Online Media & Society
Vol.5, No.3 (September, 2024), Pp.60-72
ISSN(online): 2790-1882, ISSN(print):2790-1874

Original Article https://hnpublisher.com

Machine Learning and Analytics for Performance Prediction of ODL Students: Decoding Digital Communication for Sustainability

Hanaan Sadeed Ahmad[1], Moiz Uddin Ahmed[2], Shahid Hussain[3]

[1]Research Scholar, Department of Computer Science, Allama Iqbal Open University, Islamabad, Pakistan.
[2]Assistant Professor, Department of Computer Science, Allama Iqbal Open University, Islamabad, Pakistan.
[3]Assistant Professor, Department of Mass Communication, Allama Iqbal Open University, Islamabad, Pakistan.
Correspondence: shahid.hussain@aiou.edu.pk[3]

ABSTRACT

Aim of the Study: The extensive use of digital communication in Open and Distance Learning (ODL) is generating a huge volume of academic transactions. This occurs through online platforms such as the Learning Management System (LMS) and Student Information System (SIS). The goldmine LMS and SIS transactions emphasized the need to convert this data into meaningful knowledge. The aim of the current study is to apply machine learning and analytics to decode the digital data generated by LMS and SIS, extracting meaningful knowledge for continuous improvement and sustainability in educational processes.

Methodology: This paper presents a model based on learning analytics and machine learning to predict the academic performance of students enrolled in a course offered at an open university of Pakistan. The researchers' extracted data for two semesters combined it with the best attribute subset and employed eight (08) machine learning algorithms by dividing data into four sets.

Findings: The results of the study validated the predictive ability of machine learning on a localized dataset of distant learners of the country's largest open university. Students use digital platforms for communication and learning and their usage can be analyzed to decode and predict their performance. The findings of the study demonstrate the effective utilization of Artificial Intelligence and Machine Learning technologies, effectively overcoming challenges and leveraging them to create opportunities.

Conclusion: It is concluded from the study that students use digital platforms for communication and learning, and their usage can be analysed to decode and predict their performance.

Keywords: Distance Education, Learning Analytics, Machine Learning, Student Performance Prediction, Sustainable Education.

Article History

Received:
July 27, 2024

Revised:
September 19, 2024

Accepted:
September 23, 2024

Published:
September 30, 2024

Introduction

Open and Distance Learning (ODL) is an approach of acquiring education that is independent of physical presence, time and distance (Simpson, 2018). It provides a flexible mode of learning where learning can be held at any place and at any time at the pace of learners. The flexible mode of distance education is making it popular even in the traditional institutes, which have experienced an educational lockdown during Covid-19 and thus preparing themselves for such disastrous situations in the future (Dhawan, 2020; Safdar et al., 2020). The ease of learning from office or home is also attracting more students and therefore, the rate of enrolment in distance learning programs is on the rise all over the world and especially in the developing countries. Technology has added more value to it by transforming Open and Distance Education into Technology Enabled Learning (TEL). The digital phenomena through TEL, enable institutes to store the profiles and the preferences of admitted students and record their transaction logs in the form of Learning Analytics (LA).

Learning Analytics is the collection and analysis of data about the learners in a particular educational setting with the goal of improving academic standards, learners' performances, and decision-making (Kew & Tasir, 2021). There are three main building blocks of LA i.e. data, analysis and action (Dixit, 2020). Data refers to the collection of facts and figures such as profiles, actions, and preferences of learners. Analysis is the division of complex data into the smaller parts with the objective to closely examine and discover the hidden relationship. It helps to shape an opinion and give judgment for initiating appropriate actions while taking important decisions. Therefore, LA is the combination of Learning (improving knowledge, assessment using technology), Analytics (visualization, data analysis) and Learner-Centred Design (usability) that is carried out to investigate the effectiveness of a learning environment (Goodell & Thai, 2020; Safdar et al., 2020a). The combination of LA with Machine Learning (ML) is an opportunity to optimize analytical tools to discover useful information.

Machine Learning (ML) is a sub-area of Artificial Intelligence (AI) that processes data by using specialized algorithms and inferring knowledge by imitating the human ability of learning and deducing (Burgos et al., 2018; Safdar & Khan 2020; Khalid et al., 2023). It can identify the relationship and patterns among huge datasets and extract valuable information to predict the performance of students. Such model when applied to ODL data will help in adopting the future technology of personalized e-learning and may become a role model for competitive locations. Following the emerging trends, the main objectives of the current study are (1) to propose a machine learning model based on learning analytics to predict the academic performance of ODL students enrolled in an open university of Pakistan (2) to evaluate the performance of machine learning algorithms with the relative importance of students' attributes. The model may contribute as a next step in the digital transformation of a distance learning university in a developing country like Pakistan for sustainable education.

Problem Statement

One of the significant issues in ODL is student retention within academic programs. The high enrolment in ODL programs makes it challenging to examine and track individual student performance. Since COVID, the use of LMS in ODL programs has also increased. Another emerging issue is the rapidly growing volume of data generated by SIS and LMS, due to the vast number of educational transactions. This study focuses on pre-processing educational data and forecasting ODL learners' performance using learning analytics and machine learning approaches. So, the study will provide a mechanism for decoding the digital communication of the ODL students with the LMS and SIS for prediction of their performance for a sustainable learning environment.

Study Objectives

The objectives of the study are to:

1. Analyse learning analytics features and general profiles from LMS and SIS.
2. Pre-process the educational data and find the best attribute subset important for performance prediction.
3. Propose a predictive model using machine learning for selected attributes.

Background and Related Work

Learning Analytics is an emerging field in the educational sector that concentrates on collecting and analysing extensive datasets with the explicit aim of enhancing academic standards. The experts are of the opinion that it will become an integral part of educational systems to evaluate the performance of learners. It has been identified as a new wave of development for quality assurance of academic programs from the early education to the postgraduate level of studies. It has been adopted by many institutes of higher education due to available potential of using technology tools by students, teachers and administrators (Charitopoulos et al., 2020). A number of studies have been conducted to address the implementation of LA in different regions of the world like Tsai et al. (2020) described the trends and barriers of adopting LA in Europe. The study found the teachers and the support staff as more active users of LA as compared to the students. Another study emphasized that a significant portion of the research conducted in Asia has centred around mitigating student dropout ratio and enhancing learning practices (Li et al., 2018). One study has examined the pre-conditions before adopting the LA in Southeast Asia, emphasizing that while the use of ICT is prevalent in most institutes, the polices of implementing LA are still under process (Rodrigo, 2018). Another study listed the initiatives of LA in Latin America (Cechinel et al., 2020). In continuation of these initiatives all over the world, there is a need to take benefit of LA potential and apply its methodologies to evaluate the data of distant students.

The ODL institutes can measure and analyze online transactions of distant learners and improve online services with proper academic advice well in time when needed. The huge data sets are generated by e-learning systems that are used by students to access and complete various learning activities such as reading text, listening to multimedia, downloading content, and uploading assignments using a web-based system. These activities are managed through a platform called Learning Management System (LMS), which uses various technologies, largely on the internet, to help students to participate and communicate online. The growing use of LMS has resulted in a significant amount of real-time learning data that stores students' personal data, learning progress, behaviours, and usability preferences and is a source of LA data. There is another system that is used by educational institutes to automate registration and admission data. This system is called Student Information System (SIS) which is a platform that maintains the semester wise information of students from program enrolment history to all grades obtained during the academic career. It stripes to manage the entire data to automate the academic and administrative processes at one place, easily accessed by students, parents and administration. Both LMS and SIS are the sources of data that can be used by LA tools to answer many research questions like the performances of individual and a group of students, improvement in courses/programs and the success factors in specific regions/areas. Based on customized attributes from LMS and SIS it can help to recommend targeted course offerings, curriculum development, student learning outcomes, instructor performance and post-educational employment options. It has the potential to provide new insight into the learning process, allowing for practice that will help students succeed in their studies (Dawson et al., 2019).

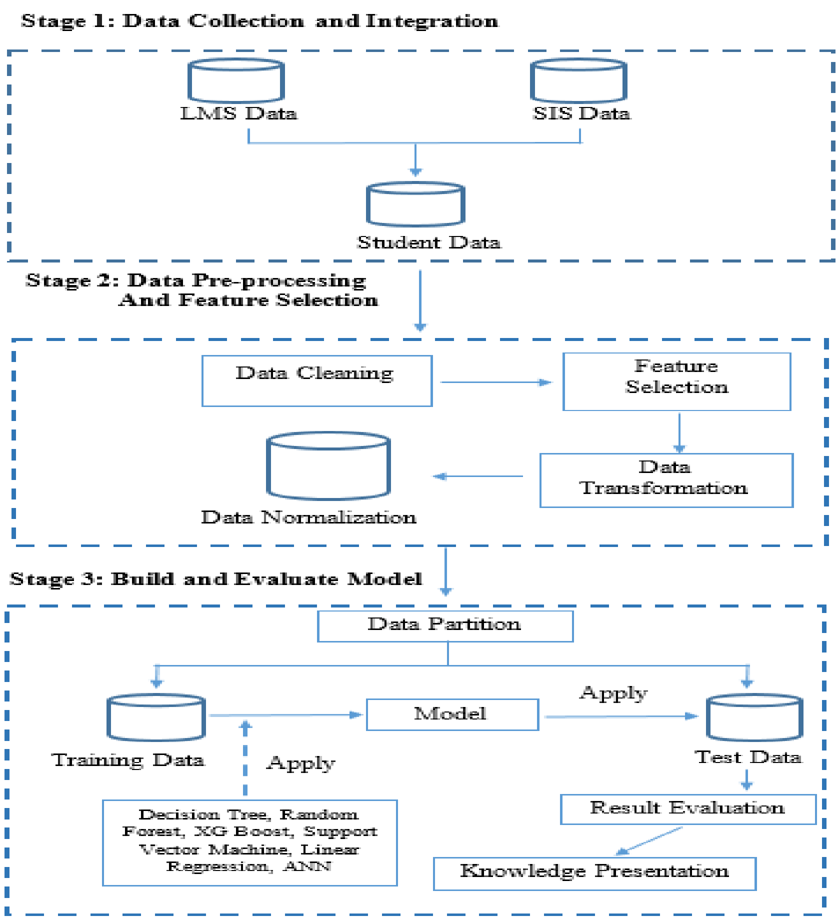
Machine Learning (ML) is a branch of AI that refers to specialized computer programs that learn from past experience without external intervention (Burgos et al., 2018). It can be seen as a sister community of LA as it can implement techniques that can automatically detect patterns in educational data and predict future outcomes. The ML comprises two main approaches supervised learning and unsupervised learning based on labelled and unlabelled datasets respectively and a number of studies have been conducted to analyze educational data (Burkov, 2019). Clustering has been used where labels within educational data were unknown (Bharara et al., 2018) to find the important features in a learning domain (Mohamed et al., 2022). The important algorithms found in cluster analysis of educational data are K-Means algorithm (Mohamed et al., 2022) and Partition around Medoids (Navarro & Ger, 2018). Classification is a part of the supervised learning technique which assigns predefined labels to incidences based on facts dug out from educational data. Regression is another method under supervised learning for predictive modelling to predict continuous outcomes on student performance. The important algorithms used are DT, RF, KNN and Logistic Regression, (Sousa, et al., 2021).

After the post Covid era, we have seen a remarkable increase in technology adoption, especially in educational institutes resulting in increased educational data which is useless unless converted into the meaningful knowledge. We took up the challenges of predicting student performance in one of the largest ODL institute of a developing country like Pakistan which is striving hard to adopt technology solutions in many fields including the education sector. We aim to analyse an educational dataset of ODL students which has been drawn from the central database of the university. The work will add a new dimension of research in a localized domain by proposing a predictive model on a hybrid set of features extracted from LMS and SIS usage and interaction logs.

Methodology

This section describes the proposed model as shown in figure 1. The model comprises of 3-stage methodological steps, as described below:

Figure 1: *Proposed Model Diagram*



Data Collection and Integration (Dataset)

Data collection is the first step where the numeric dataset of educational records and transactions was collected from LMS and SIS. Thirteen (13) attributes based on literature review were selected and availability of data in the central repository as shown in table 1.

Table 1: *Dataset Features*

S.NO.	Attribute	Description
1.	Gender	A binary attribute that consists of Male or Female.
2.	Age	A numeric attribute.
3.	Attendance	A numeric attribute that consists of attendance marks. In our dataset, its range is from 0 to 100.
4.	Assignment 1	A numeric attribute that consists of marks attained in assignment 1. In our dataset range of Assignment is from 0 to 100.
5.	Assignment 2	A numeric attribute that consists of marks attained in assignment 2. In our dataset range of Assignment is from 0 to 100
6.	Exam	A numeric attribute that consists of marks of final examination. In our dataset range of Exam is from 0 to 100.
7.	Total hits	A numeric attribute consisting of the total number of sessions on LMS.
8.	Assignment submission	A numeric attribute consisting of logs of those students who have submitted their assignments.
9.	BBN Meeting Joined	A numeric attribute consisting of logs of those students who have attended their online classes.
10.	Assignment Feedback View	A numeric attribute consisting of logs of those students who have seen feedback given by the instructor.
11.	Grade Report View	A numeric attribute consisting of logs of those students who have seen marks of attendance and assignments given by the instructor during the semester.
12.	Area type	A binary attribute consists of Semi urban or Urban.
13.	Final Aggregated Marks	A numeric attribute that consists of 70% of final exam, 30 % of assignment marks. In our dataset range of final marks is from 0 to 100.

The collected data was integrated into a single dataset by matching the primary key (Roll No.) of the students in both databases.

Data Pre-processing and Feature Selection

Raw data is always deficient and contains uncertain values. It was pre-processed to convert it into an understandable format by carrying out the following steps:

Data Cleaning

The unwanted and irrelevant attributes were removed from the dataset. The missing values were replaced with the null values using imputing techniques implemented by the Jupiter Lab tool.

Feature Selection

The Wrapper method was used to create selection models and identify the best attribute subset.

Data Transformation

This step was carried out to normalize the data set for scaling up the attributes. The data was transformed into a single dataset for applying the machine learning algorithms.

Experimentation and Interpretation

The experiments were carried out on the transformed data by employing machine learning algorithms. The dataset was divided into four sets as shown in table 2.

Table 2: *Experiment wise Features*

Experiment	Total Records	Features	Semester
1st Experiment	3800 Records	SIS (Demographics, Assignment Marks, Midterm Marks)	Autumn-2020
		LMS (Total hits, Online Classes Attended, Grade Report View)	
2nd Experiment	3800 Records	SIS (Demographics, Assignment Marks, Midterm Marks)	Autumn-2020
		LMS (Total hits, Assignment Submission, Assignment Feedback View)	
3rd Experiment	3300 Records	SIS (Demographics, Assignment Marks, Midterm Marks)	Spring-2020
		LMS (Total hits, Online Classes Attended, Grade Report View)	
4th Experiment	3300 Records	SIS (Demographics, Assignment Marks, Midterm Marks)	Spring-2020
		LMS (Total hits, Assignment Submission, Assignment Feedback View)	

The performance of each model was evaluated using error estimation measures employing different machine learning algorithms.

Performance Measures

The performance of algorithms was measured by using three metrics i.e. R Square (R^2), Percentage Absolute Difference (PAD) and Root Mean Square Error (RMSE), the formulas are given in equations 1, 2 and 3 respectively.

$$R^2 = 1 - \frac{RSS}{TSS} \qquad \text{(Eq. (1))}$$

Where

R^2 = Coefficient of determination

RSS = Sum of square of residuals

TSS = Total sum of squares

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - x_i| \qquad \text{(Eq. (2))}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - x_i)} \qquad \text{(Eq. (3))}$$

Where

x_i = Actual Value

y_i = Predicted Value

n = total number of records

Results

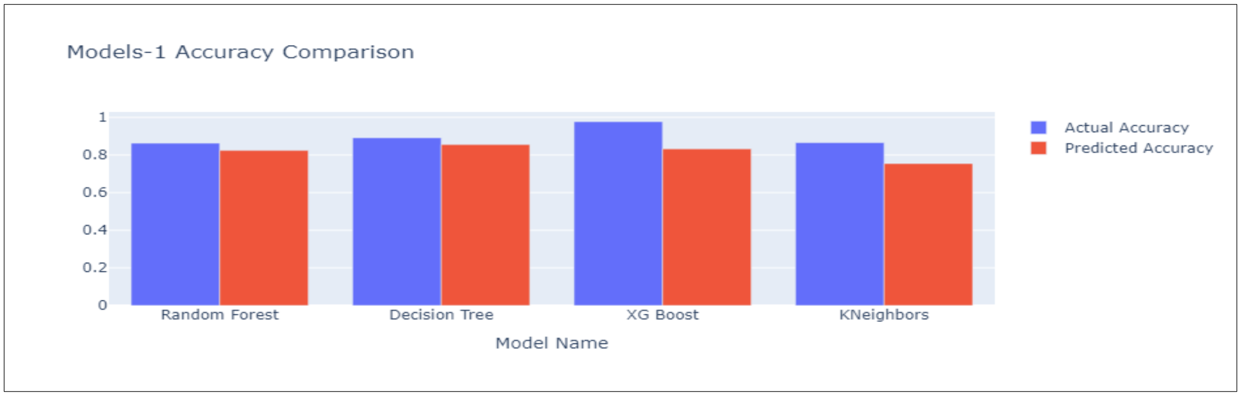
We employed eight (08) machine learning algorithms on the dataset. Four algorithms XGBoost, Decision Tree (DT), K- Nearest Neighbour (KNN) and Random Forest) were shortlisted on the basis of predictive performance. The dataset of each model was divided into 30% testing and 70% training ratio. The performance of Model 1 is shown in table 3 and figure 2.

The results reveal that XGBoost showed the best performance for training data with an accuracy of 97 % (R2 = 0.97, RMSE = 2.93 and MAE value = 2.12). Its accuracy dropped to 83% (R2 = 0.83, RMSE = 7.49 and MAE value = 5.84) for the testing data. The performance of DT was better in view of closer accuracies of 89 % (R2 = 0.89, RMSE = 6.40 and MAE=4.96) for training and 85 % (R2 = 0.85, RMSE = 6.95 and MAE=5.37) for validation. The performance of KNN and RF was almost equal for training data with 86.5 % (R2 = 0.865, RMSE = 7.11 and MAE = 4.90) and 86.2 % (R2 = 0.862, RMSE = 7.19 and MAE = 5.87) accuracies, respectively. The KNN dropped its performance for the testing data with an accuracy of 75 % (R2 = 0.75, RMSE = 9.07 and MAE = 6.13) as compared to RF which retained the accuracy up to 82 % (R2 = 0.82, RMSE = 7.67 and MAE = 6.20).

Table 3: *R-Squared, RMSE and MAE for Training and Validation for Model – 1 using 3800*

Algorithms	Input	R-Squared	RMSE	MAE
XGBoost	Training	0.97	2.93	2.12
	Validation	0.83	7.49	5.84
DT	Training	0.89	6.40	4.96
	Validation	0.85	6.95	5.37
KNN	Training	0.865	7.11	4.90
	Validation	0.75	9.07	6.13
RF	Training	0.862	7.19	5.87
	Validation	0.82	7.67	6.20

Figure 2: *Actual and Predicted Students' Results*

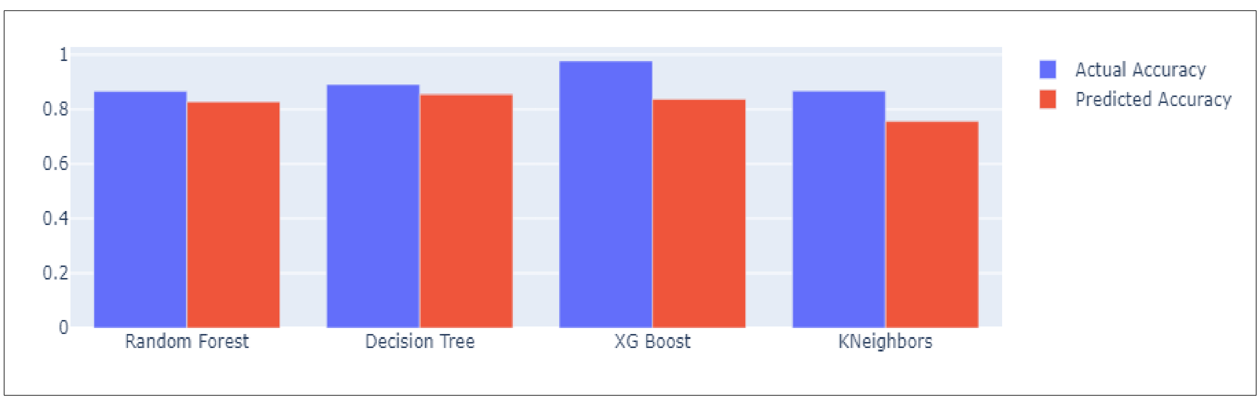


The performance of Model 2 is shown in table 4 and figure 3. The XGBoost again showed best performance for training data with an accuracy of 97 % (R^2 = 0.97, RMSE = 2.96 and MAE value = 2.16). Its accuracy dropped to 83% (R^2 = 0.83, RMSE = 7.38 and MAE value = 5.74) for the testing data. The performance of DT was again better in the view of closer accuracies of 89 % (R^2 = 0.89, RMSE = 6.42 and MAE=4.97) for training and 85 % (R^2 = 0.85, RMSE = 6.98 and MAE=5.40) for validation. The performance of KNN and RF was almost equal for training data with 86 % (R^2 = 0.86, RMSE = 7.09 and MAE = 4.87) and 86 % (R^2 = 0.86, RMSE = 7.13 and MAE = 5.83) accuracies respectively. The KNN dropped its performance for the testing data with an accuracy of 75 % (R^2 = 0.75, RMSE = 9.03 and MAE = 6.10) as compared to RF which retained the accuracy up to 82 % (R^2 = 0.82, RMSE = 7.60 and MAE = 6.16).

Table 4: *R-Squared, RMSE and MAE for Training and Validation for Model – 2 using 3800* Data Set

Algorithms	Input	R-Squared	RMSE	MAE
XGBoost	Training	0.97	2.96	2.16
	Validation	0.83	7.38	5.76
DT	Training	0.89	6.42	4.97
	Validation	0.85	6.98	5.40
KNN	Training	0.86	7.09	4.87
	Validation	0.75	9.03	6.10
RF	Training	0.86	7.13	5.83
	Validation	0.82	7.60	6.16

Figure 3: *Actual and Predicted Students' Results*

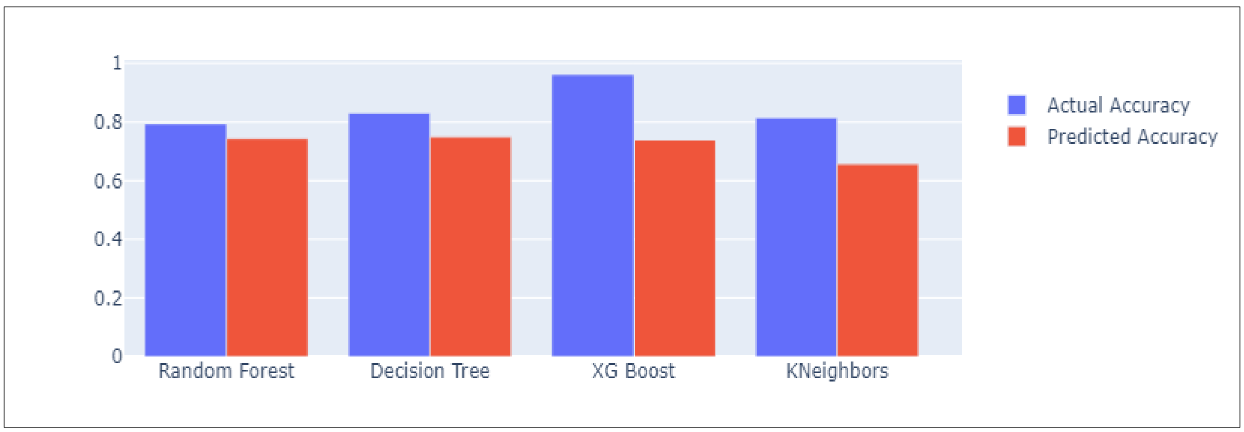


The results of Model 3 are shown in table 5 and figure 4. Overall, the XGBoost again showed the best performance for training data with an accuracy of 96 % (R^2 = 0.96, RMSE = 3.30 and MAE value = 2.36). Its accuracy dropped to 73% (R^2 = 0.73, RMSE = 8.37 and MAE value = 6.64) for the testing data. The performance of DT was although better again in view of closer accuracies of 82 % (R^2 = 0.82, RMSE = 6.96 and MAE = 5.52) for training and 74 % (R^2 = 0.74, RMSE = 8.21 and MAE = 6.46) for validation yet lowered as compared to Models 1 &2. Similar analogy as of models 1 & 2 was found on the performances of KNN and RF, but overall the accuracy dropped for the validation. The KNN showed an accuracy 81 % (R^2 = 0.81, RMSE = 7.28 and MAE = 5.39) for training which dropped to 65 % (R^2 = 0.65, RMSE = 9.61 and MAE = 6.78) for testing. The RF showed an accuracy of 79 % (R^2 = 0.79, RMSE = 7.68 and MAE = 6.32) for testing which dropped to 65 % (R^2 = 0.65, RMSE = 9.61 and MAE = 6.78) for testing.

Table 5: *R-Squared, RMSE and MAE for Training and Validation for Model – 3 using 3800 Data Set*

Algorithms	Input	R-Squared	RMSE	MAE
XGBoost	Training	0.96	3.30	2.36
	Validation	0.73	8.37	6.64
DT	Training	0.82	6.96	5.52
	Validation	0.74	8.21	6.46
KNN	Training	0.81	7.28	5.39
	Validation	0.65	9.61	6.78
RF	Training	0.79	7.68	6.32
	Validation	0.65	9.61	6.78

Figure 4: *Actual and Predicted Students' Results*



The same trend as of Model 3 was continued in the results of Model 4 as shown in table 6 and figure 5. The XGBoost again showed best performance for training data with an accuracy of 96 % (R^2 = 0.96, RMSE = 3.11 and MAE value = 2.24). Its accuracy dropped to 73% (R^2 = 0.73, RMSE = 8.41 and MAE value = 6.67) for the testing data. The performance of DT was although better again in view of closer accuracies of 82 % (R^2 = 0.82, RMSE = 6.96 and MAE = 5.49) for training and 75 % (R^2 = 0.75, RMSE = 8.20 and MAE = 6.39) for validation yet lowered as compared to Models 1 & 2. The performance of KNN and RF was not found similar as compared to results of Models 1 & 2. The KNN showed an accuracy 81 % (R^2 = 0.81, RMSE = 7.32 and MAE = 5.32) for training which dropped to 63 % (R^2 = 0.63, RMSE = 9.90 and MAE = 6.79) for testing. The RF showed an accuracy of 79 % (R^2 = 0.79, RMSE = 7.71 and MAE = 6.33) for testing which dropped to 74 % (R^2 = 0.74, RMSE = 8.36 and MAE = 6.82) for testing.

Table 6: *R-Squared, RMSE and MAE for Training and Validation for Model – 4 using 3800 Dataset*

Algorithms	Input	R-Squared	RMSE	MAE
XGBoost	Training	0.96	3.11	2.24
	Validation	0.73	8.41	6.67
DT	Training	0.82	6.96	5.49
	Validation	0.75	8.20	6.39
KNN	Training	0.81	7.32	5.32
	Validation	0.63	9.90	6.79
RF	Training	0.79	7.71	6.33
	Validation	0.74	8.36	6.82

Figure 5: *Actual and Predicted Students' Results*

The performance of the four algorithms, namely XGBoost, DT, RF and KNN were evaluated and compared. The results indicate that both XGBoost and DT performed competitively with similar results in testing. However, the DT algorithm demonstrated the best performance as it had the lowest accuracy difference between training and testing data. On the other hand, the XGBoost algorithm displayed the highest predictive performance in training data but its performance dropped in the testing data.

Discussion

The existing body of knowledge has shown the ability of machine learning algorithms to predict the performance of students and the same is also demonstrated by the current study. The researchers employed eight machine learning algorithms on a dataset comprising academic transactions from LMS and SIS. Based on performance, four algorithms i.e. DT, RF, KNN, and XGBoost were selected for further analysis. DT showed the highest performance, followed closely by XGBoost, with RF and KNN trailing behind. The findings are in accordance with the existing literature like prediction of features to improve the course design (Gupta & Sabitha, 2019), prediction of students' achievements (Nuankaew et al., 2020), analysis of dropouts (Perchinunno et al., 2021), and identification of at-risk students (Gkontzis et al., 2022). Students' attributes have been categorized based on demographics, behavioural interactions, and performance. Their importance has been demonstrated by various studies that have selected relevant attributes through feature selection and expert opinion.

Developing countries are striving hard to improve academic standards and get their children and adults into educational systems. The investment in technology enabled learning (TEL) for developing smart and effective systems can help to overcome the infrastructure barrier and outreach education to remote parts. It may help the youth to acquire technical and digital skills that can increase the socio-economic capital and reduce the poverty. The development of smart systems needs a detailed analysis of students' need and their current state of learning which can be accomplished by establishing a field of inquiry and research using analytical and predictive models in localized settings as demonstrated in the current study.

Conclusion

This research has demonstrated the use of learning analytics and machine learning for predicting academic performance of students enrolled in an open and distance learning University of Pakistan. We extracted the best attribute subset of academic transactions from the platforms of LMS and SIS. We carried out four experiments with the objective to find the best correlation among the selected attributes with the highest accuracy achieved by the machine learning algorithms. We compared our results with the existing literature within the realms of regional studies, attribute selection, and studies conducted in the (ODL environment. Our findings contribute to the understanding of the use of AI and ML from both facilitation and hindrance perspectives in the context of teaching and learning. Our results validated the ability of machine learning algorithms for predicting performances of students on a localized dataset for sustainable education and enhancing the success ratio of students. The study concludes that SIS and LMS features used in this research has positive impact on students' academic performances. It suggests replacing old traditional statistical methods with emerging trends of data analytics through machine learning. Any hinderance caused due to excessive use of technology may be converted into an opportunity through better planning and robust implementation strategies as demonstrated by the current study. The analysis may be extended to a larger dataset with more features and on more courses offered in distance learning mode to pave the way for sustainable learning.

Acknowledgements

None.

Conflict of Interest

Authors declared NO conflict of interest.

Funding Source

The authors received NO funding to conduct this study.

ORCID iDs

Hanaan Sadeed Ahmad [1] https://orcid.org/0009-0000-4754-2788
Moiz Uddin Ahmed [2] https://orcid.org/0000-0001-6384-1841
Shahid Hussain [3] https://orcid.org/0000-0001-6455-4000

References

Behr, A., Giese, M., Teguim K, H. D., & Theune, K. (2020). Early prediction of university dropouts–a random forest approach. *Jahrbücher für Nationalökonomie und Statistik*, *240*(6), 743-789. https://doi.org/10.1515/jbnst-2019-0006

Bharara, S., Sabitha, S., & Bansal, A. (2018). Application of learning analytics using clustering data Mining for Students' disposition analysis. *Education and Information Technologies*, *23*, 957-984. https://doi.org/10.1007/s10639-017-9645-7

Brandao, I. V., da Costa, J. P. C., Santos, G. A., Praciano, B. J., Júnior, F. C., & Júnior, R. T. D. S. (2019). Classification and predictive analysis of educational data to improve the quality of distance learning courses. *Proceedings of IEEE Workshop on Communication Networks and Power Systems (WCNPS), Brasilia, Brazil*, (pp. 1-6). https://doi.org/10.1109/WCNPS.2019.8896312

Burgos, C., Campanario, M. L., de la Peña, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering*, *66*, 541-556. https://doi.org/10.1016/j.compeleceng.2017.03.005

Cechinel, C., Ochoa, X., Lemos dos Santos, H., Carvalho Nunes, J. B., Rodés, V., & Marques Queiroga, E. (2020). Mapping learning analytics initiatives in Latin America. *British Journal of Educational Technology*, *51*(4), 892-914. https://doi.org/10.1111/bjet.12941

Charitopoulos, A., Rangoussi, M., & Koulouriotis, D. (2020). On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010–2018. *International Journal of Artificial Intelligence in Education*, *30*, 371-430. https://doi.org/10.1007/s40593-020-00200-8

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5-22. https://doi.org/10.1177/0047239520934018

Dixit, P. (2020). *Data – Analysis – Action: The Learning Analytics Trifecta*. https://blog.commlabindia.com/elearning-design/learning-analytics-components

Gkontzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2022). A Predictive Analytics Framework as A Countermeasure for Attrition of Students. *Interactive Learning Environments*, *30*(6), 1028-1043. https://doi.org/10.1080/10494820.2019.1709209

Goodell, J., & Thai, K. P. (2020). A learning engineering model for learner-centered adaptive systems, *Proceedings of International Conference on Human Computer Interaction, Copenhagen, Denmark* (pp. 557-573). https://doi.org/10.1007/978-3-030-60128-7_41

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, 22-32. https://doi.org/10.1016/j.compedu.2018.12.006

Gupta, S., & Sabitha, A. S. (2019). Deciphering the attributes of student retention in massive open online courses using data mining techniques. *Education and Information Technologies*, *24*(3), 1973-1994. https://doi.org/10.1007/s10639-018-9829-9

Kew, S. N., & Tasir, Z. (2022). Learning analytics in online learning environment: A systematic review on the focuses and the types of student-related analytics data. *Technology, Knowledge and Learning*, 1-23. https://doi.org/10.1007/s10758-021-09541-2

Khalid, L., Malik, S., & Malik, M. (2023). Rigid Assessment Methods and Lack of ICT Use Preventing Inclusion of Hearing-Impaired Learners in Mainstream Schools. *Online Media and Society*, *4*(3), 67-74.

Li, K. C., Wong, B. T. M., & Ye, C. J. (2018). Implementing learning analytics in higher education: the case of Asia. *International Journal of Services and Standards*, *12*(3-4), 293-308. https://doi.org/10.1504/IJSS.2018.100215

Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016, April). Big data application in education: dropout prediction in edx MOOCs. *Proceedings* of *Second IEEE International Conference on Multimedia Big Data (BigMM), Taipei, Taiwan*, (pp. 440-443). https://doi.org/10.1109/BigMM.2016.70

Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering analyssis for classifying student academic performance in higher education. *Applied Sciences*, *12*(19), 9467. https://doi.org/10.3390/app12199467

Navarro, Á. A. M., & Ger, P. M. (2018). Comparison of clustering algorithms for learning analytics with educational datasets. *IJIMAI*, *5*(2), 9-16. https://doi.org/10.9781/ijimai.2018.02.003

Nuankaew, P., Nuankaew, W., Teeraputon, D., Phanniphong, K., & Bussaman, S. (2020). Prediction Model of Student Achievement in Business Computer Disciplines. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(20), 160-181. https://doi.org/10.3991/ijet.v15i20.15273

Perchinunno, P., Bilancia, M., & Vitale, D. (2021). A Statistical Analysis of Factors Affecting Higher Education Dropouts. *Social Indicators Research*, *156*(2), 341-362. https://doi.org/10.1007/s11205-019-02249-y

Rodrigo, M. M. T. A criticAl examination of the pre-conditions of learning Analytics Adoption in developing countries in southeAst AsiA. *INCludE uS All! dIRECtIONS fOR AdOPtION Of lEARNINg ANAlYtICS IN thE glObAl SOuth*, 44. https://core.ac.uk/download/pdf/479891415.pdf#page=49

Safdar, G., Javed, M.N., Amin, S. (2020a). Use of Internet for Educational Learning among Female University Students of Punjab, Pakistan. *Universal Journal of Educational Research, 8*(8), 3371-3380. DOI: 10.13189/ujer.2020.080809

Safdar, G., Khan, A.W. (2020). E-Learning: Current Scenario of Internet and Educational Learning among University Students of Punjab, Pakistan. *Journal of Educational Research*, *23*(1), 171-185.

Safdar, G., Rauf, A., Ullah, R., Rehman, A.U. (2020). Exploring Factors Leading to Quality Online Learning in the Era of Covid-19: A Correlation Model Study. *Universal Journal of Educational Research*, *8*(12A), 7324-7329. DOI: 10.13189/ujer.2020.082515

Saykili, A. (2018). Distance education: Definitions, generations, key concepts and future directions. *International Journal of Contemporary Educational Research*, *5*(1), 2-17. https://files.eric.ed.gov/fulltext/EJ1207516.pdf

Sivakumar, S., Venkataraman, S., & Selvaraj, R. (2016). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*, *9*(4), 1-5. https://doi.org/10.17485/ijst/2016/v9i4/87032

Sousa, E. B. D., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B., & Gašević, D. (2021). Applications of learning analytics in high schools: A systematic literature review. *Frontiers in Artificial Intelligence*, *4*, 737891. https://doi.org/10.3389/frai.2021.737891

Tsai, Y. S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., ... & Gašević, D. (2020). Learning analytics in European higher education—Trends and barriers. *Computers & Education*, *155*, 103933. https://doi.org/10.1016/j.compedu.2020.103933

Zawacki-Richter, O., & Qayyum, A. (2019). Open and distance education in Asia, Africa and the Middle East: National perspectives in a digital age (p. 140). Springer Nature. https://doi.org/10.1007/978-981-13-5787-9