An Early Intervention Strategy For Identifying Students at Risk

1.D.Swathi, Assistant Professor,CSE, Gokaraju Rangaraju Institute of Engineering and Technology, Telangana, India, Swathi1681@grietcollege.com

2 Pagadala Srujana ,Btech, Department of CSE, pagadalasrujana459@gmail.com

3 Sai Lalitha Shreya Bandhakavi ,Btech, Department of CSE, bslshreya@gmail.com

4 Sunkara Anu Pallavi,Btech, Department of CSE, anu12102001@gmail.com

5 Zeba Naaz,Btech, Department of CSE, naazzeba476@gmail.com

ABSTRACT: The application of machine learning algorithms for various educational objectives has been the subject of much study in the e-learning sector. This research uses log data from learning management systems to forecast dropout rates or at-risk students in online courses, which is one of its main objectives. Students now have access to high-quality digital information as a result of the massive rise of Massive Open Online Courses (MOOCs). Because of supported learning and the instructional environment's versatility, the number of participants is growing quickly. Numerous studies have identified the high dropout rate and poor completion rate as key problems. This research enables the early detection of students who are at danger of dropping out or failing their classes. Several machine learning algorithms, including Random Forest, Multilayer Perceptron (MLP), Gradient Boosting, Generalized Linear Model, and Feed Forward Neural Network, are used to identify students who may drop out of a course as a consequence of interventions. The algorithms can swiftly identify students who run the danger of failing or dropping out of an online course. The findings demonstrate that all models' classifiers achieve high accuracy.

Keywords- *Random Forest, Multilayer Perceptron (MLP), Gradient Boosting, Generalized Linear Model, and Feed Forward Neural Network.*

1. INTRODUCTION

Modern educational institutions, especially those in higher education, now operate in a highly competitive and challenging environment as a result of the growth of the World Wide Web. Through e-learning platforms like intelligent tutoring systems (ITS), learning management systems (LMS), and massive open online courses, students may enrol in a range of online courses to study after school or learn something new (MOOC). Higher education institutions now have a variety of alternatives for surviving in the cutthroat market of today. Although education has advanced, universities are still faced with problems such increased student dropout rates, academic underachievement, graduation delays, and other enduring problems [1,2]. As a consequence, the majority of educational institutions prioritise developing automated tools for evaluating student performance, offering top-notch training, developing plans for evaluating students' academic progress, and foreseeing future demands.

One of the study's strengths appears to be its use of both time-varying and time-invariant data. However, these experiments demonstrated that using only time-varying data resulted in improved accuracy in subsequent rounds and that using only time-invariant data, such as gender or experience, did not initially lead to a reliable classification. Classification success cannot be determined using only time-varying data instead of time-invariant data. Data classification into only two categories is another common error. The primary goal of these research was to identify at-risk kids rather than to determine student performance levels. However, categorising pupils based on their performance levels (e.g., average performance, bad performance, worst performance, etc.) may be more informative. Instructors may offer more adaptable feedback to each student in this manner. As a consequence, the small number of classifications seems to be another significant drawback of these investigations.

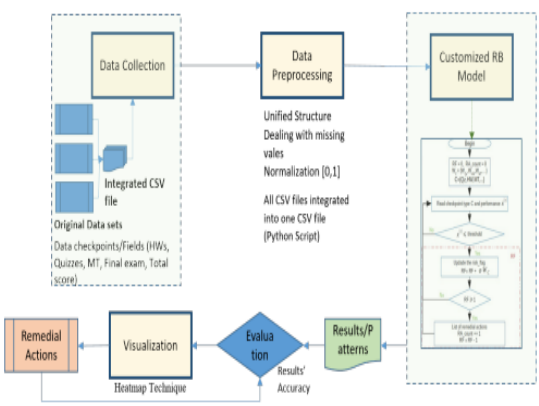

Fig.1: High-level flowchart for rule-based model

Researchers investigated the factors that lead students to drop out of classes in order to develop an accurate model for predicting students who are at risk. There are a number of possible causes for this. Lack of enthusiasm is the most common reason students drop out of online classes [6]. Researchers found that students' levels of motivation in online courses may be influenced by social, cognitive, and environmental variables [7]. A key variable in the prediction of student dropout is the trajectory of motivation. Examining how students behave throughout a course can be used to evaluate motivational trajectories.

2.LITERATURE REVIEW

2.1 The impact of information and communication technology (ICT) on educational improvement [1] :

Data were gathered using a 24-question survey designed by a scientist with a grading size of Likret type, and Cronbach's alpha was used to show its reliability and consistency, according to M. R. Ghaznavi, A. Keikha, and colleagues. In the part on illustrative insights, the recurrence conveyance table, recurrence rate, and graph drafting were used. In the region of inferential measures, the estimation scale was tested using the Khi (chi-square), U Mann-Whitney, and Kruskal-Wallis factual tests. The data were broken down using SPSS. According to the emphasis on outcomes, effective data and communications technology use boosts educational motivation, enhances communication skills, fosters a spirit of inquiry, and promotes academic performance. It significantly affects how well third graders in secondary schools are doing academically. This impact was seen in understudies who were both male and female and had a range of normal scores, ages, and callings. Despite this, secondary school and professional students were the only ones who were affected.

2.2 Student engagement in massive open online courses [2]:

Massive open online courses (MOOCs) have disturbingly low completion rates, say J. Sinclair, S. Kalvala, and others. Previous studies have focused on asset access examples and dropout predictions using learning examination. In contrast, evaluations of student commitment are increasingly being used to quantify educational project presentation in conventional higher education (HE) settings worldwide. Compared to the current definition of the term in relation to MOOCs, this commitment idea goes much further and is more shrewd. In the light of this deeper comprehension of student association, this article examines MOOC support, learning, and dropout. We examine the methods employed in MOOC teaching as well as the potential applications of HE commitment metrics in this setting. We acknowledge the need for a commitment model for MOOCs and provide recommendations for main, first-step actions that MOOC developers may utilise to further pique interest.

2.3 Identifying at-risk students for early interventions—A time-series clustering approach [4] :

Time-series clustering risk online students was utilised more efficiently, often, and earlier, according to J.-L. Hung, M. C. Wang, S. Wang, M. Abdelrasoul, and colleagues. The contextual analysis demonstrated that, in comparison to existing recurrence accumulation methods, the suggested strategy might provide models with more precision and common sense. As early as week 10, the best model might start catching at-risk students. In addition, the four phases of the understudy educational experience recognized the occasion's impact and identified activities that put the understudy in danger during a prolonged occasion break. The results also make it possible for online educators to provide nearly identical instruction via clearly plan or understudy instructor exchange.

2.4 The application of Gaussian mixture models for the identification of at-risk learners in massive open online courses [5]:

According to T. Bread Cook et al., R. Alshabandar, A. Hussain, R. Keight, and A. Regulations, early identification of at-risk student groups is becoming simpler in MOOC environments, where student withdrawal rates are high. Despite the fact that many earlier research have identified succession grouping as a problem in dropout, these studies only target a small number of conduct elements, which are generally organised on a weekly basis. They exclude crucial context-specific elements like work deadlines, which may be substantial contributors to understudy inactive commitment. Therefore, the goal of this study is to examine how to take into account such important elements using Gaussian Mixture Models. Insightful analysis of the effects of a dormant commitment on students and their possible risk of leaving the course will be provided by this. Additionally, k-closest neighbours and straight relapse classifiers were used to analyse execution.

2.5 Motivation to learn in massive open online courses: Examining aspects of language and social engagement [6]:

M. Barak, A. Watted, and H. Haick looked at social collaboration and the language of education among other things. The two criteria may be crucial in choosing what subjects to emphasise in massive open online courses (MOOCs). The review's three goals were as follows: Investigate the relationships between various methods of commitment and inspiration gain;

look at instances of MOOC participants who are concentrating on the same course but utilising a different language of teaching; and define MOOC participants in light of their learning motivation. An exploratory contextual study of a MOOC on nanotechnology and nanosensors was conducted and was given in two dialects: English and Arabic 325 English and Arabic MOOC participants (N = 289) were included in the review. Utilizing a combination of tactics, the inquiry gathered information from pre- and post-polls, message boards, and email exchanges. The results show that participants in MOOCs were driven to succeed by equivalent goals, independent of the language of teaching, which promoted intrinsic motivation and self-assurance. The results revealed a connection between the rise of inspiration and the quantity of messages submitted in online forums, focus groups, and conversations. Supporters, trend-setters, problem-solvers, organisers, and free students are the five categories of MOOC participants.

3. IMPLEMENTATION

Past exploration utilized Learning Analytics (LA) strategies to depict understudies' persuasive level utilizing Impetus Inspiration Hypothesis. This approach separates students into three classes: amotivation, outward, and characteristic inspiration. Understudy inspiration differs over the long haul and all through courses, which might impact an understudy's decision to drop the course. Understudies who are naturally determined or amotivated can't be surveyed for this dataset since they are supposed to participate in assessments in OULAD courses. Since an outcome, the in danger understudy recognizable proof methodology is solely utilized with the Harvard dataset, as the objective is to examine how inspiration directions influence in danger understudies. By following understudy ways of behaving across time, learning directions might support online course investigation. LA is utilized in this work to screen learning directions more than a few courses. Figure 3 portrays the system for in danger understudies.

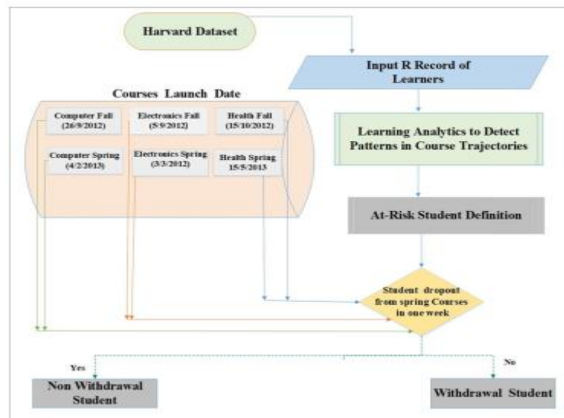

Fig.2: At-risk student framework

The creator of this review utilize an assortment of ML strategies, including Random Forests, Multilayer Perceptron (MLP), Slope Helping, Summed up Direct Model, and Feed Forward Brain Organization, to recognize understudies who might exit a course inferable from mediations (execution or terrible grades). The creator is preparing the above ml calculations utilizing datasets from Harvard and OULAD colleges. Since the dataset contains many missing qualities and superfluous highlights, the creator is utilizing the PCA include determination calculation to eliminate the unessential elements and select just applicable highlights.

```
id,code_module,code_presentation,id_assessment
0,AAA,2013J,1752,TMA,19.0,10.0,268,11391,18,0,
1,AAA,2013J,1753,TMA,54.0,20.0,268,11391,53,0,
2,AAA,2013J,1754,TMA,117.0,20.0,268,11391,115,
3,AAA,2013J,1755,TMA,166.0,20.0,268,11391,164,
4,AAA,2013J,1756,TMA,215.0,30.0,268,11391,212,
5,AAA,2013J,1752,TMA,19.0,10.0,268,28400,22,0,
6,AAA,2013J,1753,TMA,54.0,20.0,268,28400,52,0,
7,AAA,2013J,1754,TMA,117.0,20.0,268,28400,121,
```

Fig.3: Dataset

In the aforementioned dataset, we have all of the information described in the study, such as code module, gender, number of tries, exam outcome, and so on. The final result is displayed in the last section, and we can use it to determine if the student is WITHDRAWAL or NON-WITHDRAWAL. We will utilise the aforementioned dataset to train machine learning algorithms and develop a model, which will then be applied to the test dataset to predict whether a student will withdraw or not. We don't have a final result column in the test dataset below, therefore the machine learning model will predict it by evaluating test attributes.

Author identifies three distinct sorts of pupils from the aforementioned dataset

Amotivation: students who drop classes within one week after registering

Extrinsic: students who excel in the subject

Intrinsic: students who truly do inadequately in class

Following the disclosure of the previously mentioned attributes in the dataset, the creator is preparing all calculations and breaking down their presentation concerning ACCURACY, FSCORE, AUC, Sensitivity, and Specificity. How much right expectations from the whole test information attributes is alluded to as precision.

FSCORE: The F1 score is the amount of accuracy and review. Accuracy is characterized as the quantity of precise forecasts in each class, while review is characterized as the general number of right expectations from the gave information.

AUC: An illustration metric for order issues at various limit levels is the AUC-ROC bend. While ROC is a likelihood bend, AUC is the degree or proportion of distinguishableness. It demonstrates the model's ability to differentiate between classes. The model's ability to correctly predict 0s as 0s and 1s as 1s increases with the AUC. The model's ability to distinguish between patients with and without the disease is indicated visually by the higher the AUC.

Sensitivity and Specificity: relates to the classifier's performance.

4. ALGORITHMS

We are utilizing ML techniques like Random Forest, Multilayer Perceptron (MLP), Gradient Boosting, Generalised Linear Model, and Feed Forward Neural Network for this exploration to find understudies who might exit because of intercessions (execution or terrible grades).

4.1 Random Forest

A notable ML calculation from the directed learning approach is Random Forest. It has a good chance of being applied to ML issues involving arrangement and relapse. It is based on the concept of troupe realizing, which is a method that uses a few classifiers to solve a problem and make the model look better. As the name suggests, "Random Forest is a classifier that takes the normal and includes various choice trees on various subsets of the given dataset to improve the expected precision of that dataset." Instead than depending on a single option tree, the Random Forest forecasts the result based on the majority vote of expectations. Then, it aggregates each tree's theories. With more forest trees, there is less chance of overfitting and more accuracy.

The Random Forest method is seen in the picture below:

4.2 Multilayer perceptron

The multi-layer perceptron is the artificial neural network design with the most complexities. The majority of its components are various perceptron layers. TensorFlow is a famous deep learning system created by, and this journal will walk you through the method involved with making a brain network utilizing this library. To comprehend what a multi-layer perceptron is, we should make one without any preparation utilizing Numpy.

4.3 Gradient Boosting

A family of machine learning algorithms known as gradient boosting classifiers combine many weak learning models to pinpoint the main model strengths. In many circumstances, decision trees are used in angle support. Gradient boosting models are becoming more and more popular as a result of their ability to organise dense data.

4.4 Generalized Linear Model

The Generalized Linear Model, or GLM, is a cutting-edge technique for presenting data. The reaction variable y can have a wrong dispersion in addition to the usual circulation by using a trick expression for a few distinct models. In any case, GLM models enable us to establish a direct connection between the reaction and indicators even when the fundamental connection is not straight. This is achieved by utilizing a connection capability, which interfaces the reaction variable to a straight model. Not at all like linear regression models, the reaction variable's blunder dispersion doesn't need to be consistently conveyed. The errors in the reaction variable are supposed to be dispersed in a remarkable family (for example typical, binomial, Poisson, or gamma conveyances).

4.5 Feed forward neural network

When the connections between hubs don't form a circle, an artificial neural network is said to be a "feed forward neural network." A recurrent neural network is the antithesis of a feed forward neural network, which cycles explicitly. The feed forward model is the most fundamental form of brain organization because input is only handled in one way. The information may pass through a number of secret hubs, but it always moves forward and never backward.

5. EXPERIMENTAL RESULTS

The most common way of changing over raw data into a conceivable organization is known as data preparation. We can't manage crude information, subsequently this is a vital stage in information mining. Prior to utilizing ML or information mining techniques, the information quality ought to be assessed.

While making a prescient model, highlight choice is the method involved with limiting the quantity of info factors. It is desirable over limit the quantity of info factors to decrease demonstrating computational expenses and, in specific circumstances, increment model execution.

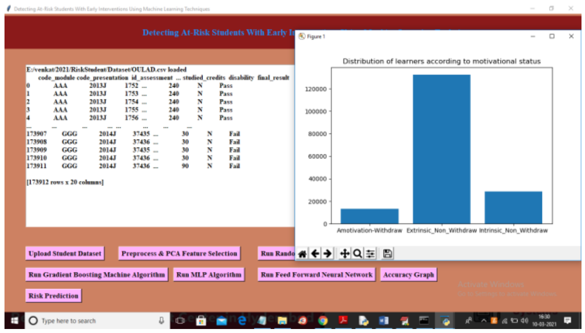

Fig.4: Uploading OULAD dataset

The dataset is loaded in the above screen, and we can see that there are many non-numeric values in the dataset, which the machine learning model will not accept, so we must process it to convert it to numeric format. In the above screen, we found three different users: AMOTIVATION, EXTRINSIC NON-WITHDRAW, and INTRINSIC NON-WITHDRAW. The x-axis in the graph above shows student type, while the y-axis reflects the number of students of that category.

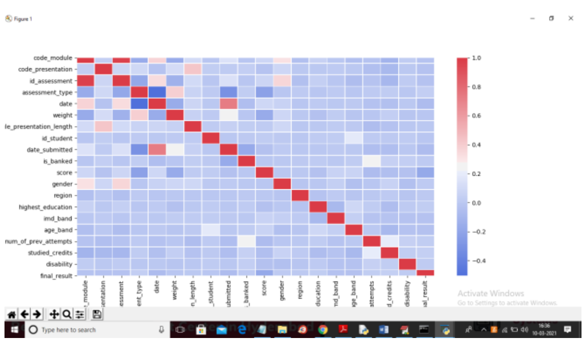

Fig.5: PCS feature selection

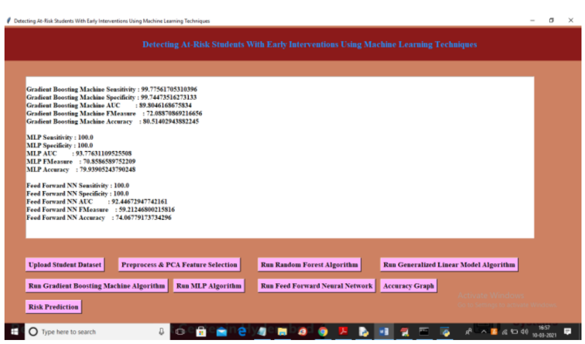

Fig.6: Algorithms result

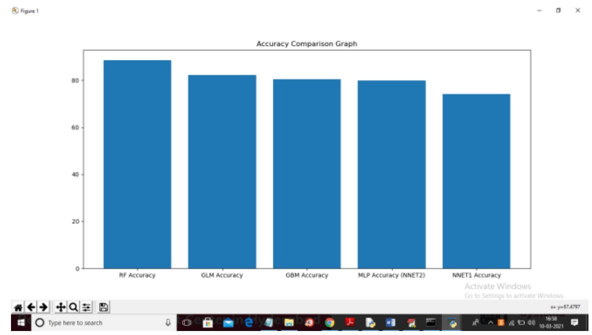

Fig.7: Accuracy comparison graph

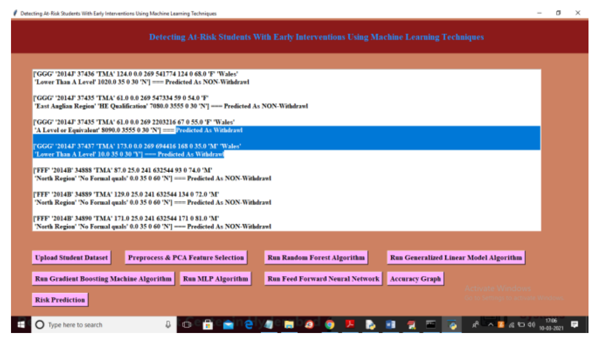

Fig.8: Risk prediction

Following dashed symbols, we can see the values of student test records and predicted results as WITHDRAW or NON-WITHDRAW. This ML model allows us to predict a student's likelihood of dropping a class simply by looking at their past performance characteristics.

6. CONCLUSION

As per the examination, understudy inspiration directions are the essential driver of understudy separation in web-based courses. Highlight choice further develops the expectation force of ML models while bringing down their registering costs. Additionally, the element choice channel approach is a suitable solution to the problem of overfitting. Teachers may be able to identify students who require additional assistance thanks to the findings of this study, which may make it easier for them to monitor changes in understudy arousal. Different factors affecting in danger understudies were broke down in the learning accomplishment model using the Harvard and OULAD datasets. The dataset discoveries show that snap stream properties are significant determinants that are considerably associated with understudy disappointment in web-based courses.

7. FUTURE SCOPE

Concerning the examination that will take place in the future, we must investigate whether or not additional datasets are required to approve the proposed structure. In order to examine subject patterns, it would be useful to collect online measurements from a few suppliers that offer seminars on comparable topics. Deep learning may be used to identify the pupils who are most likely to skip class. By deriving fleeting event groupings from numerous MOOC datasets, deep learning can distinguish attributes from understudy records.

REFERENCES

[1] M. R. Ghaznavi, A. Keikha, and N.-M. Yaghoubi, ‘‘The impact of information and communication technology (ICT) on educational improvement,’’ Int. Educ. Stud., vol. 4, no. 2, pp. 116–125, 2011.

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

[3] H. B. Shapiro, C. H. Lee, N. E. W. Roth, K. Li, M. ˙etinkaya-Rundel, and D. A. Canelas, ‘‘Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers,’’ Comput. Educ., vol. 110, pp. 35–50, Jul. 2017.

[4] J.-L. Hung, M. C. Wang, S. Wang, M. Abdelrasoul, Y. Li, and W. He, ‘‘Identifying at-risk students for early interventions—A time-series clustering approach,’’ IEEE Trans. Emerg. Topics Comput., vol. 5, no. 1, pp. 45–55, Jan./Mar. 2017.

[5] R. Alshabandar, A. Hussain, R. Keight, A. Laws, and T. Baker, ‘‘The application of Gaussian mixture models for the identification of at-risk learners in massive open online courses,’’ in Proc. IEEE Congr. Evol. Comput. (CEC), Jul. 2018, pp. 1–8.

[6] M. Barak, A. Watted, and H. Haick, ‘‘Motivation to learn in massive open online courses: Examining aspects of language and social engagement,’’ Comput. Edu., vol. 94, pp. 49–60, Mar. 2016.

[7] J. C. Turner and H. Patrick, ‘‘How does motivation develop and why does it change? Reframing motivation research,’’ Educ. Psycholog., vol. 43, no. 3, pp. 119–131, 2008.

[8] C. Geigle and C. Zhai, ‘‘Modeling MOOC student behavior with twolayer hidden Markov models,’’ in Proc. 4th ACM Conf. Learn. Scale, 2017, pp. 205–208.

[9] Altair. (2019). Improve Retail Store Performance Through In-Store Analytics. [Online]. Available: https://www.datawatch.com/in-action/usecases/retail-in-store-analytics/

[10] D. S. Chaplot, E. Rhim, and J. Kim, ‘‘Predicting student attrition in MOOCs using sentiment analysis and neural networks,’’ in Proc. 17th Int. Conf. Artif. Intell. Educ., 2015, pp. 7–12.