantly from 7.2% in 2024 to 9.6% in 2026. This shows that there is a trend of increasing interest or accessibility in higher education among high school graduates.

UT New Student Projected Growth: The Open University is projected to admit more new students each year. In 2024, it is estimated that there will be 239,901 new students, and this number continues to increase until it reaches 348,110 in 2026. This growth could reflect UT's growing popularity as a higher education option or capacity building that can accommodate more students.

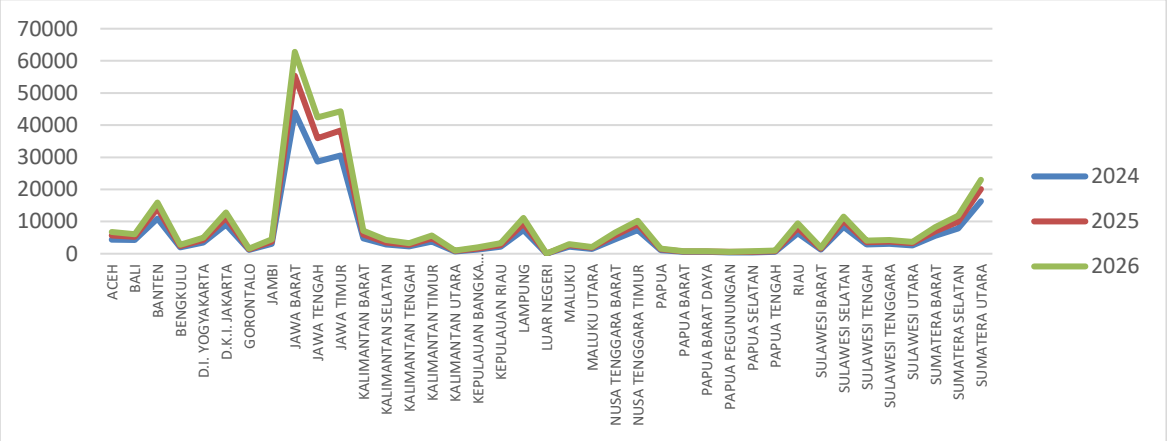

Figure 3. Projected Absorption of High School Graduates in 2024-2026
Source: Processing Data (2021-2024)

Based on the calculation of the projected trend of absorption of National High School graduates per region in Indonesia, it can be seen that the National High School graduates and the highest absorption during the projection period of 2024-2026 are found in 3 regions in Java Province, namely West Java, Central Java and East Java. With a projected absorption in 2024 of 7.2% in West Java Province, a total of 43,962 prospective new students were obtained, in 2025 it was 8.4% or a total of 55,368 prospective new students and in 2026 it was 9.6% or a total of 62,804 prospective new students in the projected absorption in that year. Meanwhile, the lowest absorption in 2024 is in the North Kalimantan region, which is a total of 653 prospective new students with a projection rate of 7.2%, in 2025 there are 863 prospective new students with a projection rate of 8.4%, and in 2026 there are 998 prospective new students with a projection rate of 9.6%.

This study demonstrates that the application of Educational Data Mining in accounting can be an effective strategy for market segmentation of prospective students at Universitas Terbuka. By utilizing national high school data from KemendikbudRistek, the Growth Ratio Analysis and Naïve Bayes Classifier (NBC) methods successfully projected enrollment trends more accurately.The analysis results indicate an increasing trend in the projected absorption rate of high school graduates, from 7.2% in 2024 to 9.6% in 2026. These findings provide valuable insights for Universitas Terbuka in developing data-driven marketing strategies and optimizing financial management for student admissions.

Furthermore, this study contributes to improving the efficiency of higher education planning through predictive analytics, allowing for better resource allocation. With the implementation of an automated *data mining* system, Universitas Terbuka is expected to enhance its competitiveness and long-term sustainability. However, this research has some limitations, such as dependence on historical data and the potential influence of external variables that were not considered. Therefore, future research is recommended to develop more complex models by incorporating broader social, economic, and educational policy factors. Based on the findings of this study, several recommendations can be implemented to increase student enrollment rates in the region. First, enhancing socialization and promotion by conducting intensive information campaigns about study

programs, facilities, and the benefits of studying in the region through social media, local media, and community events. Additionally, collaborating with high schools through seminars, workshops, or presentations to emphasize the importance of pursuing higher education at the Open University. Second, providing scholarships and financial aid, such as special scholarships for prospective students from underserved areas, tuition fee waivers, living allowances, and transportation cost assistance for financially disadvantaged students. Third, improving the quality and relevance of study programs by offering courses that align with the local job market's needs and periodically updating the curriculum to stay relevant to the latest industrial and technological advancements. Fourth, developing infrastructure and educational facilities by upgrading laboratories, libraries, and study rooms while also improving internet connectivity and transportation services to support student mobility. By implementing these measures, it is expected that student participation in higher education will increase significantly

CONCLUSION

Based on the data from the discussion, in the context of new student admissions, market segmentation through the Educational Data Mining (EDM) approach plays an important role in supporting the financial management of higher education institutions. This segmentation allows universities to understand the characteristics of prospective students from various groups, whether based on geographical, demographic, behavioral, or psychographic, in order to design appropriate strategies in resource allocation and efficient financial policy development. From here, financial management can design promotion strategies and larger investments in areas that have high potential but low conversion rates, for example through the opening of Open University service centers in areas with the most high school graduates and optimizing the distribution of region-based promotional budgets. The data of high school graduates can be segmented by age, gender, school background and socioeconomic status. This is very relevant for financing policies such as the determination of the amount of scholarships or subsidies based on economic needs and the offer of certain study programs for certain age groups or majors that have high purchasing power potential, which so far the Open University still has limited data in the determination of scholarships. With accurate segmentation, in the context of financial management, the data is expected to optimize promotional spending, design flexible financing schemes based on the purchasing power of certain segments, manage revenue projections based on regional segmentation and study interests and improve the efficiency of allocating academic development budgets according to student market preferences. With this approach, financial management strategies are no longer only reactive, but proactive and data-driven, in line with the direction of higher education's digital transformation.

REFERENCES

Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2019). *Trends in Global Higher Education: Tracking an Academic Revolution*. UNESCO

Alvi, M. (2016). *A manual for selecting sampling techniques in research*

Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1(1)*, 3–17

Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). Developing your objectives and choosing methods. *Thesis projects*, 54–70. https://doi.org/10.1007/978-1-84800-009-4_8

Bustami, B. (2013). Penerapan algoritma Naive Bayes untuk mengklasifikasi data nasabah asuransi. *TECHSI: Jurnal Penelitian Teknik Informatika, 3(2)*, 127-146

CAHYO, M. (2024). *ANALISIS PREDIKSI KELULUSAN MAHASISWA DENGAN METODE*

NAIVE BAYES CLASSIFIER (Studi Kasus: Program Studi Diploma III Teknologi Bank Darah di Poltekkes BSI Yogyakarta) (Doctoral dissertation, Universitas Teknologi Digital Indonesia)

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access, 5*, 15991–16005

Garrison, R. H., Noreen, E. W., & Brewer, P. C. (2018). *Managerial accounting*. McGraw-Hill

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

Larasati, D., Tanzil, N. D., Alfian, A., & Wardani, L. (2024). Business Intelligence Dashboard for Financial Performance Analysis of Public Service Agency Using Microsoft Power BI. *JASa (Jurnal Akuntansi, Audit dan Sistem Informasi Akuntansi)*, *8*(2), 491-499. https://doi.org/10.36555/jasa.v8i2.2649

Lu, H., Setiono, R., & Liu, H. (1996). Effective data mining using neural networks. *IEEE transactions on knowledge and data engineering*, *8*(6), 957-961

Mirza, A. H. (2019). Application of naive bayes classifier algorithm in determining new student admission promotion strategies. *Journal of Information Systems and Informatics*, *1*(1), 14-28

Patil, T. R., & Sherekar, M. S. (2013). Performance analysis of Naïve Bayes and J48 classification algorithm for data classification. *International Journal of Computer Science and Applications, 6(2)*, 256-261

Peña-Ayala, A. (2014). *Educational Data Mining: Applications and Trends*. Springer

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3)*, e1355

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

Saputri, L. N., Asroni, A., & Riyadi, S. (2021). Prediction of New Student Registration at Universitas Muhammadiyah Yogyakarta using the Naïve Bayes Classification

Senika, A. (2022). Implementasi metode Naïve Bayes dalam penilaian kinerja sales marketing pada PT. Pachira Distrinusa. *Jurnal Media Informatika Budidarma, 6(1)*. https://doi.org/10.30865/mib.v6i1.3331

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46(5)*, 30–40

Sivaram, N. (2010). Applicability of clustering and classification algorithms for recruitment data mining. *Citeseer, 4(5)*, 975–8887

Suwayudhi, S., & Eka, I. (2022). Teknik klasifikasi dalam memprediksi penerimaan siswa baru menggunakan metode Naïve Bayes. https://doi.org/10.55123/jomlai.v1i3.963

Syahrullah, S., Ngemba, H. R., & Hendra, S. (2016). Pendaftaran Mahasiswa Baru Program Pendidikan Dasar Menggunakan Algoritma Nn Backpropagation Di Upbjj-Universitas Terbuka Provinsi Sulawesi Tengah. *SEMNASTEKNOMEDIA ONLINE*, *4*(1), 4-9

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