Journal of Advanced Zoology
ISSN: 0253-7214
Volume 45 Issue 2 Year 2024 Page 481-491

Recognizing Students At-Danger With Early Intervention Using Machine Learning Techniques

Kotari Sridevi[1]*, AC Priya Ranjani[2], Syed Shabeer Ahmad[3]

*[1]*Associate Professor, Department of CSE, Muffakham Jah College of Engg & Tech, Hyderabad*
[2]Assistant Professor, Department of Computer Science and Applications, KLU, Vijayawada.
[3]Professor, Department of CSE, Muffakham Jah College of Engg & Tech, Hyderabad

***Corresponding Author:** Kotari Sridevi*
**Associate Professor, Department of CSE, Muffakham Jah College of Engg & Tech, Hyderabad*

Abstract

Students in online courses require attention as there is no much interaction between the teacher and student compared to traditional instructing methods. Due to the increase in advent of massive open online courses, there is a need to focus on identifying students at danger of withdrawl or failure. As the count of students enrolling in an online course is huge it's quite difficult to find out specific students who are at-danger of failure/withdrawal from the course. There is a need to alleviate this problem by identifying those students and help academic instructors offer support to them. The major contribution of this work is to analyze the risk associated with the dropout of student in order to aid instructors in delivering the intensive intervention support to student who is at verge of quitting from the course. The main objective is to track student performance and provide valuable information to the educator to subsequent the courses according to their learning achievement and also help academic advisors to detect the student having low academic achievement records and encourage the candidates. Data collected from OULAD datset is analyzed with the help at -risk prediction model is to identify whether a student is at verge of withdrawal or not.

Keywords: MOOCs, Dropout Prediction, Particle Swarm Optimization, Convolution Neural network

1. INTRODUCTION

ICT(Information and communication technology) is now widely used and plays an important role in education. ICT has helped to promote academic curricular and has made it possible to create a virtual classroom. ICT has the potential to improve student results by allowing teachers to assist students in completing tasks. As a result, high-quality instruction might be offered via virtual learning.
MOOCs (Massive Open Online Courses) use a wide range of multimedia techniques to create an interactive learning environment. MOOCs provide students with great digital learning resources by allowing them to access material from around the world. A number of top-ranked universities have adopted online courses as an alternative to traditional learning as a result of the collapse of financial and geographical barriers

connected with the traditional teaching methodologies. Low completion rates are a serious issue with MOOCs, given the fast proliferation of online courses in higher education [7]. One of the ways for increasing completion rates is to identify at-risk learners. Detecting at-danger students early on could aid instructors in providing educational interventions and improvising the course structure [8]. Instructors can provide real-time feedback to students with a fast intervention solution, and retention rates could improve [9]. Researchers looked at the reasons for course abandonment in order to develop an accurate at-risk student prediction model. Lack of motivation is the most common reason for students dropping out of online courses [10]. Students' motivation levels in online courses are said to drop or rise depending on the social, cognitive, and environmental aspects [11]. The motivational trajectory of a student is a key determinant of their likelihood of dropping out. Changes in student behaviour over courses can be used to measure motivational trajectories [11]. The case study presents a novel dropout predicting model that can provide at-risk students with timely intervention support. Machine learning is used to analyse student historical behaviour and detect potential patterns of learner attrition from course activities. Student engagement, as well as motivational status in previous courses, continuing to participate in the current course was also investigated.
In this research work, we collect the data from OULAD dataset and perform various operations in order to remove noisy data and then the features are aggregated and multiple machine learning classifiers are applied in order to classify the data. Performance measures specify the efficiency of each classifier used. Machine learning is used to track the student performance and provide useful information to educators so that they can continue with their courses based on their learning. It could also assist academic advisers in detecting students with low academic achievement and providing support to them. Also, from the results obtained Particle Swarm Optimization outperforms other conventional classifiers.

2. Related Work

[1]Ragha Al-Shabandari,Abir Jaafar Hussain,Panos Liatsis,Robert Keight, "Detecting At-Risk Students With Early Intervention Using Machine Learning",October 2019. Work focused on early detection of students who are at risk of withdrawl or failure was provided. Two models where constructed namely at-risk student model and the learning achievement model. Harvard and Oulad(Open University Learning Analytics Dataset) datasets where considered for data analysis. A total of eight courses were taken into consideration out of which four belonged to Harvard dataset and four of Oulad dataset.This models had the potential to detect the students who are in danger of failing and withdrew at the primary stage of the online course.
[2] Luis M.Romero-Rodriguez,Maria Soledad Ramirez-Montoya, and Jaime Ricardo Valenzuela Gonzalez,"Gamification in MOOCs: Engagement Application Test in Energy Sustainability Courses" ,March 2019.A gamification board with challenges, badges, and leaderboards was used, and at the same time, this platform was analyzed using the integrated theoretical gamification model in e-learning environment. The courses are completed by high school students, they may be find it to be too difficult, while for users with engineering degrees ,they may be very basic.

[3]W.Xing and Ddu, "Dropout prediction in MOOCs for personalized intervention,". This research work aimed to optimize the dropout prediction models focusing on personalizing and prioritizing intervention for academically at-risk students in MOOC's considering data of course held by Canvas in August 2014 which lasted roughly for 8 weeks .The course had 11 modules and 14 discussion forums and multiple MCQS. It relied on a weekly temporal prediction mechanism, this proposed using a deep learning algorithm to build dropout models and also produce individual student dropout probabilities for intervention and substantiation. By using deep learning, this approach not only built more accurate dropout prediction models when compared to baseline models but also introduced a valid approach to inform intervention design thereby personalizing and prioritizing support for at-risk students using MOOC dropout probabilities[3].
[4] O.Zughoul, F.Momani,O.H.Almasri,A.A.Zaidan, B.B.Zaidan ,M.A.Alsalem,O.S.Albahri,and M.Hashim, "Comprehensive insights into the criteria of student performance in various educational domains," . An in-depth insight was considered on surveying the literature on criteria of student performance in different educational domains so that it can figure out the gap on this study. This search for articles focused on the (a) evaluation of student (b) education-related and (c) criterion and domain.
[5] J.L.Hung,M.C.Wanf,S.Wang,M.Abdelrasul,Y.Li, and W.He, "Ídentifying at-risk students for early interventions A time-series clustering approach", In time-series clustering approach,the data was analyzed at regular intervals depending on holiday effect i.e., it focused on detecting student behaviour before and after a long holiday break.

[6] R.Al-Shabandar, A.Hussain, A.Laws, R.Keight, J.Lunn,N.Radi, “Machine learning approaches to predict learning outcomes in massive open online courses”. The common feature selection techniques are employed.Behavioral features where taken in conjunction with demographic features.Linear and non-linear classifiers were used for classification of data in dataset.The dataset was divided into two sets where first set consistedof all the features whereas in second set only highly ranked features were used.
[7] H.B.Sharpio,C.H.Lee,N.E.W.Roth,K.Li,M.Cetinkaya-Rundel,andD.A.Canelas, “Understanding massive open online course(MOOC) student experience :An examination of attitudes,motivations,and barriers The main aim was to understand the impact of instructional design on quantitative outcomes.This analyzed the text of interview transcripts to gain deeper understaning of motivation for and barriers to course engagements for students participating in MOOC’s.Both internal and external factors of course setting which impacted engagement and learning were considered.36 participants were interviewed who varied in age,belonged to different locations.80% of the interview statements were found to be positive.In addition,when demographic features were taken into consideration interviewe statements having bachelor’s degree were found to be more postive.Lack of time was one of the most common barrier faced by almost all the students,others were previous bad experience,inadequate background knowledge,linguistic competence and communication skills.

[8] J.Sinclair and S.Kalvala, “Student engagement in massive open online courses. Completion rates in massive open online courses were disturbing low.Analysis mainly focused on the patterns of resource access and prediction of dropout using learning analytics[8]. Massive Open Online Courses have experienced many enrollments with lower certification rates and highly educated registrants from all over the world[9]. 68 open online courses offered on edX by Haravard and MIT were taken into consideration.

3. SYSTEM DESIGN AND IMPLEMENATION

3.1 Dataset Collection

A dataset is collected from Open University in UK, a dashboard was released known as Open University Learning Analytics Dataset(OULAD) Demographic, behavioural and temporal features are captured in dataset. This dataset consists data about different courses,students enrolled and their interactions with VLE(Virtual Leaning Environment) for four selected courses which are known as modules. Electronics Fall, Circuits and Electronics Spring,Social Science Fall and Social Science Spring.These courses are represented with the help of alphabets “A” stands for course held in January,”B” for February and so on . VLE (Virtual Learning Environment) data was collected on a day to day basis and features were extracted. The VLE features extracted rely on clickstream data. This OULAD dataset consists of eleven VLE activity types.Figure provides an overview of the OULAD dataset.

Table: 3.1: Data Set Overview

Features	Description
Student_id	Learner identity number
Age_band	Age of learner
Gender	Learners gender
Highest_education	Education level of learner
Region	Lerner’s geographic area
Studied_credits	No of credits for the module in which learner is involved
Disability	Indicator of student disability
Num_of_prev_attempt	No of times that student undertook the course
Imd_band	Socio-economic indictor measure of student economic level
Learning_activity	The type and number of daily activities that the student undertakes
Grades	The students assessment marks
Date_registration	Date of learner registration in the course
Date_unregistration	Date that learner quit the course

3.2 **Architecture**

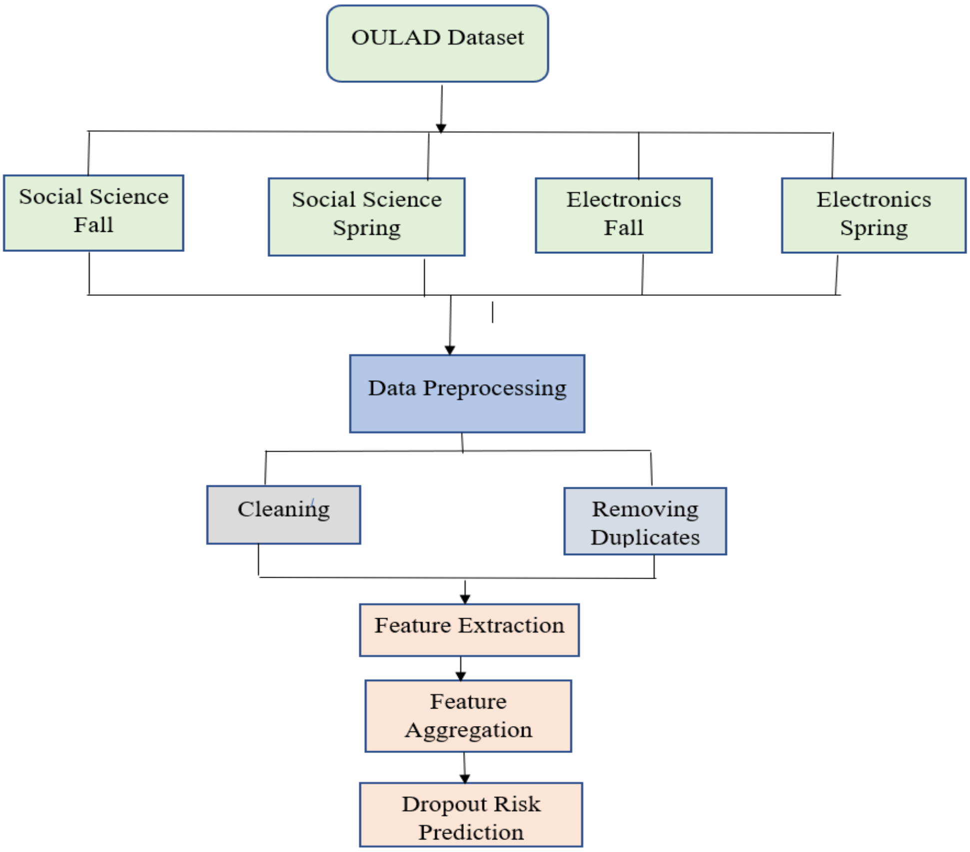

Fig 3.1: System Architecture

3.3 At-Risk Student Framework

In previous work[17],Learning Analytics where used to categorize students motivational status as amotivation,extrinsic and intrinsic.An algorithm is proposed to detect at risk students with respect to the course trajectories comcept.The intervals defined in this algorithm are T1,T2.Learners engaging in fall course are selected in T1 whereas learners of both falls and spring courses are considered in T2.Low student performance and learning achievement outcomes are considered to be importan factors of students withdrawl from online courses.While investigating the critical factors a data driven approach should be considered which affect the student learning outcomes.To detect these factors,student learning achievement model is proposed.

Let R*i* €V represent the ith student record,given as:

R_i=<s_i,g_i,d_i,e_i,c_i >

Where

s_i.Identity of student for the ith record
g_i -Grade of the ith student record
d_i-Start date of student interaction with course
e_i-End date of student interaction with course
c_i.Identity of course associated with ith entry

3.4 Learning Achievement Framework

Learning achievement is regarded as a key indicator of the MOOCs platform's performance. To forecast whether students will pass or fail an online course, a student performance predictive model is developed. The framework's goal is to assess bad student performance and look into the impact of learning activities on students' decisions to continue with a course in the future which will assist instructors in drawing inferences about performance of the student and give us deeper insights into the learning process.Factors are examined using linear algorithm as it gives in-depth insight nto truth bhind learners succes or failure in MOOC's platform.The Virtual Learning Environment activities are used to construct behavorial features from OULAD dataset. In terms of temporal characteristics, the number of days that students interact with OULAD online courses is calculated by subtracting the dates of student enrollment and deregistration from MOOCs.This results in weak association between learning outcomes and demographic features hence,demographic features are excluded in the analysis.

3.5 Data Pocessing

Raw data are full of noise, misspellings, and contain numerous abbreviations. Such noisy characteristics often involve the performance of dropout prediction analysis approaches. Thus, some preprocessing approaches are applied prior to feature extraction. The pre processing of OULAD dataset include the following steps:

Cleaning

Cleaning is done to remove the values that are of no use and creates a mess while analyzing using a model as it consists of missing values, out-of-range values, etc. Hence, this step is needed. Removing the student records with duplicated rows.

Normalization

To deal with diverse units normalization is done. It is the process of rescaling the attributes to the range of 0 or 1.Hence, Normalization of the raw data is done.

Feature Selection

After data preprocessing we proceed ahead with the feature selection. Since the existence of redundant features could affect results as it makes the model learn based on irrelevant features as well as various combinations of features would have different results this step is needed. Feature selection is used to select those features automatically or manually that contributes towards the prediction or output in which one is interested. Principle Component Analysis is being used for feature selection in our study. PCA is generally used to reduce the number of variables of a large number of interrelated variables or to reduce the dimensionality while retaining as much of the information as possible. It selects a subset of variables from a large set of variables such that the subset of variables explains most of the variability of the dependent variable in a multiple regression problem. The main objective of PCA is to reduce the predictor variables and to detect the structure in the statistical relations that might exist between the variables. In this, the linear combinations of p initial variables ($y1, y2, \ldots\ldots, yn$) are created to produce principle components ($PC1, PC2, \ldots\ldots, PCn$) . Each of this principle component is expressed by the equation below

$$PC_1 = w_{11}y_1+w_{12}y_2+\ldots\ldots w_{1n}y_n$$
$$PC_2 = w_{21}y_1+w_{22}y_2+\ldots\ldots \quad w_{2n}y_n$$
$$PC_n = w_{n1}y_1+w_{n2}y_2+\ldots\ldots w_{nn}y_n \qquad \text{eq.(3.1)}$$

In the above equation,

PC_i= Principal Component;

w_{ij} = coefficient of the principal component and the initial variable; and

y_i = Initial variable.

The coefficient is estimated in such a way that the first principal component (PC1) measures the largest possible variance, and the second principal component (PC2) measures the second largest possible variance not accounted for by the first principle component. The PCA process is continued until the last principle component (PC_n) completes the entire variance. Once the data is prepared, key features from different dimensions are identified by primary correlation analysis and then training and testing datasets are generated. Different modeling techniques can be used to process the training and testing data sets. For our analysis we have spilt the data into 80% for training set and 20% for testing set for each indicator.

Risk evaluation

Risk Evaluation is by calculating Mean Absolute error,

Mean Absolute Error (MAE)

The mean absolute error is one of the ways to measure the accuracy of the model which is being calculated for each model implemented is the system. It is calculated as the difference between actual outcome and predicted outcome.

The equation of calculating MAE is as given below

$$MAE = \sum |yi\text{-}xi|/n \qquad \text{(eq.3.2)}$$

where
y_i is the predicted value
x_i is the actual value
n is the total no of observations

Risk prediction
After risk modeling, and risk evaluation the last step is the risk prediction. The output showing the water quality risk is predicted by taking into consideration all the parameters selected while feature section. The models are being applied to predict the risk and the results are generated displaying whether the risk is predicted or not.

Algorithm 1 displays the overall procedure used in the preprocessing of the data set in this study.

3.5.1 Algorthim Preprocessing
Begin
Input dataset
For data in dataset, Do:
Procedure Pre-processing:
Remove missing values
Remove duplicated rows
Remove near-zero variance features
Apply data-transformation technique
Return preprocessed data
End Procedure
Procedure Dropout Risk Classification:
Classify data using machine learning
techniques(RF,GLM,PSO,GBM.MLP,FFNN)
End Procedure
End

3.6 Algorithms
3.6.1 At-Risk Students Agorithm
1. BEGIN
2:Ci -Set of courses
3:Sv-Set of students who enrol in course
4:Let gi be the score obtained, Pi be the ith student record
5. ∀Pi € Sv :

if gi ≤40 and and ei <8
Then
Pi = “Withdrawl Student”
Else
Pi = “Non-Withdrawl Student”

3.6. At-Risk Student Prediction Algorithm

Algorithm 2 At-Risk Student Prediction Algorithm

Input: S is a set of n samples
Max-iteration is the maximum number of iterations
Output: Let PM is set of performance metrics
1: **for** i = 1 … Max-iteration, **do**
2: **for** j = 1 … n, **do**
3: Calculate the feature weights
end for
4: Let ML is set of machine learning models where
ML={ GLM,GBM,RF,FFNN,MLP,PSO}
5: Let PM to be a set of performance metrics where
PM = { Accuracy, F1 Score, Sensitivity, Specificity, AUC }
6: Training = { tr € S→tr S}
7: Test = { ts € S→ts S & ts € Training}
8: **for** all ML **do**
9: Compute Perfomance Metrics
10: E[PM] = { S: S→ML(Training, Test)}
End for

4. EXPERIMENT RESULTS

The result analysis is done considering accuracy, Mean Absolute Error (MAE), F-measure, Sensitivity, Specificity,AUC..

Accuracy:
The accuracy is calculated for each algorithm and the comparison graph for accuracy is as shown in the Figure 5.1 for dataset 1where one can see that PSO,CNN has gained the highest accuracy compared to Random Forest,GLM,GBM,MLP. The accuracy of each algorithm is given in the Table 5.1 for dataset.

Table 5.1 Accuracy values of dataset

Algorithm	**Accuracy**
PSO	98.88
GLM	87.04
Random Forest	81.56
MLP	80.0
CNN	98.0
GBM	88.56

The comparison graph of accuracy is shown in the Figure 5.1 where the x-axis represents the names of the algorithms and the y-axis represents the accuracy score.

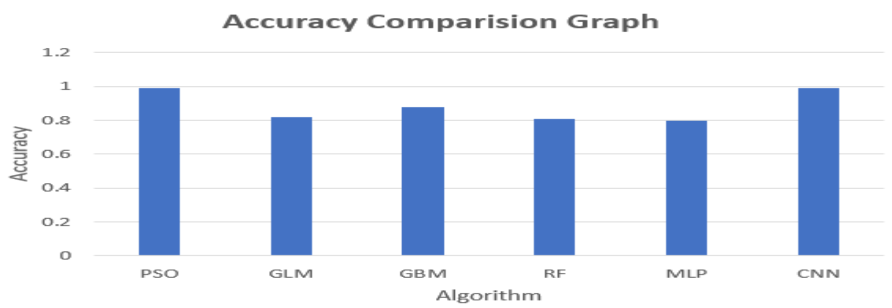

Figure 5.1: Comparative analysis of accuracy for dataset

Mean Absolute Error (MAE):
It is a measure that has been used to assess the error rate of the algorithms. It is calculated differences between the actual outcome and the predicted outcome. MAE is calculated for each algorithm and a comparison graph

is drawn. The lower the MAE the better the algorithm. The equation of calculating MAE is as given in eq. (5.1)

$$\text{MAE}= \sum \frac{|y_i - x_i|}{N} \qquad \text{eq. (5.1)}$$

Where,
y_i is the predicted value,
x_i is the actual value, and
N is the total number of observations.

Table 5.2 MAE of Algorithms

Algorithm	**MAE**
PSO	0.007
GLM	0.242
Random Forest	0.270
MLP	0.270
CNN	0.0057
GBM	0.270

From the above table we observe that the random Forest algorithm has obtained the highest mean absolute error and PSO,CNN has obtained the lowest mean absolute error compared to other algorithms. The comparison graph of MAE is as shown in the figure 5.2 where the x-axis represents the algorithms and y-axis represents the MAE error.

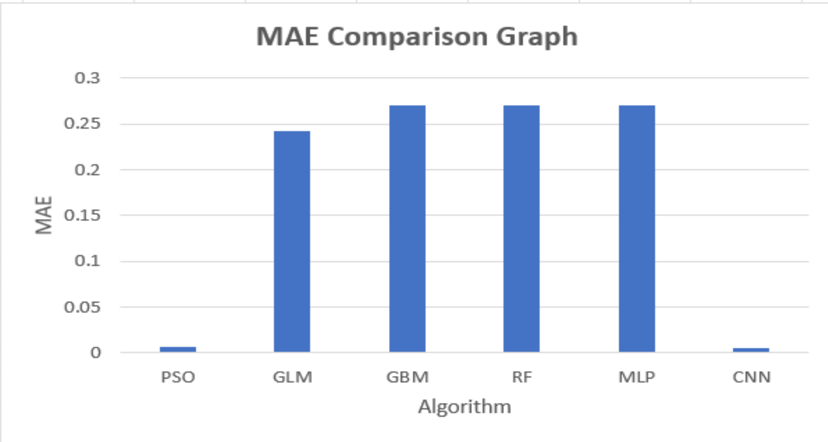

Figure 5.2 MAE Comparison graphs

F-Measure:

This is calculated using precision and recall. The F-measure is calculated using the eq. (5.2). It is calculated for each algorithm and the results are listed in the Table 5.3 for dataset .

$$\text{F-measure}= 2*\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \qquad \text{e.q (5.2)}$$

Table 5.3 F-measure values of dataset

Algorithm	**F-Measure**
PSO	0.99
GLM	0.82
Random Forest	0.79
MLP	0.80
CNN	0.98
GBM	0.80

AUC:
The AUC is calculated for each algorithm and the results are listed in the Table 5.4 for dataset.

Table 5.4 AUC values of dataset

Algorithm	AUC
PSO	0.99
GLM	0.94
Random Forest	0.89
MLP	0.93
CNN	0.98
GBM	0.89

Specificity:
The specificity is calculated for each algorithm and the resulted values are listed in the Table 5.5 for dataset .The equation for calculating specificity is given in eqn 5.3 .The higher the specificity value the better the algorithm.

$$\text{Specificity} = \frac{\text{True negative}}{\text{True Negative+False Positive}} \quad \text{eq. (5.3)}$$

Table 5.5 : Specificity values of dataset

Algorithm	F-Measure
PSO	0.99
GLM	1
Random Forest	0.99
MLP	1
CNN	1
GBM	1

Sensitivity:
The sensitivity is calculated for each algorithm and the results are listed in the Table 5.6 for dataset. The equation for calculating sensitivity is given in eqn 5.4

$$\text{Sensitivity} = \frac{\text{True positive}}{\text{True positive+False Negative}} \quad \text{eq. (5.4)}$$

Table 5.6 Sensitivity values of dataset

Algorithm	F-Measure
PSO	0.99
GLM	1
Random Forest	0.99
MLP	1
CNN	1
GBM	1

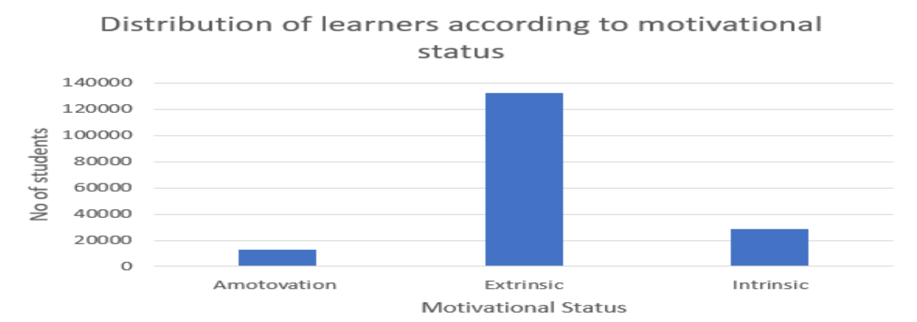

Figure 5.3 Distribution of learners based on motivational status
Above figure clearly depicts the extrinsic and intrinsic non withdraw students and motivated students.

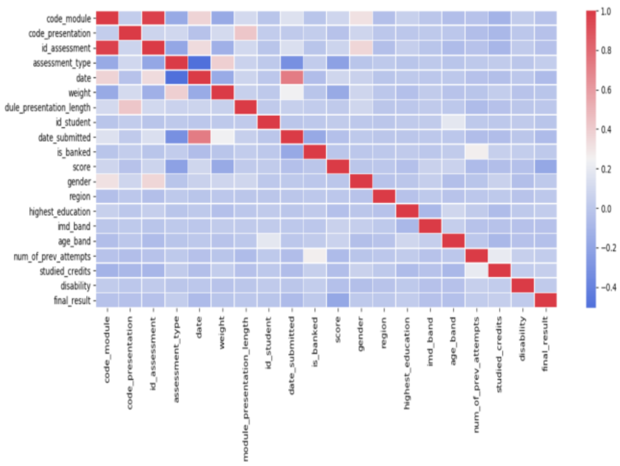

Figure 5.4 Heat Map

Heat Map is 2D graphical representation of data where the individual values contained in matrix are represented as colors.It gives a 2D correlation matrix between two discrete dimensions using coloured cells. Results show that PSO obtained highest Fmeasure value and random forest obtained the lowest Fmeasure value.The learning achievement model revealed nearly ideal sensitivities and specificties for all classifiers. Classification performance for learning achievement model is shown below:

Table 5.7 Performance Classification

Classifier	Accuracy	F-Measure	Sensitivity	Specificity	AUC	MAE
GLM	0.82	0.82	1	1	0.94	0.242
RF	0.81	0.79	0.99	0.99	0.89	0.270
MLP	0.80	0.80	1	1	0.93	0.270
PSO	0.99	0.99	0.99	0.99	0.99	0.007
CNN	0.98	0.98	1	1	0.98	0.0057
GBM	0.88	0.79	1	1	0.89	0.270

The table shows the overall comparison between the performance measures for various machine learning classifiers. This concludes that the PSO algorithm has better overall performance when compared to other algorithms giving an acccuracy of about 0.99.

In addition, PSO (particle swarm optimization) features selection algorithm which is an advance features selection algorithm is utilized. PSO will analyse all dataset columns and then apply linear regression classifier to estimate importance of each column or attribute and the attribute which contribute more in getting high accuracy will be selected and the attribute which is contribute less will be removed out and due to this selection of important features it helps PSO in obtaining high accuracy compared to all other machine learning algorithms.

5. CONCLUSION & FUTURE WORK

In the context of this work, an approach that aims to extract Dropout Prediction analysis was done by pre-processing, feature extraction and utilizing transformation technique to address non-normally distributed data using machine learning classifiers. Classifiers, such as Linear Model, Gradient Boosting machine, Convolutional Neural network, Particle Swarming Optimization, and Random Forest, where used. This approach was optimized using OULAD data set that is publicly available.

Experimental results indicate that Generalized Linear Model and Gradient Boosting Machine have an almost similar accuracy. However, the Particle Swarm Optimization, Convolution Neural network gives the highest accuracy at 98.81%. Experimental results concluded that the proposed model can detect students at risk of withdrawal or failure using machine learning methods with good accuracy result. In regards to future research, we intend to consider the validation of the proposed framework with additional datasets. We can improvise our approach by attempting to use bigrams and trigrams. Furthermore, we intend to investigate different machine learning techniques and deep learning techniques, such as Deep Neural Networks, and Recurrent Neural Networks.

6. REFERENCES

1. Ragha Al-Shabandari,Abir Jaafar Hussain,Panos Liatsis,Robert Keight, “Detecting At-Risk Students With Early Intervention Using Machine Learning”,October 2019.
2. Luis M.Romero-Rodriguez,Maria Soledad Ramirez-Montoya, and Jaime Ricardo Valenzuela Gonzalez,“Gamification in MOOCs: Engagement Application Test in Energy Sustainability Courses” ,March 2019.
3. W.Xing and Ddu, “Dropout prediction in MOOCs for personalized intervention,” J. Educ. Comput. Res., vol. 57, no. 3,pp. 547570, 2018.
4. O.Zughoul,F.Momani,O.H.Almasri,A.A.Zaidan,B.B.Zaidan ,M.A.Alsalem,O.S.Albahri,and M.Hashim, “Comprehensive insights into the criteria of student performance in various educational domains,” IEEE Access, vol. 6, pp. 73245-73264, 2018.
5. J.L.Hung,M.C.Wanf,S.Wang,M.Abdelrasul,Y.Li, and W.He, “Ídentifying at-risk students for early interventions A time-series clustering approach”, IEEE Trans. Emerg. Topics Comput., vol. 5, no. 1,pp. 4555, Jan./Mar. 2017.
6. R.Al-Shabandar,A.Hussain,A.Laws,R.Keight,J.Lunn,N.Radi, “Machine learning approaches to predict learning outcomes in massive open online courses”, in Proc. Int. Joint Conf. Neural Netw. (IJCNN),May 2017, pp. 713-720.
7. H.B.Sharpio,C.H.Lee,N.E.W.Roth,K.Li,M.Cetinkaya-Rundel,andD.A.Canelas, “Understanding massive open online course(MOOC) student experience :An examination of attitudes,motivations,and barriers,”Comput. Educ., vol. 110, pp. 35-50, Jul. 2017.
8. J.Sinclair and S.Kalvala, “Student engagement in massive open online courses,”. J. Learn. Technol., vol. 11, no. 3, pp. 218-237,2016.
9. E.Summary, “Haravard and MITx:Two years of open online courses”, Electron. J., no. 10, pp. 1-37, 2015.
10.]Uma .N.Dulhare, Udai Kumar Kudikyala, “Using Scrum and Wikis to Manage Student Major Projects”,2015.
11. J.Reich,S.Nesterko,D.Seaton,T.Mullaney,J.Waldo,I.Chuang, and A.D.Ho, “PH207x:Health in numbers &PH278x:Human health and global environmental change, ”USA, Tech. Rep., 2014.
12. Dr. K.Udai Kumar,Dr. Uma N. Dulhare, “Enhancing Engineering Education in an Indian Education System: WebBlogs”,2014.
13. 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.
14. M.Barak,A.Watted,and H.Haick, “Motivation to learn massive open online courses:Examining aspects of language and social engagement,” Edu., vol. 94, pp. 49-60, Mar. 2016.
15. C.Geigle and C.Zhai, “Modeling MOOC student behavior with two layer hidden Markov models”in Proc. 4th ACM Conf. Learn. Scale, 2017,pp. 205-208.
16. Altair , “Improve Retail Store Performance Through In-StoreAnalytics”,2019. Available: https://www.datawatch.com/in-action/usecases/retail-in-store-analytics/
17. 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.
18. J.He,J.Bailey,B.I.P.Rubinstein, and R.Zhang, “Identifying at-risk students in massive open online courses,” in Proc. 29th AAAI Conf. Artif.Intell., 2015, pp. 1749-1755.
19. B.Minaei-Bidgoli,D.A.Kashy,G.Kortemeyer, and W.F.Punch, “Predicting student performance:An application of data mining methods with an educational web-based system,” in Proc. 33rd Annu. Frontiers Educ.(FIE), vol. 1, Dec. 2003, pp. T2A13T2A18.
20. A.Ho,J.Reich,S.Nesterko,D.Seaton,T.Mullaney,J.Waldo,and I.Chaung, “Harvardx and MITx:The first year of open online courses,fall 2012-summer 2013,” SSRN Electron. J., no. 1, pp. 1-33, 2014.
21. K.Kuzilek,M.Hlosta,D.Herrmanniva,Z.Zdrahal,and A.Wolff, “OU analyse:Analysing at-risk students at the open university,” inProc. 5th Int. Learn. Anal. Knowl. (LAK), 2015, pp. 1-16.
22. J.W.D.Seaton,J.Reich,S.Nesterko,and T.Mullaney, “Introduction to Computer Science and Programming MITx on edX-2012 Fall,” York, NY, USA, 2014.
23. P.F.Mitros,K.K.Afridi,G.J.Sussman,C.J.Terman,J.K.White,L.Fischer, and A.Agarwal, “Teaching electronic circuits online:Lessons from MITx’s 6.002x on edX,” in Proc. IEEE Int. Symp. Circuits Syst.,May 2013, pp. 2763-2766.