PREDICTING AT-RISK STUDENTS AT DIFFERENT PERCENTAGES OF COURSE LENGTH FOR EARLY INTERVENTION USING MACHINE LEARNING MODELS

K.RENUKA[1], PABBATHI DEEPTHI REDDY[2], MOTHE SNIHITHA[3], P.BHANU PRAKASH REDDY[4]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT-Online learning platforms such as Massive Open Online Courses (MOOCs), Virtual Learning Environments (VLEs), and Learning Management Systems (LMS) provide students with the flexibility to learn based on their individual interests, regardless of time or location. Despite their many advantages, these platforms face several challenges, including lack of student interest, high dropout rates, low engagement, and issues with self-regulated behavior, where students are often required to set their own learning goals. To address these challenges, this study proposes a predictive model designed to identify at-risk students early on, enabling instructors to intervene and help boost student engagement and performance. The predictive model uses various machine learning (ML) and deep learning (DL) algorithms to analyze the learning behavior of students, based on a variety of study-related variables. The performance of different ML algorithms is compared using several metrics such as accuracy, precision, recall, support, and F-score. The ML algorithm yielding the best results is selected for creating a predictive model that operates at different points throughout the course. This model can assist instructors in identifying at-risk students early, thus facilitating timely interventions to prevent student dropouts. The study's findings indicate that students' assessment scores, engagement intensity (such as clickstream data), and time-dependent variables are critical factors affecting online learning outcomes. The experimental results revealed that the Random Forest (RF) algorithm performed the best, providing consistently strong results across all evaluation metrics: precision, recall, F-score, and accuracy. Specifically, the RF model achieved averaged precision scores of 0.60%, 0.79%, 0.84%, 0.88%, 0.90%, and 0.92%, averaged recall scores of 0.59%, 0.79%, 0.84%, 0.88%, 0.90%, and 0.91%, and averaged F-score values of 0.59%, 0.79%, 0.84%, 0.88%, 0.90%, and 0.91% at various stages of course length (0%, 20%, 40%, 60%, 80%, and 100%). These results suggest that predictive models, particularly those based on Random Forest, can be highly effective in improving student outcomes by allowing instructors to intervene before students become disengaged or at risk of dropping out.

INDEX TERMS- Online learning, predictive modeling, at-risk students, student engagement, machine learning, deep learning, dropout prediction, Random Forest, course length, learning behavior, academic performance, intervention, clickstream data.

I.INTRODUCTION

Rapid advancements in online learning platforms, such as Massive Open Online Courses (MOOCs), Virtual Learning Environments (VLEs), and Learning Management Systems (LMS), have revolutionized access to education by overcoming the limitations of time and space. These platforms allow students to learn according to their interests, providing them with unprecedented flexibility. However, despite the benefits, online learning platforms face several challenges, including student disengagement, high dropout rates, and the need for students to take responsibility for setting and achieving their learning goals. In response to these issues, this study proposes a predictive model to address the problems faced by at-risk students, thereby enabling instructors to intervene early in the course. By analyzing various student learning behaviors and data, the model aims to predict at-risk students and facilitate timely interventions, ultimately increasing student engagement and reducing dropouts. The predictive model is trained and tested using multiple machine learning (ML) and deep learning (DL) algorithms, comparing the performance of different models using metrics such as accuracy, precision, recall, and F-score. The goal is to identify at-risk students as early as possible in the course, allowing instructors to take action before student performance declines significantly.

The results of the study show that the use of machine learning techniques, such as Random Forest, can effectively predict student performance at different stages of the course, with significant accuracy at various percentages of the course length. The study highlights the importance of factors like assessment scores, engagement levels (clickstream data), and time-dependent variables in predicting at-risk students. By accurately identifying these students early in the course, instructors can deliver personalized feedback and take preventative measures to help students stay on track. This intervention could significantly reduce dropout rates by offering students the support they need to improve their performance. Moreover, the study emphasizes the potential of integrating personalized persuasion techniques into the predictive model, which could help motivate students to engage more fully in their studies. The study concludes with a discussion on how these findings could lead to the development of more effective intervention strategies in online learning environments, improving student retention and academic success.

II.LITERATURE SURVEY

A) L. P. Macfadyen and S. Dawson, ‘‘Mining LMS data to develop an ‘early warning system’ for educators: A proof of concept,’’ Comput. Edu., vol. 54, no. 2, pp. 588–599, Feb. 2010.

In this paper, the authors explore the use of data mining techniques applied to Learning Management System (LMS) data to create an "early warning system" for educators. The aim is to predict students at risk of underperforming or dropping out by analyzing various behavioral patterns and engagement metrics within the LMS. Using real-time data, such as login frequency, assignment submissions, and participation in discussions, the authors propose a model that can detect early signs of student disengagement. The study utilizes a variety of data mining methods to identify patterns and predict student performance, demonstrating the potential of LMS data in improving educational outcomes. The results show that this early warning system can significantly aid educators by highlighting students who may need additional support or intervention before it is too late. Additionally, the paper discusses the implications of such a system in terms of practical applications and the potential for improving student retention rates. The proposed system represents an important step forward in integrating technology and data-driven strategies into educational practice, offering a way to proactively support students.

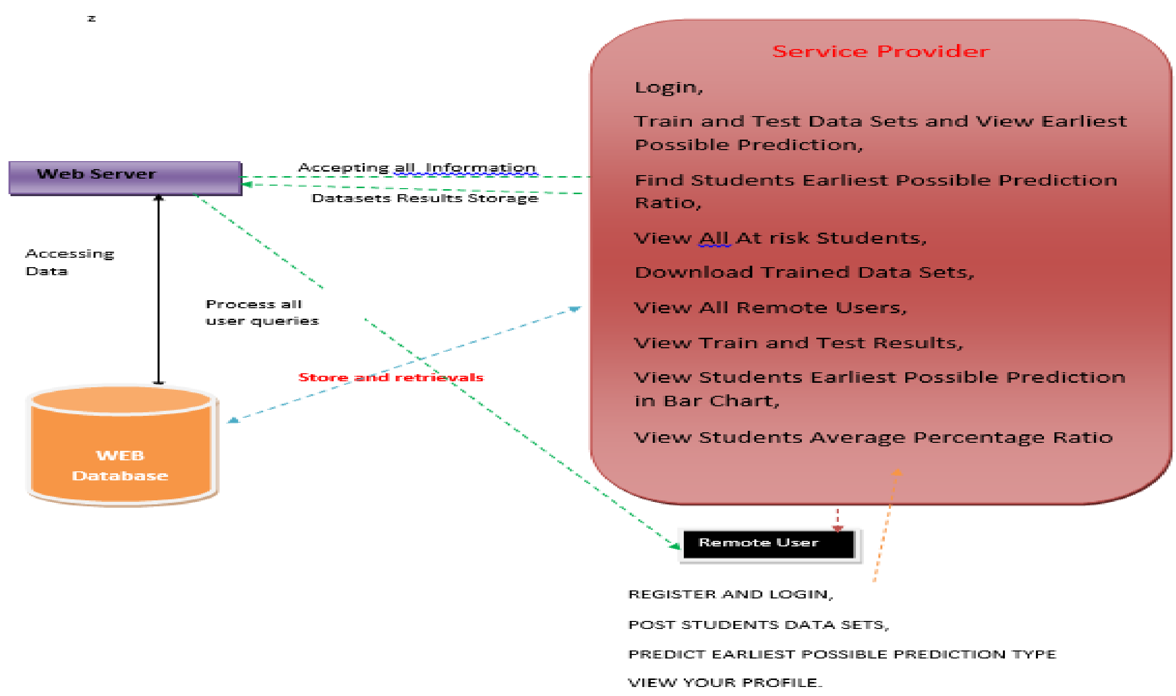
B) C. Romero, S. Ventura, and E. García, ‘‘Data mining in course management systems: Moodle case study and tutorial,’’ Comput. Edu., vol. 51, no. 1, pp. 368–384, Aug. 2008.

In this paper, the authors present a case study on the application of data mining techniques to course management systems (CMS), specifically focusing on Moodle, a widely used open-source learning platform. The study explores how data mining can be used to analyze students' behavior and interactions within the Moodle platform, with the goal of identifying patterns that can inform educational strategies. The authors detail various data mining techniques, such as clustering, classification, and association rule mining, to extract meaningful insights from student activity logs, forum participation, and quiz performance. By applying these techniques, the study demonstrates how educators can gain a deeper understanding of student engagement, learning behaviors, and potential academic risks. Furthermore, the paper includes a tutorial that guides readers through the process of using Moodle's built-in features alongside external data mining tools to perform similar analyses. The results highlight the potential of using data mining to improve learning outcomes by providing personalized feedback and early interventions for at-risk students. The paper concludes by discussing the broader implications of integrating data mining into CMS to enhance the learning experience, optimize course delivery, and foster more effective student-teacher interactions.

C) S. Valsamidis, S. Kontogiannis, I. Kazanidis, T. Theodosiou, and A. Karakos, ‘‘A clustering methodology of Web log data for learning management systems,’’ J. Educ. Technol. Soc., vol. 15, no. 2, pp. 154–167, 2012.

In this paper, the authors propose a clustering methodology for analyzing web log data from Learning Management Systems (LMS), aiming to uncover patterns in student behavior and enhance the understanding of their interaction with online courses. The study focuses on applying clustering techniques to web log data, which includes students' activity, page views, logins, and time spent on the LMS platform. By using unsupervised machine learning algorithms, such as k-means clustering, the authors categorize students based on their usage patterns, identifying distinct groups of learners with similar behaviors. The results show that clustering can help in recognizing different types of learners, such as active, passive, and disengaged students, which can inform personalized interventions and improve course design. Additionally, the study emphasizes how clustering techniques can be used to detect potential issues like low engagement or at-risk behavior, enabling instructors to take proactive measures to support students. The paper also discusses the challenges of applying clustering to large-scale web log data and the importance of selecting relevant features for accurate analysis. The authors conclude that this methodology can be an effective tool for LMS administrators and educators to better understand student interactions, optimize course content, and provide tailored learning experiences.

III.PROPOSED SOLUTION



Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

Predicting and intervening with students at different stages of a course provides significant benefits for both students and instructors. This approach allows instructors to identify at-risk students early on, offering the opportunity to make timely interventions that can help improve study behavior and reduce dropout rates. Among the various factors analyzed, clickstream and assessment variables were found to have the most significant impact on student performance. The study demonstrated that feature engineering plays a critical role in enhancing the performance of predictive models. The students' performance was initially predicted using only demographic data and then at different stages of the course (20%, 40%, 60%, 80%, and 100% of the course length). At 20% course completion, the Random Forest (RF) predictive model achieved a promising performance with an average precision, recall, F-score, and accuracy of 79%. At the 60% mark, the RF model's performance improved to 88% across all metrics. Finally, by the end of the course, the model reached its peak performance, with average precision, recall, F-score, and accuracy scores of 92%, 91%, 91%, and 91%, respectively. Interestingly, the Fail class performed better after feature engineering, likely due to an imbalanced class distribution, with 17,208 Fail students and 15,385 Pass students. Overall, the results of the RF predictive model showed effectiveness in predicting student performance early in the course, making it a valuable tool for VLE administrators and instructors to assist in decision-making. However, further research is needed to assess how various online activities can influence student performance, as well as to explore the implementation of different early intervention techniques. Future plans include investigating the impact of students' feedback using deep learning models and natural language processing techniques.

REFERENCES

[1] L. P. Macfadyen and S. Dawson, ''Mining LMS data to develop an 'early warning system' for educators: A proof of concept,'' Comput. Edu., vol. 54, no. 2, pp. 588–599, Feb. 2010.

[2] C. Romero, S. Ventura, and E. García, ''Data mining in course management systems: Moodle case study and tutorial,'' Comput. Edu., vol. 51, no. 1, pp. 368–384, Aug. 2008.

[3] S. Valsamidis, S. Kontogiannis, I. Kazanidis, T. Theodosiou, and A. Karakos, ‘‘A clustering methodology of Web log data for learning management systems,’’ J. Educ. Technol. Soc., vol. 15, no. 2, pp. 154–167, 2012.

[4] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, ‘‘Using machine learning to predict student difficulties from learning session data,’’ Artif. Intell. Rev., vol. 52, no. 1, pp. 381–407, Jun. 2019.

[5] O. E. Aissaoui, Y. E. A. El Madani, L. Oughdir, and Y. E. Allioui, ‘‘Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles,’’ Procedia Comput. Sci., vol. 148, pp. 87–96, Jan. 2019.

[6] J. Y. Chung and S. Lee, ‘‘Dropout early warning systems for high school students using machine learning,’’ Children Youth Services Rev., vol. 96, pp. 346–353, Jan. 2019.

[7] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, ‘‘Mining in educational data: Review and future directions,’’ in Proc. Joint EuropeanUS Workshop Appl. Invariance Comput. Vis. Cairo, Egypt: Springer, 2020, pp. 92–102.

[8] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, ‘‘A systematic review of deep learning approaches to educational data mining,’’ Complexity, vol. 2019, May 2019, Art. no. 1306039.

[9] K. S. Rawat and I. Malhan, ‘‘A hybrid classification method based on machine learning classifiers to predict performance in educational data mining,’’ in Proc. 2nd Int. Conf. Commun., Comput. Netw. Chandigarh, India: National Institute of Technical Teachers Training and Research, Department of Computer Science and Engineering, 2019, pp. 677–684.

[10] S. M. Jayaprakash, E. W. Moody, E. J. M. Lauría, J. R. Regan, and J. D. Baron, ‘‘Early alert of academically at-risk students: An open source analytics initiative,’’ J. Learn. Analytics, vol. 1, no. 1, pp. 6–47, May 2014.

[11] C. Márquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. Mousa Fardoun, and S. Ventura, ‘‘Early dropout prediction using data mining: A case study with high school students,’’ Expert Syst., vol. 33, no. 1, pp. 107–124, Feb. 2016.

[12] S. Palmer, ‘‘Modelling engineering student academic performance using academic analytics,’’ Int. J. Eng. Edu., vol. 29, no. 1, pp. 132–138, 2013.

[13] Z. Papamitsiou and A. Economides, ‘‘Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence,’’ Edu. Technol. Soc., vol. 17, no. 4, pp. 49–64, 2014.

[14] A. Peña-Ayala, ‘‘Educational data mining: A survey and a data miningbased analysis of recent works,’’ Expert Syst. Appl., vol. 41, no. 4, pp. 1432–1462, Mar. 2014.

[15] N. Z. Zacharis, ‘‘A multivariate approach to predicting student outcomes in Web-enabled blended learning courses,’’ Internet Higher Edu., vol. 27, pp. 44–53, Oct. 2015.