Jurnal Scientia
Volume 13, Number 04, 2024, DOI 10.58471/ scientia.v13i04
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

Utilization Of Big Data For Personalized Online Learning: An Empirical Study In Higher Education

Usanto[1*], Christine Sientta Dewi[2]
[1*,2]Program Studi Sistem Informasi, Institut Teknologi Dan Bisnis Swadharma

Article Info
Keywords:
Big Data,
Data Anayltic,
Online Learning,
Higher Education

ABSTRACT
This study explores the use of Big Data in personalizing online learning in higher education, focusing on student access and engagement patterns in e-learning platforms. The main problem faced is the inefficiency in monitoring student engagement, which impacts academic outcomes. The solution offered is learning analytics analysis using clustering and classification techniques to personalize learning materials. Data is taken from student activities on e-learning platforms for one semester. Data processing is done using machine learning tools such as K-Means Clustering and Decision Tree. The results show that active engagement in e-learning platforms is associated with better academic performance, where students with higher access frequencies tend to have better grades. The visualization graph shows the trend of access intensity in the evenings and weekends, as well as the positive relationship between access duration and exam scores. With a Big Data-based system, institutions can improve the online learning experience and provide more personalized recommendations to support student academic success

Corresponding Author:
Usanto
Program Studi Sistem Informasi, Institut Teknologi Dan Bisnis Swadharma
usanto.s@swadharma.ac.id

INTRODUCTION

In recent decades, the development of information and communication technology has had a significant impact on various sectors, including education. One of the technologies that is increasingly dominating is Big Data, which allows the analysis of large amounts of data to provide deeper insights. In the field of education, Big Data plays a vital role in personalizing the learning process through the analysis of learning patterns, student engagement, and their academic performance (Alfiah et al., 2023; Lin, 2024; Sopian et al., 2023; S. Usanto et al., 2024; U. Usanto et al., 2023). Online learning, which is becoming increasingly common in higher education institutions, generates large amounts of data that can be analyzed to improve the learning experience (Anin et al., 2023; Ibrahim et al., 2023; Sahyunu et al., 2023).

Universities implementing e-learning systems collect data from various student interactions with learning platforms, such as logins, access times, discussion participation, downloading materials, and exam scores. This data has the potential to be used to personalize each student's learning experience (Aristamy et al., 2021; Wiratama et al., 2022). One of the main challenges in online learning is ensuring that students remain actively engaged

Jurnal Scientia
Volume 13, Number 04, 2024, DOI 10.58471/ scientia.v13i04
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

throughout the semester, as such engagement has a direct impact on academic outcomes. Without sufficient engagement, students often struggle to achieve optimal learning outcomes.

In this context, Big Data can be an important solution to overcome the problems faced in monitoring and increasing student engagement on online learning platforms. By utilizing machine learning and learning analytics technology, educational institutions can analyze student behavior in more detail and personalize learning for each individual (Kurniati et al., 2021; Puspitasari et al., 2023; U. Usanto, 2022). This technology enables e-learning systems to automatically recommend the most relevant materials to students, based on analysis of their access and engagement patterns.

This study aims to develop and test a Big Data-based learning personalization model at one of the universities in Indonesia. The focus of the study is to study the relationship between students' access patterns to e-learning platforms and their academic outcomes, and to provide recommendations for interventions that can be carried out by teachers to improve student engagement and performance. By using Big Data Analytics (Dewi et al., 2024; Lin, 2024; Saputro et al., 2024; Wibowo & Kraugusteeliana, 2024), this research is expected to provide a real contribution to improving the quality of online learning in higher education.

METHODS

Data collection

The data used in this study were collected from student activity logs on the e-learning platform, namely Google Classroom, for a full semester. The data collected included student login activities, duration of platform access, number of materials downloaded, participation in discussion forums, quiz scores, and final exam scores. In addition, student demographic data such as age, gender, study program, and year of entry were also included to provide additional context to the analysis. There were 50 students participating in this study, with the amount of data collected reaching thousands of entries. Data was automatically retrieved from the e-learning system using automated data scraping techniques. Furthermore, raw data was processed using data analysis tools for data cleaning and analysis (Sudipa et al., 2023; Wiguna et al., 2022).

Data Cleaning and Processing

The first stage in data processing is cleaning, where duplicate data is removed, missing data is filled with mean values, and the data format is standardized to facilitate further analysis. After the cleaning process, the prepared data is then transformed into features relevant for analysis, such as login frequency per week, access duration per session, number of downloaded materials, and participation in discussion forums.

Data analysis

Two main methods used in data analysis K-Means Clustering is used to group students based on their access behavior to the e-learning platform. The features used in clustering include login frequency, average access duration, discussion participation, and the number of materials downloaded. This method produces three main groups of students: Active Group, Moderate Group, and Passive Group. The Decision Tree method is used to predict students'

Jurnal Scientia
Volume 13, Number 04, 2024, DOI 10.58471/ scientia.v13i04
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

academic outcomes based on their access patterns (Afifuddin & Hakim, 2023; Herlina, 2024). The independent variables used in this model include login frequency, discussion participation, amount of material downloaded, and average access duration.

RESULTS AND DISCUSSION

The analyzed data is then visualized using heatmap, bar chart, scatter plot, and line chart. These visualizations are used to provide further insight into student access patterns and their impact on academic performance.

Student Access Intensity

The heatmap graph below shows the intensity of student access to e-learning platforms based on hours and days of the week.

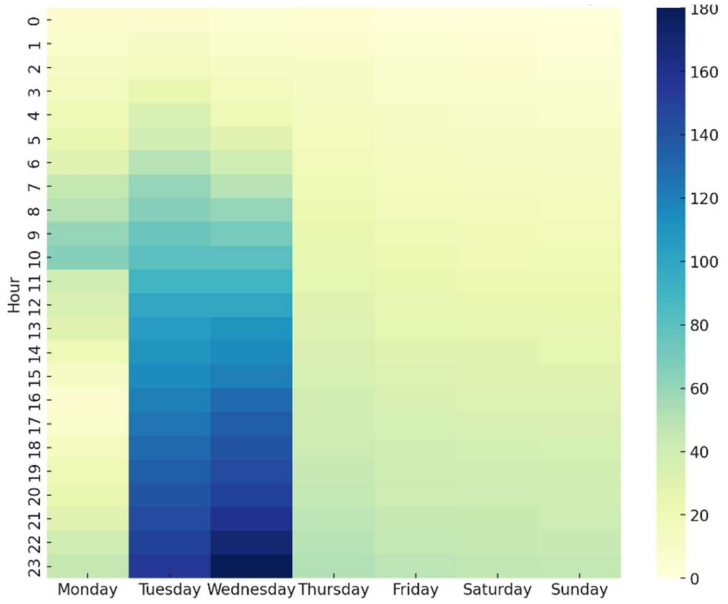

Figure 1. Intensity of Student Access to E-Learning Platforms

Based on Figure 1, it can be explained that the highest access occurs at night, especially between 19:00 and 23:00 on weekdays. This indicates that students access the platform more outside of work hours or formal lectures. On weekends, the intensity of access is higher throughout the day, indicating increased flexibility in study time.

Relationship between Access Frequency and Exam Scores

The following graph shows the correlation between the frequency of access to e-learning platforms and student exam scores.

Jurnal Scientia
Volume 13, Number 04, 2024, DOI 10.58471/ scientia.v13i04
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

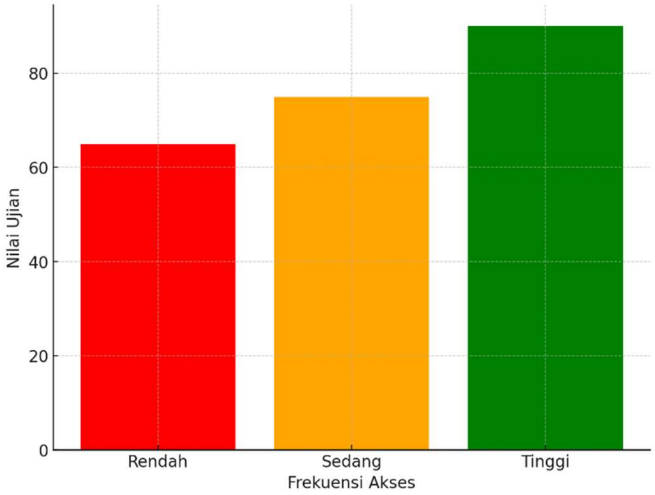

Figure 2. Relationship Between Access Frequency and Academic Grades

Based on Figure 2, it can be explained that students with higher access frequency tend to get better academic grades. Students who actively access the platform have higher average exam scores (90), while those who rarely access have lower scores (65).

Student Activity Trends During the Semester

This graph shows the trend of student activity over a semester.

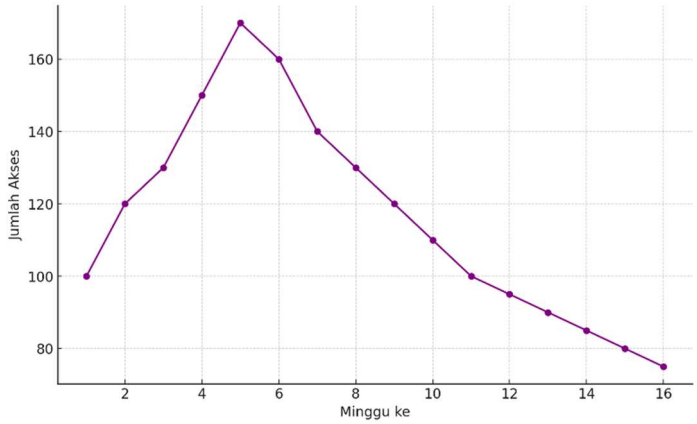

Figure 3. Student Activity Trends During the Semester

Based on Figure 3, it can be explained that this trend shows that student activity increases at the beginning of the semester, peaking in the 6th week, before starting to decline until the end of the semester. This decline may be caused by academic fatigue or decreased motivation approaching the final exam.

CONCLUSION

This study shows that Big Data has great potential in enhancing the personalization of online learning. Analysis of students' access patterns to e-learning platforms provides valuable insights into the relationship between student engagement and their academic performance.

Jurnal Scientia
Volume 13, Number 04, 2024, DOI 10.58471/ scientia.v13i04
ESSN 2723-7486 (Online)
https://infor.seaninstitute.org/index.php/pendidikan

Students who are more active and engaged in the platform tend to have better academic outcomes. In addition, data visualization helps educators and institutional administrators monitor student engagement and take faster and more appropriate intervention actions. By implementing Big Data-based systems, educational institutions can improve the online learning experience and provide more personalized recommendations to students, in order to support their academic success.

REFERENCE

Afifuddin, A., & Hakim, L. (2023). Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4. 5. *Jurnal Krisnadana*, *3*(1), 25–33.

Alfiah, F., Usanto, S., Setiadi, A., Supriadi, A., Suhanda, Y., & Nurlaela, L. (2023). Data Mining to Predict the Ability of Prospective Customer Credit Payments. *2023 11th International Conference on Cyber and IT Service Management (CITSM)*, 1–6.

Anin, K., Kelen, Y. P. K., & Nababan, D. (2023). Sistem Pendukung Keputusan Pemilihan Siswa Berprestasi Menggunakan Metode Profile Matching Berbasis Web (Studi Kasus: SMK Negeri 1 Kefamenanu). *Jurnal Krisnadana*, *2*(3), 388–402.

Aristamy, I. G. A. A. M., Sudipa, I. G. I., Yanti, C. P., Pratistha, I., & Waas, V. D. (2021). An Application of a Decision Support System for Senior High School Scholarship with Modified MADM Method. *2021 6th International Conference on New Media Studies (CONMEDIA)*, 54–59. https://doi.org/https://doi.org/10.1109/CONMEDIA53104.2021.9617180

Dewi, N. L. P. T. K., Nilawati, N. K. U., & Anandita, I. B. G. (2024). Visual Analysis of Marketplace Sales Data for Strategic Decision Making Using Tableau. *TECHNOVATE: Journal of Information Technology and Strategic Innovation Management*, *1*(3), 156–169. https://doi.org/https://doi.org/10.52432/technovate.1.3.2024.156-169

Herlina, H. (2024). Utilization of Big Data for SWOT Analysis in Improving Business Sustainability of MSMEs. *TECHNOVATE: Journal of Information Technology and Strategic Innovation Management*, *1*(2), 89–95. https://doi.org/10.52432/technovate.1.2.2024.89-95

Ibrahim, M. B., Sari, F. P., Kharisma, L. P. I., Kertati, I., Artawan, P., Sudipa, I. G. I., Simanihuruk, P., Rusmayadi, G., Nursanty, E., & Lolang, E. (2023). *METODE PENELITIAN BERBAGAI BIDANG KEILMUAN (Panduan & Referensi)*. PT. Sonpedia Publishing Indonesia.

Kurniati, I., Dewi, C. S., & Juniantika, R. (2021). Penerapan Data Mining dengan Algoritma Neural Network pada Sistem Informasi Prediksi Kasus Balita Gizi Buruk di Provinsi Nusa Tenggara Barat. *JRIS: Jurnal Rekayasa Informasi Swadharma*, *1*(1), 20–27.

Lin, A. K. (2024). The AI Revolution in Financial Services: Emerging Methods for Fraud Detection and Prevention. *Jurnal Galaksi*, *1*(1), 43–51. https://doi.org/10.70103/galaksi.v1i1.5

Puspitasari, T., Tubagus, M., Aziz, F., & Dewi, C. S. (2023). The Analysis of Utilizing Pinterest Application in Making Interactive Teaching-Learning Media for Early Childhood Students. *Journal of Childhood Development*, *3*(2), 81–88.

Sahyunu, S., Moedjahedy, J., Adhicandra, I., Suhanda, Y., Usanto, S., & Rahim, R. (2023).

Evaluating the Suitability of Online Courses using the ELECTRE Method. *AL-ISHLAH: Jurnal Pendidikan*, *15*(3), 2946–2954.

Saputro, J., Saini, K., & Valentine, H. M. (2024). Data Visualization of Higher Education Participation Rates in Indonesia Provinces. *Jurnal Galaksi*, *1*(2), 101–109. https://doi.org/https://doi.org/10.70103/galaksi.v1i2.20

Sopian, A., Usanto, S., & Sauri, R. S. (2023). The Leadership Of The Head Of The Study Program In Efforts To Improve Lecturer Performance. *JHSS (JOURNAL OF HUMANITIES AND SOCIAL STUDIES)*, *7*(3.), 1222–1229. https://doi.org/https://doi.org/10.33751/jhss.v7i3..9452

Sudipa, I. G. I., Sarasvananda, I. B. G., Prayitno, H., Putra, I. N. T. A., Darmawan, R., & WP, D. A. (2023). *Teknik Visualisasi Data*. PT. Sonpedia Publishing Indonesia.

Usanto, S., Sopian, A., Sucahyo, N., Syahrial, R., & Hiswara, I. (2024). INTEGRASI IOT DAN BIG DATA UNTUK OPTIMASI LOGISTIK DAN RANTAI PASOKAN. *JRIS: Jurnal Rekayasa Informasi Swadharma*, *4*(2), 91–99.

Usanto, U. (2022). Big Data Implementation and Use for Business. *Devotion: Journal of Research and Community Service*, *3*(14), 2845–2851.

Usanto, U., Sucahyo, N., Warta, W., Khie, S., & Fitriyani, I. F. (2023). Transformasi Kepemimpinan Yang Bersifat Profetik Dan Pemberdayaan Masyarakat Di Era Society 5.0 Yang Berkelanjutan. *Community Development Journal: Jurnal Pengabdian Masyarakat*, *4*(2), 5287–5301.

Wibowo, G. W. N., & Kraugusteeliana, K. (2024). Exploratory Data Analysis: Visualization of Average Wages of Workers in Indonesia by Region of Residence using Google Data Studio. *TECHNOVATE: Journal of Information Technology and Strategic Innovation Management*, *1*(3), 110–116. https://doi.org/https://doi.org/10.52432/technovate.1.3.2024.10-116

Wiguna, I. K. A. G., Mustafida, A. N., Santika, P. P., Ariantini, M. S., & Sudipa, I. G. I. (2022). Customer Satisfaction Analysis of PLN Mobile Services Using the Naïve Bayes Classifier Method. *INFOKUM*, *10*(5), 52–58. https://doi.org/10.58471/infokum.v14i01

Wiratama, I. K., Welda, W., Permana, I. P. H., Aristana, M. D. W., & Sudipa, I. G. I. (2022). RECOMMENDATION FOR HIGH SCHOOL DETERMINATION BASED ON ACADEMIC POTENTIAL USING NAÃ VE BAYES METHOD. *JIKO (Jurnal Informatika Dan Komputer)*, *5*(2), 108–117. https://doi.org/10.33387/jiko.v5i2.4668