

DESIGN OF STUDENT SUCCESS PREDICTION APPLICATION IN ONLINE LEARNING USING FUZZY-KNN

**Selly Anastassia Amellia Kharis¹, Gatot Fatwanto Hertono^{2*}, Endang Wahyuningrum³,
Yumiati⁴, Sam Rizky Irawan⁵, T. Ahmad Danial⁶, Dimas Septian Saputra⁷**

¹Mathematics Department, Faculty of Science and Technology, Universitas Terbuka

^{3,4}Mathematics Education Department, Faculty of Education and Teacher Training, Universitas Terbuka
Cabe Raya Street, South Tangerang, 15437, Indonesia

^{2,5,6,7}Mathematics Department, Faculty of Mathematics and Science, Universitas Indonesia
Pondok Cina, Depok, 16424, Indonesia

Corresponding author's e-mail: *gatot-fi@ui.ac.id

ABSTRACT

Article History:

Received: 16th January 2023

Revised: 25th April 2023

Accepted: 29th April 2023

Keywords:

Fuzzy-KNN;

Prediction;

Machine Learning

Effective evaluation of student performance is crucial. Hence, many techniques are used, such as statistics, physical examination, and current data mining techniques, to evaluate student performance. Data mining techniques, as known as Educational Data Mining (EDM), collect, process, report, and are used to find the unseen patterns in the student dataset. EDM uses machine learning techniques to dig out useful data from multiple levels of meaningful hierarchy. Various data from intelligent computer tutors, classic computer-based educational systems, online classes, academic data in educational institutions, and standard assessments can be processed for EDM. This led universities to include open and distance learning (ODL) to collect large volumes of student and learning data in their learning management systems (LMS). Students in ODL are relatively familiar with LMS and many learning activities, such as the number of accessing materials and student participation in discussion forums recorded in LMS. The processes of using EDM to improve the quality of educational policymakers with data-based models have become a challenge that institutions of higher education face today. Therefore, this study aims to design applications that predict student performance in online learning using machine learning techniques based on EDM. The machine learning technique used in this research is Fuzzy-KNN. Testing using Fuzzy-KNN produces an accuracy of 92.5%.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

How to cite this article:

S. A. A. Kharis, G. F. Hertono, E. Wahyuningrum, Yumiati, S. R. Irawan, T. A. Danial and D. S. Saputra., "DESIGN OF STUDENT SUCCESS PREDICTION APPLICATION IN ONLINE LEARNING USING FUZZY-KNN," *BAREKENG: J. Math. & App.*, vol. 17, iss. 2, pp. 0969-0978, June, 2023.

Copyright © 2023 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/barekeng/>

Journal e-mail: barekeng.math@yahoo.com; barekeng_journal@mail.unpatti.ac.id

Research Article • Open Access

1. INTRODUCTION

The use of technology, information, and communication (ICT) has grown rapidly, including in education [1]. Since the development of internet technology, the learning process that utilizes the internet network has become an alternative method of delivering a learning process. Online learning has made it possible for students to take part in the learning process anywhere and anytime [2]. The advantages of online learning in terms of expanding access to various learning resources have made online learning widely applied in various educational institutions. Online learning, whose implementation allows the separation between lecturers and students and emphasizes student-centered learning [3], has had a different impact on student success rates compared to learning that is fully carried out face-to-face [4]. One of the issues identified in online learning is student engagement in the learning process [5].

Various educational institutions that carry out online learning processes feel the need to develop a learning analytics (LA) system [6]. Learning analytics is a multi-disciplinary approach that uses predictions that generate information based on data processing, improvement of learning technology, educational data mining, and data visualization [7]. Learning analytics is needed in online learning, especially to measure the level of success of students in following a course [8]. This can be accomplished by monitoring and evaluating student online activity, particularly that recorded in the learning management system (LMS). The results of the LA evaluation are expected to be used to predict student success at the end of lectures. Various parties, ranging from education providers, lecturers, and students, have benefited from LA. For education providers, the results of LA can be a recommendation for making policies related to the implementation of education in their institutions. For lecturers, LA results can be used to identify student performance during the learning process [9]. Thus, lecturers can find students who need special attention. Learning analytics can serve as a warning sign for students about their learning progress [10].

In the development of artificial intelligence technology, the development of machine learning models is widely used to predict and classify problems. Research conducted by Kovacic makes early predictions of student success [11]. Using 450 students' data from 2006 to 2009, A profile of the typical successful and unsuccessful students is provided, along with the most crucial elements for student achievement. The research uses classification tree models to make successful and unsuccessful classifications of student data. Another study was conducted by Oyelade et al., who used the k-Means Clustering algorithm to predict students' academic performance [12]. The k-Means Clustering algorithm was used to predict 79 students with nine courses offered. The outcome improves academic planners' ability to judge candidates' performance on a semester-by-semester basis by raising the bar for subsequent academic sessions' future academic outcomes. Furthermore, research conducted by Okubo et al. predicts student performance using Recurrent Neural Network (RNN) [13]. Unlike previous research, this study developed an application based on Fuzzy-KKN to predict and classify student success in an online course. The predictions and classifications carried out are divided into three classes (multiclass), namely, the predictions of students who pass, students who fail, and observers. The model was developed using data from student learning activities in the LMS. This application is expected to be used as one of the devices that complement the LA system operated at educational institutions.

2. RESEARCH METHODS

This study uses knowledge discovery in databases (KDD) as a research method. KDD is the extraction of potential, implicit, and unknown information from a set of data [14]. In general, the algorithm used in building the model in this study is shown in **Figure 1**. The total number of students whose data was used was 110. The research began by collecting data on student activities at the LMS. The final grade of the student is used as the target variable to classify student performance. The collected data is divided into training and testing data. The training data is used to form the initial model.

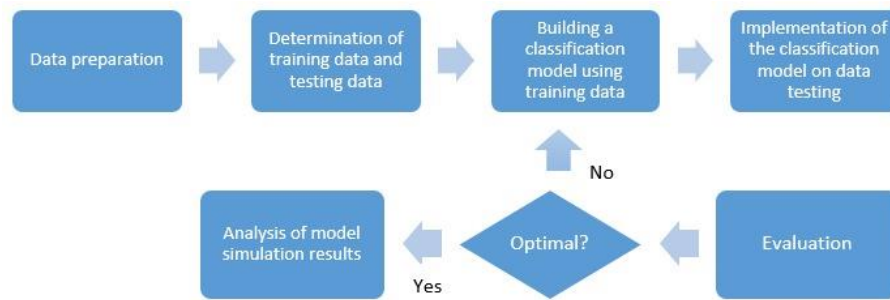


Figure 1. Flowchart of Classification Prediction Model Development

The experiment was carried out with several compositions of training data and testing data, where the training data used activity data before the midterm exam, and the testing data used data after the midterm exam. **Equation (1)** is used to find membership values in Fuzzy-KNN.

$$U_j(p) = \frac{\sum_{j=1}^K (u_{i,j}) \times d(p,q_j)^{\frac{2}{m-1}}}{\sum_{j=1}^K d(p,q_j)^{\frac{2}{m-1}}}, m > 1 \quad (1)$$

With

$$u_{i,j} = \begin{cases} 0,51 + \left(\frac{n_j}{K} \times 0,49\right), & J = i \\ \frac{n_j}{K} \times 0,49, & J \neq i \end{cases}$$

where $U_j(p)$ is the membership value of the observation p in the testing data classified into a class J with $J = \{0,1\}$, $u_{i,j}$ is the membership value of the training data to a class J , i is the class label contained in the training data with $i = \{0, 1\}$, K is the number of nearest neighbours, n_j is the number of classes J in K , and m determine the distance weight and the chosen m is $m = 2$ [15].

The model used to predict the application is built using Fuzzy-KNN. In order to provide the framework with accurate information for assessing the certainty of the choice, the fuzzy KNN classifier works by giving the unlabeled signature a membership value [16]. This membership value can describe how far the observations in the testing data fall into a classification. The model built is tested for accuracy using data testing. If the model built does not provide optimal results in classifying student graduation, then the improvement process is carried out again using a different composition of training data. Refinement iterations are carried out until a model that gives optimal results is obtained. An application using Python programming is built to simulate the resulting model. A dashboard interface called LeADS (Learning Analytics Dashboard System) was developed to be able to provide a prediction of the percentage of student graduation rates in a course.

3. RESULTS AND DISCUSSION

The predictions made in this study resulted in three types of categories: pass, fail, and observer. The "pass" category is given to students who meet the minimum threshold of a predetermined passing grade. The "failed" category is given to students who fail to meet the minimum threshold of the predetermined passing grade. The "observer" category is given to students who only access the LMS but have never participated in the course assessment component. **Figure 2** shows the results of machine learning modeling predictions that have undergone development for multiclass problems.

Accuracy	= 92.5%
Recall (Active Fail)	= 63.33%
Recall (Active Pass)	= 96.33%
Recall (Observers)	= 100%
Precision (Active Fail)	= 68.33%
Precision (Active Pass)	= 92.33%
Precision (Observers)	= 100%

Figure 2. Multiclass machine learning prediction results that have been developed

After the modeling is adjusted, the model is simulated again using student data taken from the LMS. The features of the data retrieved are also adjusted for the modeling. The student data features used in this study include the number of accesses to learning files, the number of accesses to discussion forums, the number of accesses to learning videos, quiz grades, and assignment scores. Then the activity log data is downloaded, and data processing is carried out by retrieving the data until mid-semester (before UAS). **Figure 3 (a)** shows the final results of student activity data. Furthermore, data collection is carried out for value data. Data is taken on the Gradebook menu on the LMS. **Figure 3(b)** shows the student gradebook used for model making.

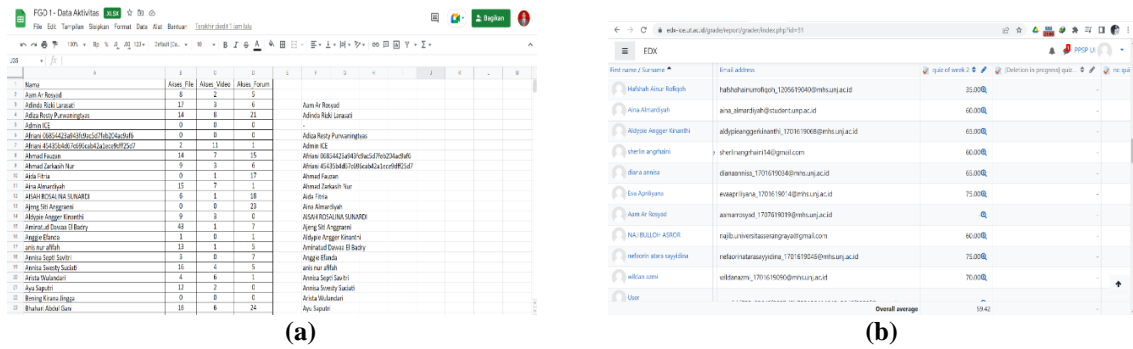


Figure 3. (a) Final Results of Student Activity Data, (b) Student Gradebook

The gradebook menu data is downloaded, and the data format is converted into Excel format. Furthermore, activity data and value data are combined into one Excel file. The final result data that has been put together is then tidied up by adjusting the feature naming column in the modeling that has been done. **Figure 4** is the final result of the data whose feature names have been adjusted.

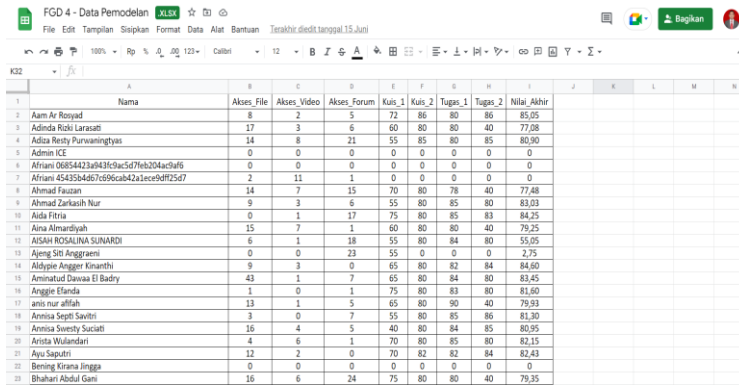


Figure 4. Final Data Results with Feature Name Adjustments

The final data, whose name has been adjusted, is cleaned so that it is ready for use. This study uses student history data and ongoing data. History data is student data from completed courses that are used for machine learning. The history data used is only up to mid-semester data and the final grades of each student. The historical data used will be used as a pattern in conducting analysis using machine learning to calculate student graduation based on ongoing data. Duplicate historical data will be deleted when the modeling is processed. Ongoing data is student data on courses that are being held and student data that will predict graduation. The ongoing data used is also taken until the middle of the semester, and there is no final grade. Students in ongoing data will be predicted to graduate based on historical data. Duplicate ongoing data will not be deleted and will be maintained. **Figure 5(a)** shows the history data format that is ready for use. **Figure 5(b)** shows the ongoing data format ready to be used.

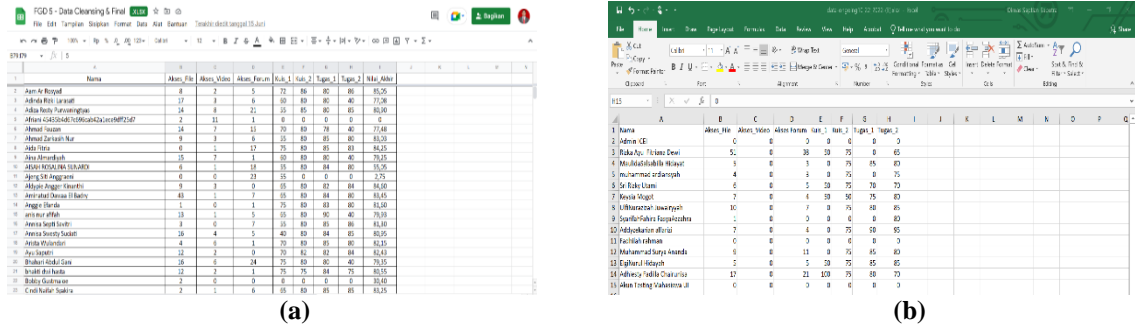


Figure 5. (a) Ready-to-use history data format, (b) Ready-to-use on going data format.

Historical and ongoing data were previously collected manually by opening the LMS and then processed using Excel. This is, of course, ineffective and inefficient for supporting lecturers who want to check the class they are in charge of because the data retrieval process is difficult and time-consuming, and there is the possibility of errors due to wrong formatting that can cause the modeling to not work. Based on this problem, an initiative was made to make tools that are easier to use for lecturers so that they don't take up time and there are no data format errors that can cause modeling not to work as it should. To make it easier to retrieve historical and ongoing data, a database query is performed on the LMS. The way this query works is simply by reading a valid class ID on the LMS and then pulling feature data according to the modeling format, both historical data and ongoing data. The first step was to create a stand-alone query system to ensure that the results of pulling the data obtained were correct and free of errors. Data withdrawal can be accessed using a simple command, namely, <https://idols.ui.ac.id/query/test.php?cid=473>. The number 473 in the command can be replaced with a valid class ID on the LMS. If the class ID is valid, then the Excel file will be downloaded automatically according to the modeling format. Figure 6(a) shows an example of a class ID on an LMS. After the initial development is complete, the integration of queries with the LeADS application is carried out. Figure 6(b) shows an example of an LMS view whose class ID is valid.

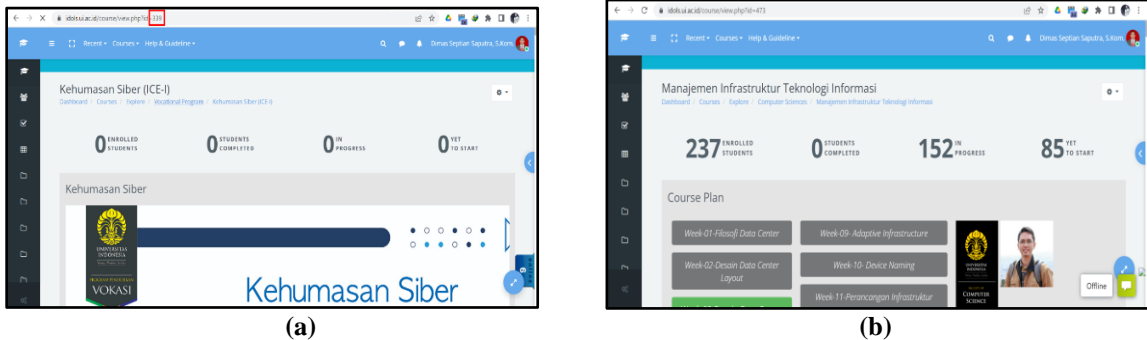


Figure 6. (a) Example of Class ID in LMS, (b) Example of an LMS View whose Class ID is Valid.

After the Excel data history file is downloaded, the display shown in Figure 7(a) will appear. The file downloaded using the query already has the appropriate modeling format, so there is no need to preprocess the data again. Figure 7(b) shows the ongoing data that has been downloaded using a query that also has the appropriate modeling format.

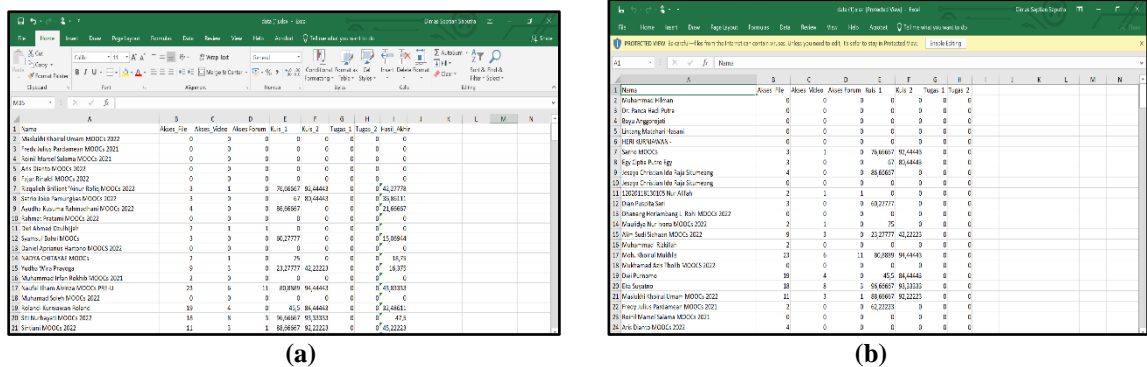


Figure 7. Example of Downloaded History Data Results in Accordance with the Modeling Format using Query, (a)History Data, (b) On Going Data

In addition to making queries to make it easier for lecturers, the research team also created a web-based application called LeADS (Learning Analytics Dashboard System). The LeADS application is an embodiment of web-based modeling that has been made using Fuzzy-KNN. The LeADS application can be accessed via the <https://ileads.my.id> page. The data flow carried out in the LeADS application is shown in **Figure 8**.

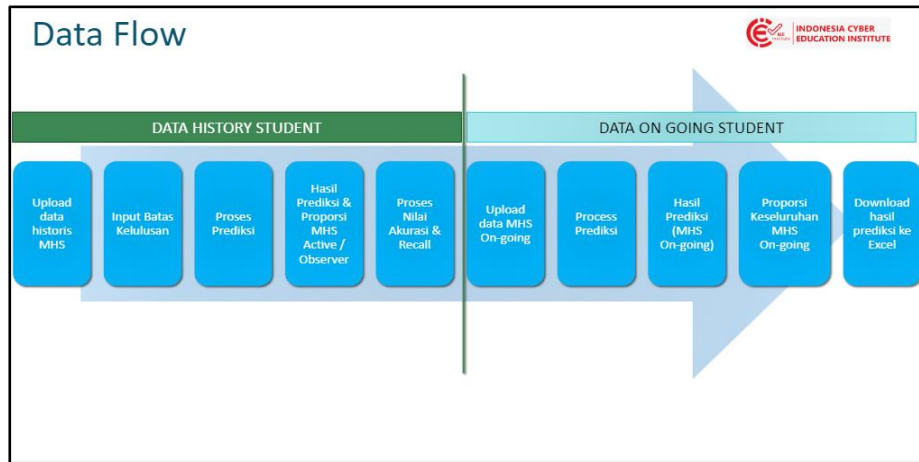


Figure 8. Display of On-Going Data Prediction Results.

In the first step, student history data is uploaded, and the graduation limit is determined. Then the prediction process and calculation of accuracy and recall values are carried out. Next, input the data ongoing. From the modeling results obtained from historical data, predictions are made for ongoing data. To be able to enter the LeADS application, users must first log in by typing their username and password in the fields provided and then clicking Login. **Figure 9** shows the initial appearance of the LeADS application.

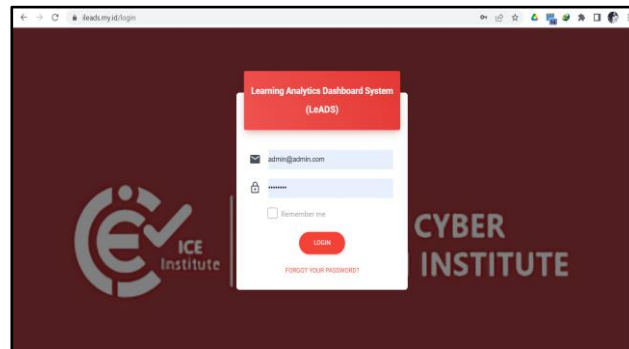
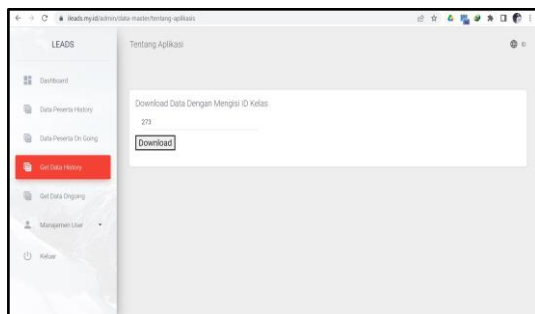
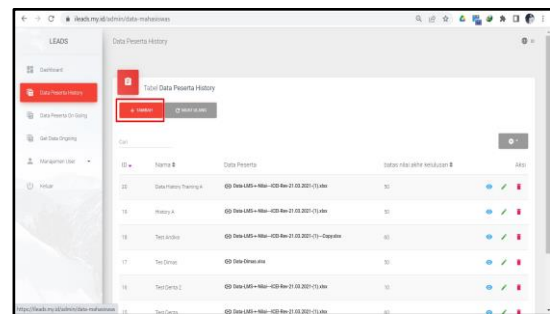


Figure 9. Initial display of the LeADS application.

After log in, please click the Get Data History menu to download history data then fill in a valid Class ID to retrieve the history data. The Get Data History menu for the LeADS application is shown in **Figure 10(a)**. After that, to upload the history data that has been downloaded, click on History Participant Data. Click 'add,' then enter the name of the course in the 'name' field, the history data file that has been downloaded in the Participant Data field, and the passing grade limit for the course in the Pass Grade Limit field, then click Save.



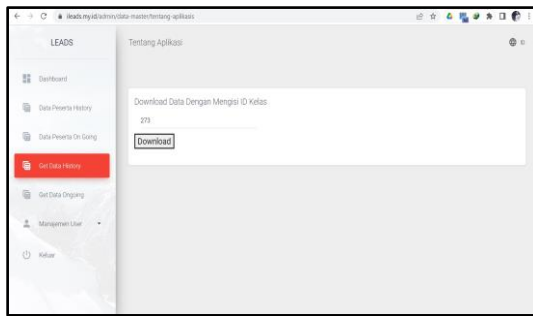
(a)



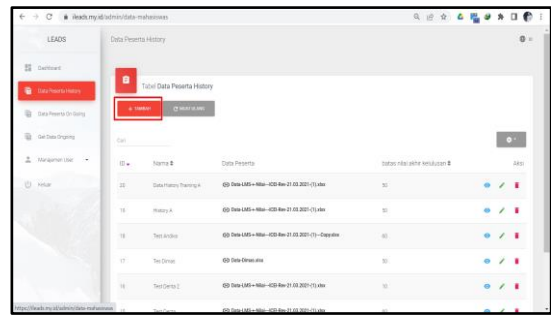
(b)

Figure 10. (a) LeADS Application Get Data History menu, (b) Add Participant Data LeADS Application History Menu

Figure 10(b) shows the contents of the added menu data for participants in the history of the LeADS application. If successful, a green notification will appear, and it will appear in the list of recently added data. Click the eye on the right to see the results of the modeling data history. When the history data has been successfully added, a notification will appear, as shown in **Figure 11(a)**. To view the history participant data that has been successfully added, you can click on 'history participant data' as shown in **Figure 11(b)**.



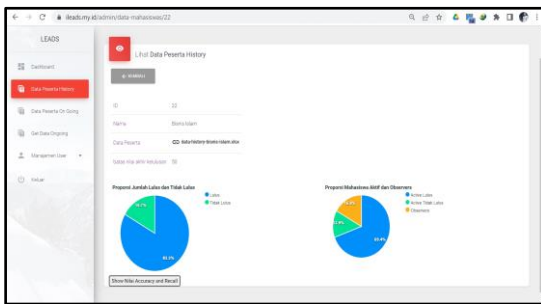
(a)



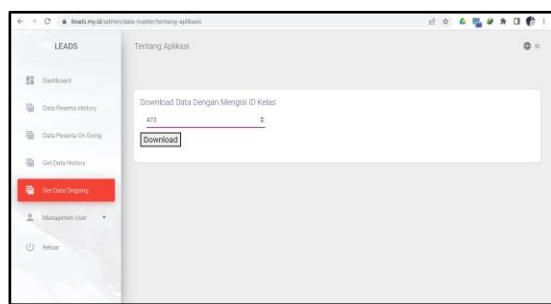
(b)

Figure 11. (a) Fill in Menu Add Participant Data LeADS Application History, (b) History Data Successfully Added.

Then click the Get Data Ongoing menu to download the ongoing data, and then fill in a valid class ID to retrieve the ongoing data. The display of historical data modeling results is shown in **Figure 12(a)**. The display of the Get Data on Going menu for the LeADS application is shown in **Figure 12(b)**.



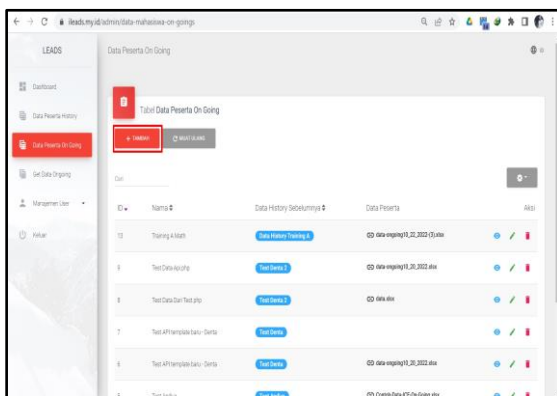
(a)



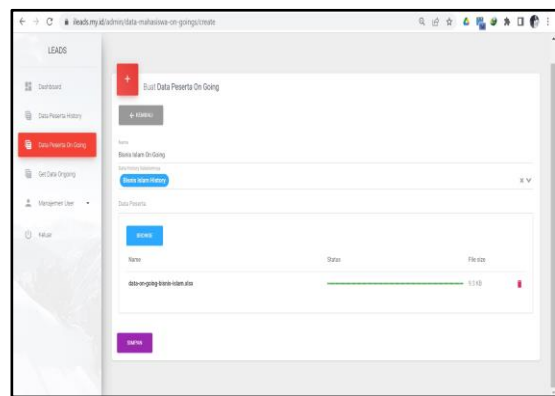
(b)

Figure 12. (a) Display of History Data Modeling Results, (b) Menu Get Data on Going Application LeADS.

After downloading the ongoing participant data, go to the Ongoing Participant Data menu and upload the data. Click 'add,' then fill in the name of the course in the Name field, select history data from the same course that was previously uploaded in the Data History field and the ongoing participant data file in the Participant Data field, then click Save. **Figure 13(a)** shows the menu for adding data when going for the LeADS application. To see the ongoing participant data that has been successfully added, you can click on 'ongoing participant data' as shown in **Figure 13(b)**.



(a)



(b)

Figure 13. (a) Menu Add Data Ongoing LeADS application, (b) Menu Entries for On Going Participant Data Add the LeADS Application.

If successful, a green notification will appear, and it will appear in the list of recently added data. Click the eye icon on the right to see the prediction results of the ongoing data. **Figure 14** shows the appearance of the prediction results of ongoing data.

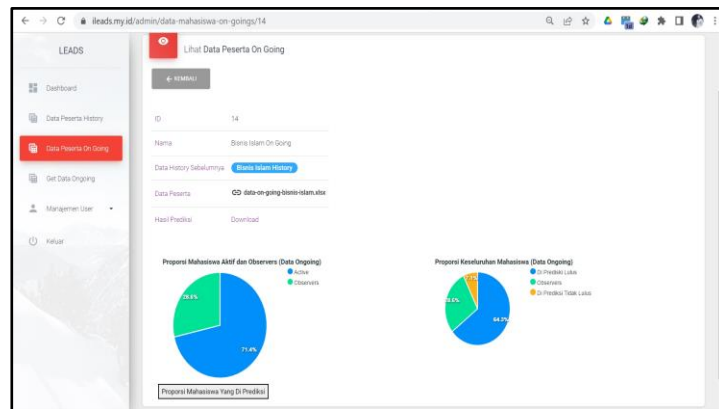


Figure 14. Display of On-Going Data Prediction Results.

4. CONCLUSIONS

This research has developed a learning analytics model using Fuzzy-KNN to predict and classify students' potential success in online learning. Our study also includes the development of a dashboard interface called LeADS, which enables monitoring and predicting students' performance in ongoing MOOCs courses. The LeADS tool utilizes various features: 1) the number of accesses to file materials, 2) the number of videos watched, 3) the number of visits to discussion forums, 4) quiz grades, and 5) test scores to make accurate predictions for the pass, fail, and observer participants. Furthermore, our study emphasizes the importance of considering a minimum MOOCs standard policy across various universities or LMSs for different subjects to enhance the accuracy of modeling and predictions made on student data. By providing these tools and insights, we hope to support educators and institutions in improving online learning experiences and outcomes for their students.

ACKNOWLEDGMENT

This research was funded by the Institute for Research and Community Service at the Open University No. 23284/UN31.LPPM/PT.01.03/2021 We would like to thank the Indonesia Cyber Education Institute (ICE Institute), without which this research would not have taken place.

REFERENCES

- [1] K. A. Razak, T. Norhayati, T. Othman, M. I. Hamzah, and H. Zulkifli, "Information and Communication Technology among Excellent Islamic Education Teachers in Selangor Malaysia," *International Education Studies*, vol. 7, no. 13, 2014, doi: 10.5539/ies.v7n13p146.
- [2] S. Dhawan, "Online Learning: A Panacea in the Time of COVID-19 Crisis," *Journal of Educational Technology Systems*, vol. 2020, no. 1, pp. 5–22, doi: 10.1177/0047239520934018.
- [3] "The Impact of E-Learning Technologies on Student's Motivation: Student Centered Interaction in Business Education," *International Journal of Research in Tourism and Hospitality*, vol. 6, no. 2, 2020, doi: 10.20431/2455-0043.0601002.
- [4] D. Spencer and T. Temple, "Examining students' online course perceptions and comparing student performance outcomes in online and face-to-face classrooms," *Online Learning Journal*, vol. 25, no. 2, pp. 233–261, 2021, doi: 10.24059/olj.v25i2.2227.
- [5] M. Chakraborty and F. Muiyia Nafukho, "Strengthening student engagement: What do students want in online courses?," *European Journal of Training and Development*, vol. 38, no. 9, pp. 782–802, Oct. 2014, doi: 10.1108/EJTD-11-2013-0123.
- [6] S. Anastasia Amellia Kharis and A. Haqqi Anna Zili, "Learning Analytics dan Educational Data Mining pada Data Pendidikan," *Jurnal Riset Pembelajaran Matematika Sekolah*, vol. 6, 2022.
- [7] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "IJTEL)-Special Issue on "State-of-the-Art in TEL." [Online]. Available: <https://tekri.athabascau.ca/analytics/>

- [8] N. Sclater, A. Peasgood, and J. Mullan, "Learning Analytics in Higher Education A review of UK and international practice Full report," 2016.
- [9] W. Greller and H. Drachsler, "International Forum of Educational Technology & Society Translating Learning into Numbers: A Generic Framework for Learning Analytics," *Source: Journal of Educational Technology & Society*, vol. 15, no. 3, pp. 42–57, 2012, doi: 10.2307/jeductechsoci.15.3.42.
- [10] G. Siemens *et al.*, "Open Learning Analytics: an integrated & modularized platform Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques Project Overview," 2011. [Online]. Available: www.solaresearch.org
- [11] Z. J. Kovačić, "Early Prediction of Student Success: Mining Students Enrolment Data."
- [12] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance," 2010. [Online]. Available: <http://sites.google.com/site/ijcsis/>
- [13] F. Okubo, A. Shimada, T. Yamashita, and H. Ogata, "A neural network approach for students' performance prediction," in *ACM International Conference Proceeding Series*, Mar. 2017, pp. 598–599. doi: 10.1145/3027385.3029479.
- [14] P. Dehning, K. Lubinetzki, S. Thiede, and C. Herrmann, "Achieving Environmental Performance Goals - Evaluation of Impact Factors Using a Knowledge Discovery in Databases Approach," in *Procedia CIRP*, 2016, vol. 48, pp. 230–235. doi: 10.1016/j.procir.2016.03.108.
- [15] T. M. Mohamed, "Pulsar selection using fuzzy knn classifier," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 1–6, Jun. 2018, doi: 10.1016/j.fcij.2017.11.001.
- [16] M. el Bakry, S. Safwat, and O. Hegazy, "Big Data Classification using Fuzzy K-Nearest Neighbor," 2015.

