ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.3 Sep 2023

Using Machine Learning Approach, Detecting At-Risk Students and Providing Early Intervention

Dr. A. AVANI

Abstract- Information and communication technology (ICT) is now widely used and plays an important role in education. ICT has helped to promote academic curricula and has made it possible to create a virtual classroom. ICT has the potential to enhance student results by allowing teachers to assist students in completing tasks. As a result, virtual learning might be used to give high-quality instruction. Massive Online Courses have grown in popularity in higher education as a result of the current ICT boom. MOOCs use a wide range of multimedia techniques to create an engaging learning environment. MOOCs provide students with important digital learning tools by enabling them to access material from across the globe.

A number of top-ranked colleges have implemented classes online as an alternative to conventional learning as a result of the collapse of financial & geographical barriers connected only with traditional teaching style. With the fast advancement of technology. Low completion rates are a key concern with MOOCs, which is growing in popularity in higher education. One of the ways for increasing completion rates is to identify at-risk students. Detecting at-risk students early on might aid instructors in providing educational interventions and improving course structure . Instructors may offer real-time instruction to learners with a fast intervention solution, and retention rates might increase.

Index Terms- MOOC, CNN , ICT

I. INTRODUCTION

Researchers looked at the causes for course cancellation.A variety of variables have been blamed for this.Lack of motivation is the most common cause for dropout out of virtual classrooms. Students' motivation levels in virtual classrooms are said to diminish or rise depending on aspects. A student a key determinant of their likelihood of dropping out. Changes in student behaviour over courses may be used to evaluate motivational trajectories. Most academics haven't looked at the link between psychological, performance, context until now.

Anticipating students' learning in MOOCs may offer instructors with vital information to assist them identify at-risk individuals early. Despite a number of papers in the literature offering robust learning methods for learning, it remains difficult to obtain good accuracy of graduation rates across numerous datasets in the long run.

In this project, two empirical studies are undertaken. The first research presents a unique dropout predicting model that might provide timely intervention help. Used to analyse student historical behaviour and predict prospective instructor. The influence of student involvement, as well as motivational status in prior courses, on students continuing to participate in the current course was investigated. In the given case study, a model for predicting student performance is provided. The methodology provides fresh insight into the most important aspects of learning activities or may help instructors track student progress. Used to monitor offer useful educators so that they may continue with their courses based on their learning progress. It might also assist academic advisers in detecting students with poor academic success and providing assistance to them.

* Assistance Professor, Department of Computer Science and Engineering, ANUBOSE Institute of Technology, Palvancha.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 14, Issuse.3 Sep 2023

Multilayer Perceptron (MLP), Gradient Boosting, Generalized Linear Model and Feed Forward Neural Network to detect students who may withdraw from course due to interventions (performance or bad grades). We are using OULAD(Open And distance Learning Analytics Dataset) university dataset to train above machine learning algorithms and this dataset contains lots of missing values and contains irrelevant features so author using PCA feature selection algorithm to remove irrelevant features and select only relevant features. From above OULAD dataset identifying 3 different types of students

1) Amotivation: students who withdraw from courses within one week of register
2) Extrinsic: students who perform well in the course
3) Intrinsic: students who underperform in course

After finding above features from dataset author is training all algorithms and then evaluating their, FSCORE, AUC, Sensitivity Specificity.

II. Existing System

In MOOCs, student disengagement and learning outcomes are key concerns. We present a summary of the most recent studies on detecting at-risk kids in terms of dropout and failing in this section.

In, neural network models were used to identify at-risk users in MOOCs utilising baseline information such as student emotions and clickstream. In 2014, the gathered undergraduate as well as postings on website. One of the key challenges in this work was dealing with an unbalanced dataset. Instead of accuracy, Cohen's Kappa criteria were used to overcome this. Since both sets of characteristics were used, the findings showed a 74 percent accuracy rate. When sentiment characteristics were removed, the percentage dropped to 70%.

Researchers looked at a variety of factors to determine the amount of student learner accomplishment in an online context, including spent interacting, the tries made. The feature set was optimised using Genetic Algorithms (GA) in. The results revealed that high-ranking traits are linked to behavioural characteristics rather than demographic characteristics. To predict student performance, four classifiers were considered: decision tree, perceptron, Nave Bayes, and k-nearest neighbour. When employing the GA-optimized feature set, showed increased. Before decision tree well with full feature set, accuracy was 83.87 percent, but, accuracy was 94.09 percent. Machine(svm models were utilised to see how latent and observable factors interact to affect student performance in online courses. Suggested a multiple infer dormant student behavioural patterns. The ability of TL-HMM to identify the micro-behavioralpupils transitions sets it apart from traditional HMM.

TABLE 1. Overview of previous research in the identification of at-risk students in MOOCs.

Author	Year	Features	Results
Minaei-Bidgoli et al.,[13]	2003	Click stream features	GAimproved by 12% for All classifiers.
Chaplot et al., [10]	2015	Sentiments , click stream features	Neural network attain higher performance , when using sentiment features.
He et al., [11]	2015	Click stream features	Regularized logistic regression acquired the best AUC.
Geigle et al., [8]	2017	Behavioural attributes	TL-HMM is able to infer latent behavioural patterns
Wasli et al., [12]	2018	Behavioural attributes	Deep learning is able to extract features automat

.

III. PROPOSED SYSTEM

DESCRIPTION OF THE INFORMATION:

In this project, we used two datasets. The first collection comes from online courses at Open University online courses and second one is take as a test dataset.

This dataset includes demographic, behavioural, and temporal data. It contains a number of tables relating to student performance, personal information about students, and student participation with online courses. The learner may engage with a variety of digital content, including, admittance. The Tutor MarkingAssessment and the Computerized Marked Assessment are the two forms of assessments (CMA). The weighted total of all exams (50 percent) and final examinations is used to get the final average rating (50 percent). The "Student Assessment" table contains information on student test scores, namely the date the assessment was filed and the grade received. In the dataset, the evaluations are required. As a result, if students wish to stay in the course, they must complete evaluations (including a final exam). A student will pass the course if his or her total grade is more than 40%. The OULAD dataset is briefly summarised in Table 3.

Data from the then features were extracted. Clickstream features are used to extract VLE features. There are eleven different VLE activity types in the OULAD dataset. From the first moment a student enrolled in the course to the final day they dropped out, we tallied the clicks and conversions they made on each activity for each

ISSN2454-9940
www.ijsem.org
Vol 14, Issue.3 Sep 2023

student. Similar to prior work twenty-two characteristics are retrieved from the VLE.

IV. PROBLEM DEFINITION

Learning Analytics (LA) methods were used in prior work to describe the students' motivated level using Incentive Motivation Theory. According to the abbreviation for the course. Framework for at-risk students. Learners are divided into three types in this theory: involved in this industry, motivation. varies and throughout, which may influence a student's choice to drop out.

Based on the notion of course trajectories, we present a method (Algorithm 1) for identifying at-risk consumers in online courses. In our method, two intervals are specified (T1, T2). Learners having participated in just autumn semester courses are classified in T1, whereas those who participate in both fall & spring courses are included in T2.

According to, there are three types of learners: fundamental (RL), external (CLsc, CLsn), &amotivation (Al). To discriminate between failed and successful extrinsic learners, the assignment threshold grade (40 percent) was used. Amotivation students are those that dropped out of a course within the first seven days. During the spring semester, motivating student is classified as withdrew. By finding patterns underlying student motivation trajectories, the programme provides a substantial contribution. The suggested methodology may help course instructors provide.

Low student learning and learning accomplishment outcomes have been highlighted as major reasons in students dropping out of online courses. In the present case study, however, students are associated with a greater risk if they drop out of spring classes within one week. This is due to the fact that evaluating teaching experiences in a really short amount of time is impossible.

When looking at the most important aspects that influence student learning outcomes, a data-driven solution should be explored. A student learning accomplishment model is presented to investigate how such variables impact students who are already at risk of failing.

ci - The name of the course that is linked to the ith item.

li - Date of the start of the course mentioned by ci

wi - ci awarded a certification with a wrap date.

V. PRE-PROCESSING OF DATA

Cleaning the data will be the first task in pre-processing, which involves detecting missing values. The "Nevent", "nplay video", "Nchapters", "nforum post", "YOB", "Gender", and "LoE DI" properties, for example, have null values in the Harvard dataset. Missing values or other errors are removed from the data. Additionally, redundant entries in student data are deleted.

The Princeton dataset is not distributed regularly. The use of transformation techniques was used to solve this issue. The data distribution was transformed into normal using the BOX COST transformation.

TABLE 5. Box-Cox transformation harvard dataset.

Features	Sample Skewness	Estimated Lambda
userid_DI	0.0135	0.1
final_cc_cname_DI	-0.569	1.2
LoE_DI	-0.163	0.7
YoB	-1.4	2
start_time_DI	-0.107	0.7
last_event_DI	0.0376	0.7
nevents	3.18	-0.1
ndays_act	1.76	0
nplay_video	6.21	0.1
nchapters	1.07	-0.4

The OULAD dataset's extracted behavioural traits and demographic variables are subjected to pre-processing in order to attain the best results. The investigation of highly correlated variables is the initial stage in pre-processing the data. We used a correlation cutoff of 0.8, which means that if the correlation characteristics is more than 0.8, they are deemed strongly associated. Given that the issue of feature redundancy might be overcome.

Because collection, transformation techniques are used to solve the issue. One of the statistic transformation techniques is Yeo-Johnson, which performs a comparable performance to that same.

VI. IMPLEMENTATION

MODULES :

Upload Student Dataset :

Click the button and choose the dataset to upload.Selecting and uploading the 'OULAD.csv' dataset , then clicking the 'Open' button to load the dataset.

Preprocess& PCA Feature Selection

The dataset is loaded, and we can see that there are many non-numeric values, which the data scientist will not accept, so we must process it to convert it to numeric format. We found three different users: amotivation, extrinsic non-wiithdraw, and intrinsic non-wiithdraw.

Amotivation: students who withdraw from courses within one week of register

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 14, Issue.3 Sep 2023

Extrinsic: students who perform well in the course

Intrinsic: students who underperform in course

The x-axis in the graph above reflects student type, while the y-axis represents the number of students of that category. Close the above graph, then select the 'Preprocess& PCA Feature Selection' button to convert non-numeric data to numeric format, and then use the PCA technique to acquire the results.
we can see that all non-numeric values have been replaced with numeric values, and then we can view the total columns before and after PCA, and PCA will eliminate unimportant columns, resulting in a smaller column size with just the most significant characteristics. The total

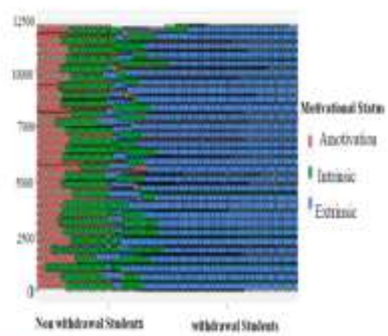


FIGURE 3. Distribution of learners according to motivational status.

records & record s
used for training and certification ML algorithms.

The relationship between motivating sculptures and at-risk students is well seen in Figure 3. It reveals that students drop out within a week of starting the spring semester. In the next course, around 31% of business confidence students withdrew, whereas the percentage of withdrawal students for intrinsically & extrinsically motivated students dramatically climbed. In the spring semester, 84 percent of intrinsically driven.
The OULAD dataset employs to eliminate duplication caused by the existence of strongly correlated characteristics across the retrieved features. Because only learners' actions are used to measure student success, this is solely applied to behavioural aspects. The Kaiser technique is used to calculate the is shown in Figure4.In this dataset, the best number of main components was determined to be 10. Figure 5 displays the Kaiser method's findings, which reveal that nine sections are chosen as the best.

Result and Analysis :

fig 8.1 output view for the project

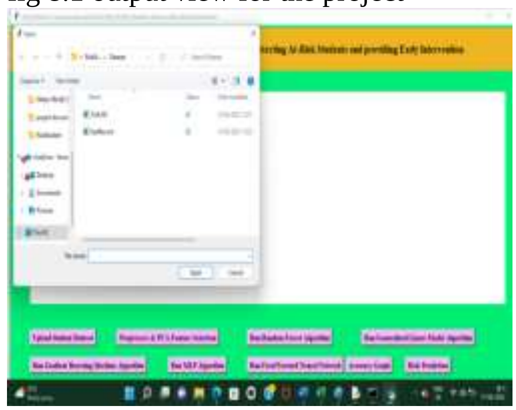


Fig 8.2 Upload Student Dataset

fig 8.2.1 after upload dataset motivational status graph

fig 8.3 preprocess data

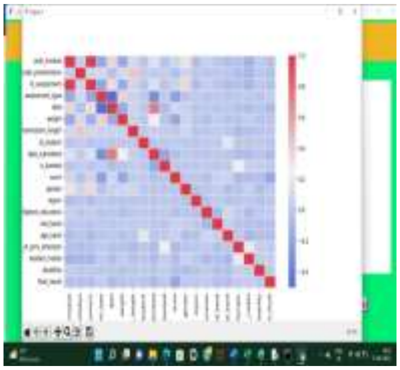


fig 8.3.1 preprocess data graph

fig 8.4 click on random forest algorithm tab

fig 8.5 click on linear model algorithm tab

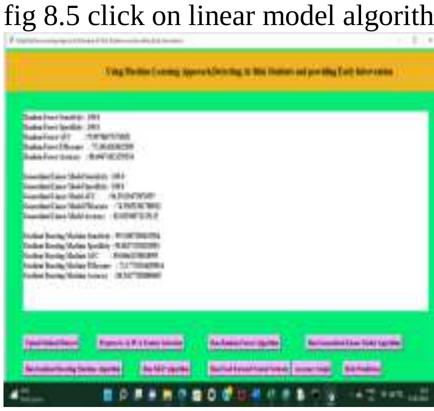


fig 8.6 click on gradient boosting machine algorithm tab

fig 8.7 click on mlp algorithm tab

fig 8.8 click on feed forward neural network algorithm tab

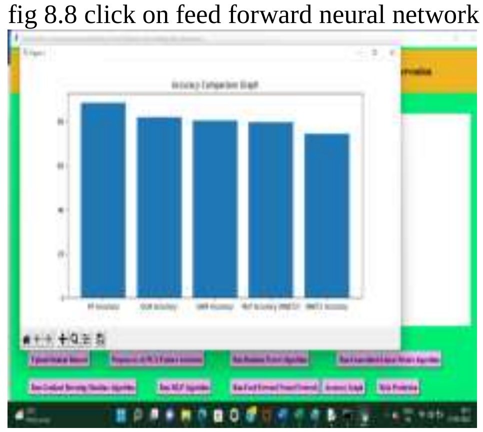
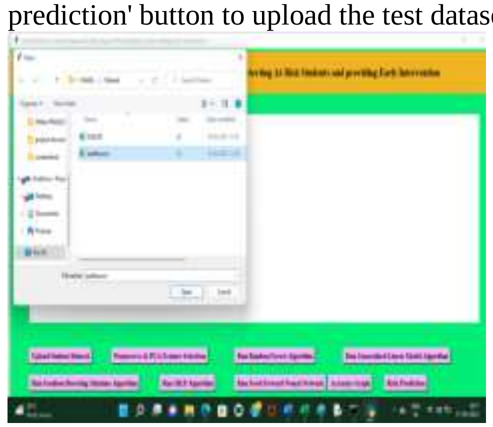


fig 8.9 click on accuracy graph tab now, click the 'risk prediction' button to upload the test dataset.



ISSN2454-9940
www.ijsem.org
Vol 14, Issue.3 Sep 2023

Fig 8.10 Risk Prediction

Fig 8.11 Prediction Data Withdrawor Non-Withdraw

VI. CONCLUSION

Educators may use the drop prediction model to provide help. Data suggest engagement leading cause of online course abandonment. Feature selection improves machine learning models' prediction abilities while lowering their computing costs. Furthermore, the feature selection filter approach seems to be a potential solution to the overfitting issue. The findings of this research may aid instructors in tracking changes in student motivation, allowing them to identify children who need further assistance.

In the learning accomplishment model, many variables impacting analysed. Both datasets show that clickstream properties are major determinants that are strongly linked to students failure in virtual classrooms.

FUTURE ENHANCEMENT:

In terms of future research, we want to examine the suggested framework's validity using more datasets. To analyze subject trends, it would to collect statistics several that give within themes. By trying to imply the sequences between temporal events across diverse MOOC datasets, supervised characteristics. As a result, coevolutionary may utilized assess & motivational state, as well as determine the influence of these factors on at-risk students.

.

REFERENCES

[1] Dr. Ghaznavi, A. Keikha, with N.-M. Yaghoubi, "The influence of technological tools (ICT) on educational progress," International Journal of Educational Research , "Student involvement in massive open online courses," Int. J. Teach.

[2] J. Sinclair and S. Kalvala, "Student engagement in massive open online," Int. J. Learn. Appl., vol. 11, no. 3, pp. 218–237, 2016.

[3] "Understanding the web based course (MOOC) student skill: An examination of sentiments, possible motives, and barriers," H. B. Wright, C. H. Cho, N. E. W. Roth, K. Li, M. etinkaya-Rundel, and Alan Canelas,

[4] J.-L. Hue, M. C. Wang, S. Wang, Prof. Abdelrasoul, Y. Li, & W. He, "Identifying at-risk students for brief detection time-series clustering method," Acm Trans. Emergency. Topics Comput., vol. 5, n. 1, pp. 45–55, January/March 2017.

[5] A. s. Alshabandar, A. Hasan, R. Keight, Aspergillus Laws, and M. a. Baker, "The use of Sparse representation models for such assessment of at-risk trainees in web - based learning," in Proc. International Congr. Evol. Computing. (CEC), pp. 1–8.

[6] M. Barak, A. Watted, and H. Haick, "Motivation to learn in web - based learning: Examining vocabulary and grammar and social engagement," in "Motivation to learn in massive open online: Examining vocabulary and grammar and social engagement," in "Motivation to learn in web - based learning:

[7] Altair is a fictional character (2019). Improve the performance of your retail store using in-store analytics. [Online]. Visit https://www.datawatch.com/in-action/usecases/retail-in-store-analytics/ for more information.

[8] R. Chaplot, E. Rhim, et J. Kim, "Using text analytics and neural networks to predict student attrition in MOOCs," in Proc. 17th Int. Confer. Artif. Operation is accomplished. Educ., 2015, vol. 7–12.

[9] "Identifying at-risk students in online classes," in Proc. 33rd AAAI Conf. Artif. Intell., 2015, pp. 1749–1755; J. He, J. Bailey, B. I. P. Rubinstein, and R. Zhang, "Identifying at-risk students in online learning,"

[10] B. Close to the actual values, D. A. Kashy, G. Kortemeyer, & W. F. Punch, "Predicting student performance: An use of data mining approaches with or without an educational web-based system," in Proc. in whatever Annu. Frontiers Educ. (Excellent), vol. 1, December 2003, pp. T2A13–T2A18.

AUTHORS

Dr. A. Avani was born in Khammam ,Telangana. She Post Graduated from the Jawaharlal Nehru Technological University, Hyderabad . Presently she is working as an Associate Professor in Anubose Institute of Technology, Palvancha. So far she is having 15 Years of Teaching Experience in various reputed engineering colleges. She special fields of interest included Data Base Management Systems, Computer Networks, Data Warehousing Data Mining, Cloud computing and Big data Analytics.