Journal of Theoretical and Applied Information Technology
15th February 2023. Vol.101. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# AN IDENTIFICATION OF THE PROMINENT LEARNER BEHAVIORAL FEATURES TO PREDICT MOOC DROPOUTS USING HYBRID ALGORITHM

**S. NITHYA[1] and Dr S.UMARANI[2]**

[1] Research Scholar, Department of Computer Science, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.
[2] Professor, Department of Computer Science, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.

E-mail: [1]nithyaresearch19@gmail.com, [2]ravania@gmail.com

## ABSTRACT

Knowledge is derived throughout the world, which has launched plenty of online courses in response to the crisis. A growing number of learners are ready to enroll in online programs that help them adapt to this dynamic place. However, in some circumstances, the learner's level of retention rate should be reduced. The findings of a recent systematic review add to the many patterns that have been found in substantial research on learner digital interaction. This work used a set of distinctive characteristics derived from real-time datasets to measure learner dropout. The learning analytics framework is frequently used to examine user contributions and performance. Additionally, it supports higher education students' decision-making. To identify a learner's dropout, many researchers apply a variety of techniques, such as machine learning approaches and statistical methods. In this work, we propose Machine Learning-Artificial Intelligence (ML-AI), a novel hybrid approach that combines Random Forest (RF) and Artificial Neural Network (ANN) that requires minimal recurrent training to overcome these issues. The random forest is made up of a number of decision trees that can be used to classify data. It also assigns higher weights to the specified features in order to improve their classification capabilities. Then optimize by connecting the random forest to an ANN based on the feature score to obtain consistent prediction performance. Then we compared the metric values with the hybrid approach and some baseline models like Logistic Regression (LR) and Support Vector Machine (SVM). Our attempts are focused exclusively on diminishing early dropout rates and raising learner retention.

**Keywords:** *Dropout Prediction, Educational Data mining, Performance Evaluation, Artificial Neural Network, Multilayer Perceptron, Machine Learning*

## 1. INTRODUCTION

The Massive Open Online Course (MOOC) is the latest advancement in quality education and virtual education[1] that uses the internet and numerous technological breakthroughs to provide free exposure and cognitive development to people all over the world, including job advancement, career change, skills and knowledge, activation secondary learning, continuous learning, professional E-Learning, etc.,. They provide a unique experience by combining course materials (videos, documents, etc.) with instructive exercises (quizzes, forums, etc.).

The courses are open to anyone with an internet connection across the world. MOOC courses on a variety of topics and expertise have already been launched by a number of lecturers[2]. The online learning courses are mostly used pattern by the educational society and update career requirements for the professionals. At the present, a lot of the platforms organize online learning programs like Coursera, edx, Udacity, and Great learning, etc.

The benefits of online learning were used as a supplement to the video[3], which looked into students' perspectives on MOOCs. Furthermore, the evidence suggests that increasing students' added benefits will raise their interest in learning with MOOCs when deciding between learning priority and learners' motivation. Many MOOC systems provide varied outcomes [4]in terms of time, intended learning, examination, community interactions and educator presence.

In a MOOC, the prediction can focus on a variety of goals, such as learning outcomes and characteristics preferred and mode of the prediction as represented in Figure 1.

The learning outcome indicates the completion of the course, as well as that the student will be able to predict a grade, a certificate, and a dropout. Grade prediction establishes that learners' score depends on the progress in their study (i.e. .estimate attempt to complete the task on Correction Fist Attempt (CFA)). Few of the learners could not attend the exam in proper timings; therefore they fail to attain Certificate. The most common query is how to track down course dropouts'. As a result, the learners' retention rate is exceptionally low when compared to the registered users, and our work focused on the learning outcome's dropout factors.

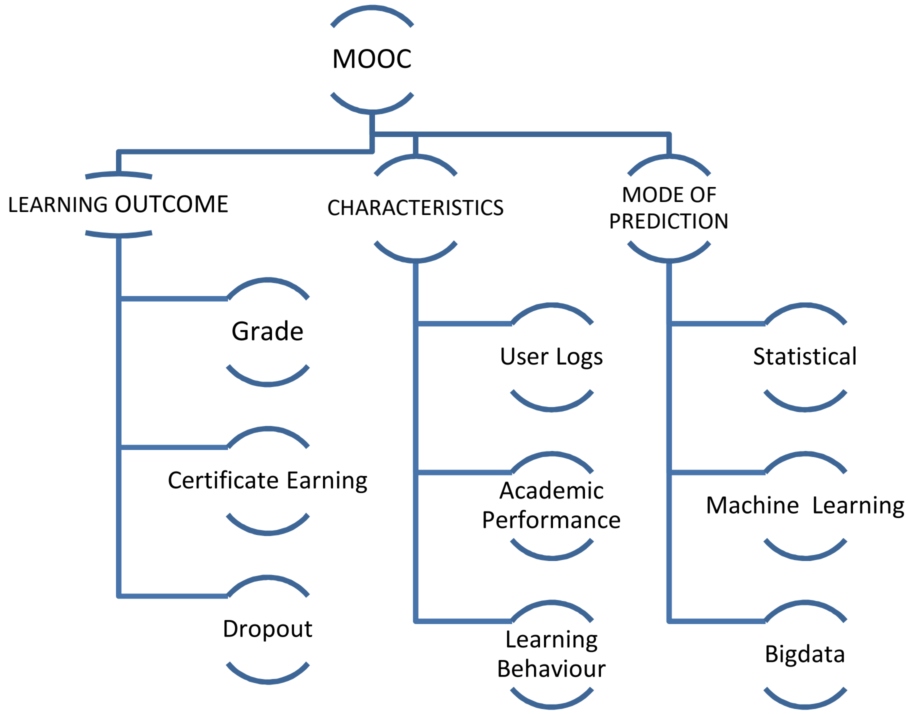


*Figure 1. MOOC Analysis*

The parameters for the evaluation were chosen since the user log contains basic information about the learners, and their academic performance includes their past course and assignment scores. Based on their activities, the learner procures data from many resources, such as reading documents and viewing videos, to predict the learners' behavior in the course.

Multiple methodologies are utilized to evaluate the metrics, while statistical methods are employed to analyze the survey questions, construct learning models using various machine learning approaches, and extend the volume of data using big data.

MOOCs are popular because of their potential for unlimited enrollment, freedom from geographical limits, and unrestricted access to the bulk of programs, and structural similarity to conventional learning. More than 10,000 MOOCs were available in 2020, with over 200 million students enrolled. Moreover, owing to the increased student-to-instructor ratio in a virtual learning environment. In such environments, teachers are unable to track students' learning behavior, leading to rejection or retention. Effective dropout prediction will assist various education systems[5]. This will aid in the improvement of curriculum planning, materials, and instruction.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The growths of Self- Learning [6], most of the learners are interested to register for the courses. At the same time, because of the availability of digital information and other resources, some students are less interested in finishing their education, and the dropout rate is rising. Dropout is a major problem for both the provider and the learner. Most of the researchers' focus is on predicting the dropout rate of the learners using various machine learning techniques.
The following are the sections of the article: Section. 2 reviews the background details; Section. 3 describe the methodology; Section. 4analyzes the results of the experiments; and the conclusion & future scopes for research are discussed in Section. 5.

## 2. BACKGROUND

The numerous learner logs gathered on MOOCs; including assignment submission and activity measures, learner demographic information, curriculum forum entries, interactive information, clickstream data, and others, reflect the various learning patterns of each participant. Different data mining techniques are applied to educational statistics to estimate learning outcomes, assess slow learners, and predict dropouts. The validated attributes in the provided dataset are found using the feature extraction method for competent assessment.

Shobana et al. [7]used a densely integrated deep feed-forward neural network that effectively capitalizes on asymmetric input vectors. Additionally, recent research employs distinct learning behavior in dropout forecast due to performance variance of online courses. In a virtual learning environment, a lack of standard description and comprehension of learning behavior will emerge in a unified decision. It only paid attention to consumer ratings and comments, and it plans to add additional features in the future.

S. Nithya et al.[8] propose using a priori concepts for feature extraction to predict dropouts and artificial neural network approaches to find metric values. In feature selection, it implements the various frequent itemset; therefore, it focuses on the combination of the attributes, but individual feature scores are not concentrated.

Students are more motivated and interested in their studies when using the gamified classroom, which Abdul Rabu et al.[9] Claim is more engaging than the traditional one. The use of both blended and flipped learning elements contributes to the complexity of gamification in the learning process.

Alberto Rivas et al. [10]state that the tree-based model and various kinds of ANN methods are used to apply an automatic learning technique to a real-time dataset and to determine student achievements based on accessing resources on the VLE platform. In order to assess if a student passed or failed the final exam, it merely looks at their academic record.

Concepción Burgos et al. [11]use several machine learning algorithms to analyze dropout prediction using 100 students' history and course grade data from ten distinct courses. In comparison to the other methodologies, LR gives a valid result, according to this experiment. The analysis only makes use of historical data from five different computer engineering courses offered by the Madrid Open University (UDIMA).

Bowei Hong et al. [12]combine multiple machine learning algorithms with a two-layer cascade classifier. They can use the “Synthetic Minority Oversampling Technique (SMOTE) method” on the dataset to get better results. A learning access record, course content, and student enrollment were combined to construct the database, and this method has a very low recall value.

According to Lin Qiu et al. [13](2018), the Feature Selection Predictive framework is used to select features and apply logistic regression and support vector machines for prediction. They can create algorithms for feature extraction and selection, utilizing the dataset XuetangX KDD CUP 2015 for evaluation. The ensemble feature selection is done using similarity metrics and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

random forest techniques. The lowest rated features are excluded using the correlation-based feature removal method. The dataset only contains 30 days of data for each course, and each feature type can have a maximum of 30 features.

Support vector machine (SVM)[14] and Logistic Regression (LR)[15] are often used methods for predicting dropouts in educational data mining traditional binary classification problems. By modeling textual characteristics related to students' feedback using deep learning models and natural language processing methods, the researchers intend to investigate whether activity-wise relevance will have a significant impact on the students' performance in the future. Improved models can guide better course designs and increase student retention.

In both univariate and multivariate digital information, machine learning algorithms have been applied to overcome prediction problems. SVM is the most prominent approach for classification problems in the ML technique because it emits a very low error rate and it organizes the data into linear and non-linear hyper planes[16].
Our strategy aims to combine feature selection with rapid training in order to produce an optimal dropout forecast. Because of its tree structure and theoretical foundation, our work uses the random forest algorithm to address feature selection. Moreover, based on the random forest structure, the identified features are strengthened with varying weights. The mission is to promote the features that allow for accurate classification. For dropout prediction, we used an ANN technique to accomplish quick training. An ANN is a feed-forward neural network with a single hidden layer that improves the gradient method without requiring multiple trials to optimize the parameters.

## 3. METHODOLOGY

In this segment, we explore the framework of the proposed model for online learning dropout prediction. With the goal of enhancing performance, such as predicted accuracy, visualization, and concision of learnt material, feature extraction and selection algorithms are utilized alone or in combination. The advantage of feature selection is that critical information about a specific feature is not lost; however, if only a small collection of features are required and the unique features are quite varied.

**Framework:**

In the first phase, a number of attributes are created and extracted from learners' learning activity records. A random forest is used to build the multiple decision trees in the algorithm, each of which has a score that indicates its relevance or significance. As a result, the algorithm's relevance of the attributes was pooled across numerous decision trees. The feature and the label matrix are the module's outputs.

In the second phase, the RF-ANN approach is utilized to forecast dropouts, which is based on the retrieved learner's learning activity features. Assign multiple thresholds for picking features depending on their values to discover the optimal number of attributes. The Random Forest layer is used to choose features and create an ANN structure around them based on their feature scores.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

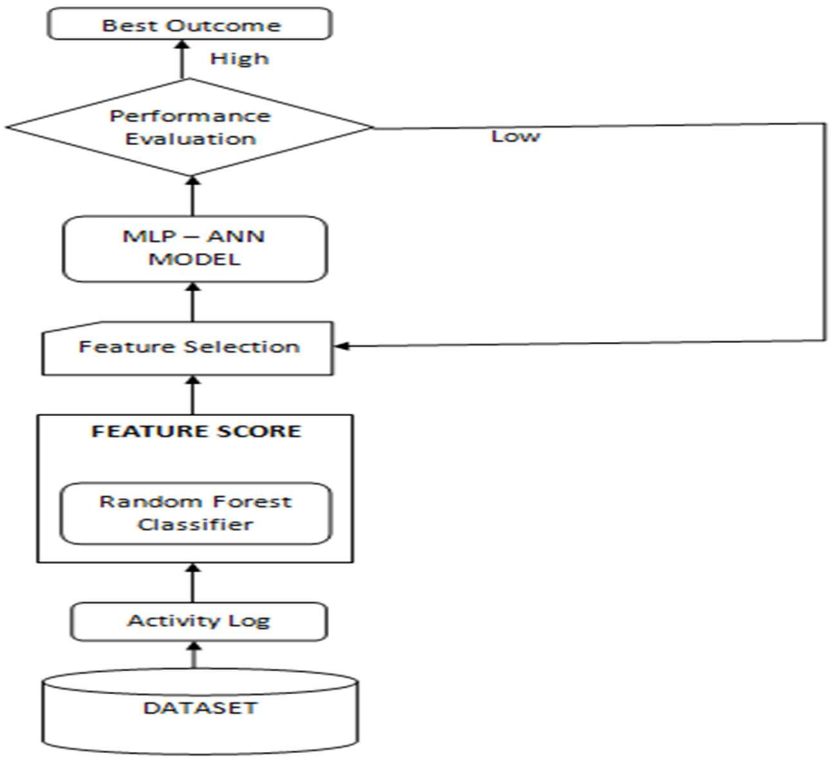


*Figure 2.The Framework of the proposed model*

In this model, Figure 2 illustrates the retrieved activity log of the dataset collected from the UCI Repository, which contains the digital interaction of the learner. The individual feature scores are measured by the random forest classifier. The multilayer perceptron is built and it executes depending on the extracted features[13] that have the highest feature score. The performance evaluations are calculated and compared based on the feature selection, which one achieves the best accuracy value in the shortest time span.

**Feature Extraction:**

Our work extracted the features from the learning activity logs of learners. L represents the set for each sort of learning behavior record. The type of learning behavior is denoted by m and m = 1, 2... f.

The feature matrix $x = [x_1, x_2,...x_a]$ is obtained after feature extraction, where a is the number of enrolled learners. The label matrix is $y = [y_1, y_2,...,y_a]$, and $y = [y_1, y_2,..., y_n] \in L^n$ is the dropout label of the learner. A specified threshold value has been set for the feature selection.

**Enhanced ANN Approach:**

Feature selection has been given a predefined threshold value, which is supplied at the input layer. The proposed structural design is based on the bridging of a random forest and a multi-layered perceptron with one hidden layer and an output layer, as shown in Figure 3.

The number of nodes in the random forest's output layer corresponds to the number of neurons as in the ANN's input neurons. A large numbers of features are combined in the hidden layer to get an output. Because set of possible classes throughout the random forest is the same as the number of neurons in the ANN technique with ReLu activation output layer, the learner will either drop out or not.

Algorithm 1 depicts the feature extraction procedure. It generates the feature and label matrices.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**Algorithm 1: Feature Extraction**

***Input:***

a: The number of enrolled learners

b: Activities centered on the learner

f: The number of distinct actions

***Output:***

Magnitude of the feature matrix, x: a*f

Magnitude of the target matrix, y: a*n.

1. The course's set of learning activities. The set of the learner's behaviors is $L = b_1, b_2... b_m$, where m = 1, 2...f.
2. On the dataset, calculate individual feature scores.
   F.S= Random Forest (L).
   a. If there are L input variables, a number b is supplied, and the best split on this b is used to split the node, with variables randomly taken from the L at each node. The value of b is kept constant during the creation of the forest's countless trees.
   b. To classify a new item from the input vector (b), place it on each of the trees in the decision forest. Everyone is assigned a classification, and the tree is considered as a "voting" member of the class. The forest determines the classification.
3. From the feature matrix, $x= [x_1,x_2,\ldots, x_a]$, where $x= [x_1,x_2, \ldots ,x_f]$ is the f types of learners' activities.
4. $y= [y_1, y2, \ldots, y_a]$ from the label matrix, where $y=[y_1+y_2+\ldots+y_n] \in L^n$.

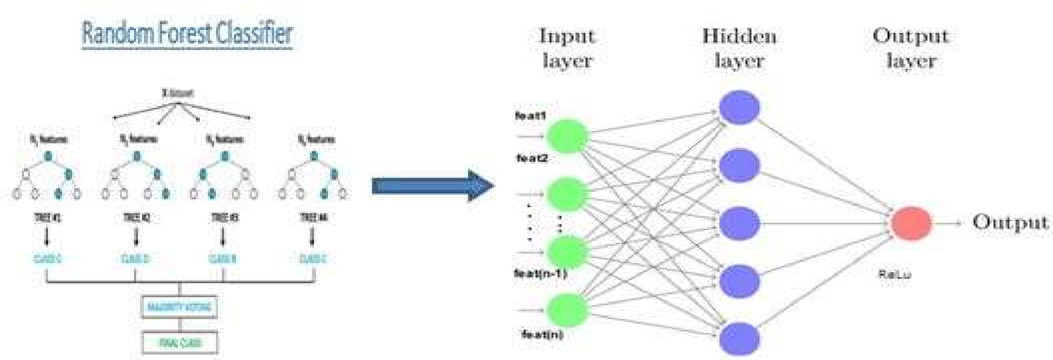


*Figure 3.Architectural Design of RF - ANN*

**Algorithm 2: Enhanced ANN Method:**

1. A specific threshold value has been assigned to feature selection, which is supplied at the input layer.
2. The training set Tr=($x_i$ , $y_i$) $x_i \in L^k$, $y_i \in L^m$ , the activation function g (x), and the number of hidden neurons "h" should all be provided.
3. Assign the input weight vector w and bias vector b at random, except for the connectionless weights and bias between the input and hidden layers.
4. Calculate the Hl output matrix for the hidden layer.
5. Calculate the output weight vector O = Hl'y, where Hl' is the value of the Relu activation function.
6. Using the supplied variables, calculate the anticipated values.

The input layer has been given a predefined threshold value for feature selection. There are K random instances ($x_i$,$y_i$), where $x_i$ = [$x_{i1}$,$x_{i2}$,…..,$x_{ik}$]$^{Tr}$ Є $L^k$, $y_i$= [$y_{i1}$,$y_{i2}$,….., $y_{im}$]$^{Tr}$Є $L^m$. A MLP with
R hidden neurons can be represented as follows:

$$\sum_{j=1}^{R} Ojg\,(x_i.w_j+b_j) = Fd_i,\ i=1,2,\ldots,k \quad (1)$$

The activation function of a hidden neuron is denoted by g(x). $w_j = [w_{j1},w_{j2},\ldots,w_{jn}]^{Tr}$ is the weight vector of input neurons connecting to $i_{th}$ hidden neuron. $x_i \cdot w_j$ is the internal combination of $w_j$ and $x_i$. The $j_{th}$ hidden neuron's bias is bj. $O_j = [O_{j1}, O_{j2}, \ldots, O_{jm}]^{Tr}$ is the sum of the weights of the $j_{th}$ hidden neuron corresponding to the output nodes.
A conventional MLP's goal is to estimate these k samples with a minimum error, which is denoted by (2), where Fdi is the target value and yi is the actual work.

$$\sum_{i=1}^{k} ||Fd_i - y_i|| = 0 \quad (2)$$

In other words, proper wj, bj, and Oj exist such that

$$\sum_{j=1}^{R} Ojg\,(x_i.w_j+b_j) = y_i,\ i=1,2,\ldots,k. \quad (3)$$

Equation (3) can be written as follows.

$$H1O = Y \quad (4)$$

Where

$$H1=\begin{bmatrix} g\,(x1.w1+b1) & \cdots & g\,(x1.wR+bR) \\ \vdots & \ddots & \vdots \\ g\,(xk.w1+bR) & \cdots & g\,(xk.wR+bR) \end{bmatrix} k*R \quad (5)$$

$$O=\begin{bmatrix} O_1 \\ \vdots \\ O_R^{Tr} \end{bmatrix} R*M \quad (6)$$

$$Y=\begin{bmatrix} y_1^{Tr} \\ \vdots \\ y_k^{Tr} \end{bmatrix} k*M \quad (7)$$

The outcome vector of the hidden layer H1 is dynamically computed while wj and bj have been identified. Locating a least-squares response to a set of linear equations can be translated into a learning algorithm (3).

The RF-ANN Algorithm is used to find dropout predictions that can be used as primary characteristics in the development of a learning approach. This feature-based hybrid method [17] may minimize the dataset's lowest interacting characteristics and analyze performance metrics more quickly. We present a hybrid technique for analyzing learners' activities in algorithm 2, which relies on refining the combination of features by applying statistical transformations with multiple dimensions. This hybrid technique identifies the dropout prediction that secures the most salient features while reducing the optimization time.

## 4. EXPERIMENTAL RESULTS

**Dataset Revelation and Preprocess:** To execute our research, we used a real-world dataset. We will leverage XuetangX's MOOC user dataset to forecast if a user will discontinue a course in KDD Cup 2015[18]. As shown in Table 1, this dataset contains five types of information about 39 courses: enrollment information (showing that a student has enrolled in a course), object information (showing course id details), behaviour records information (showing a student's studying records on a course), date information (showing the starting and ending dates of the programs), and truth information (showing whether a student will drop out or continue studying).

Our research focuses on the learners' cognitive techniques, including a priority on activity log data. The activity log in Table 2, which includes assignments [19], watching course videos, accessing digital sources [20], wikis, online interaction, navigation, and closing the web page from the repository data file, is utilized to test the effectiveness of our proposed algorithm. The record was cleaned and transformed to real values using the encoding approach prior to being implemented in the proposed model. In the dataset, there are 72,142 records indicating enrolled users. The binary class label is used in the outcome analysis, with "0" indicating that learners will continue their studies and "1" indicating that they will drop out.

*Table 1. Dataset Depiction*

| Raw Data | Description |
|---|---|
| Enrolment | Registered learners. |
| Object | Program and module relationships |
| Activity Log | Keeping track of one's behavior. |
| Date | Start and finish durations for the program. |
| Truth | Completion or dropout rate of learners. |

*Table 2. Extracted and Selected behavioral features*

| Features | Description |
|---|---|
| $x_1$ | The number of enrolled learners |
| $x_2 – x_{10}$ | The behavior is based on the numbers of accesses, page closes, problems, videos, etc., from the browser respectively. |
| $x_{11} – x_{18}$ | The behavior is based on the numbers of accesses, page closes, problems, videos, etc., from the server respectively. |

**Experimental Framework:** The Google Collaborator and the Scikit-learn Python package have been used to run the samples on a computer system including an Intel Core-i3 central processing unit as well as 8 GB of RAM. The feature score from the dataset is determined using the random forest approach. It creates a number of decision trees that are used to select random feature combinations. This method includes a number of attributes that result in a tree that is widely used for analysis and earns the greatest possible score based on the entropy and information gain values.

**Feature Importance:** In machine learning-based approaches, attribute selection strategies are essential. It gives attributes values based on how significant they are, mostly for strengthening classification outcomes.

The employment of random forest (RF) would be after the bagging trees were assembled, and how frequently a given parameter could be used for a wide selection of elements would be determined [21]. Overall, importance is calculated for every attribute included in the model's bagging decision trees, indicating its efficiency or relevance. Such significant relevance is estimated for every attribute in the data file, which allows for attribute rating and comparability. The relevance of the attributes was therefore aggregated across several decision trees in the algorithm.

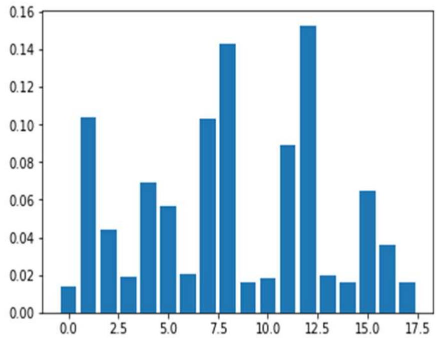


*Figure 4.Individual Feature Importance On The Analysis*

In Figure 4 represents the scores on the individual features in the y-direction and the feature number in the x-direction. The simplest and most efficient approach to the feature extraction while utilizing to boosted trees as the machine learning approach is to directly choose the attributes only with the strongest correlation values. To identify the best number of qualities, we used interaction leads and looked at a number of different thresholds for selecting features depending on their importance. Our work discovered that the model's accuracy declined as the number of features selected grew.

In the score, the feature's threshold value has been set to be greater than 0.02. Ten features should be chosen, and seven features should be excluded during this process. Enrollment id is one of the seven features, with two browser-related behaviors and three server-related behaviors.

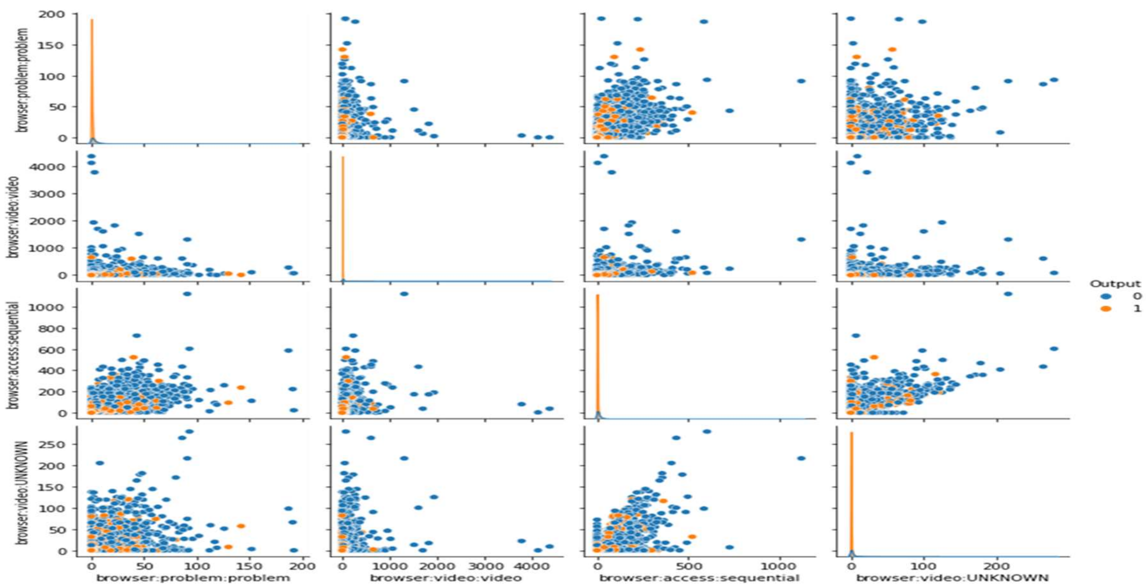


*Figure 5.Browser-Based Feature Analysis*

The dataset was separated into two groups for training and testing: 80 percent and 20 percent. The input layer receives the 10 features that were chosen, and the yield of this layer is integrated and supplied into the hidden layer. The activation function of a neural network's output layer is a rectified linear unit (ReLu) [22], the loss function is binary cross-entropy, the optimizer is Adam, and the learning ratio is 0.01.The five-fold cross-validation is an aid to evaluating the performance measures in this model.

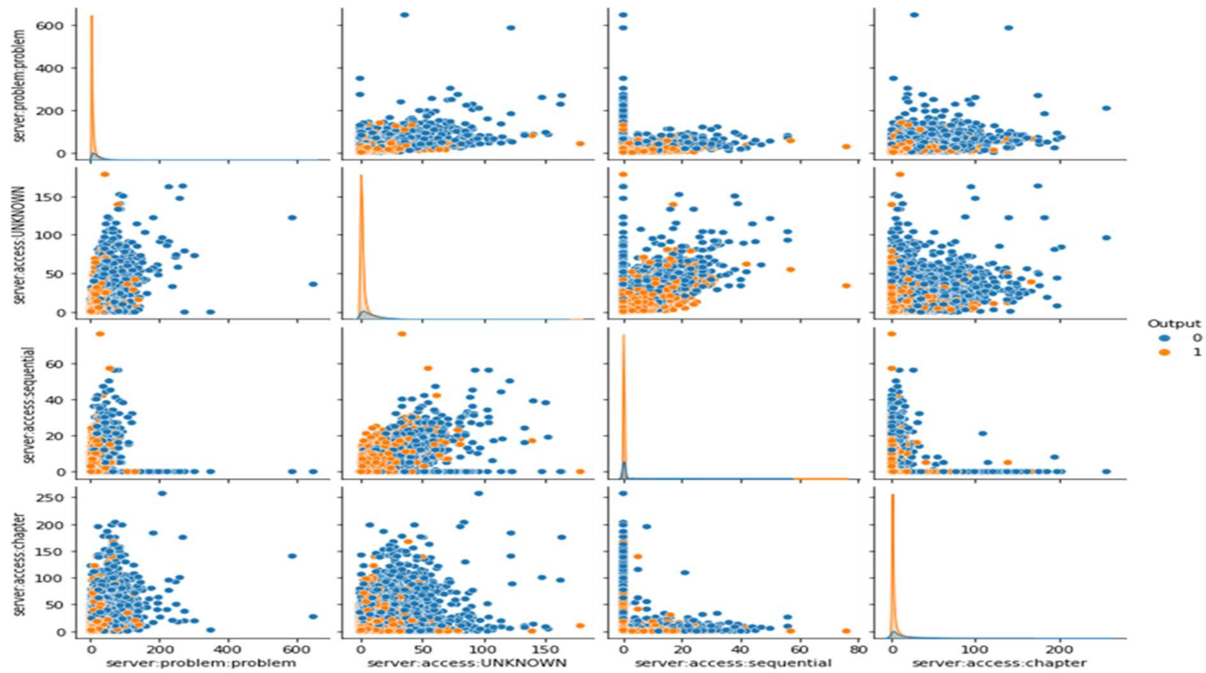


*Figure 6 Feature Analysis Based On Servers*

Figure 5 and 6 exhibits a pair plot of the feature analysis based on the browser and server. It comprises sequential access for learners, difficulties that arise throughout the learning process, and so on, all of which are based on the server and browser.

In this work, we used to find the dropout prediction by frequently used machine learning methods like Logistic Regression (LR) and Support Vector Machine (SVM) [23].

**Evaluation metrics:**

The performance measures are evaluated the metrics as accuracy, precision, recall, and F1 – score. The contingency table is a list of all occurrences of a class that have been classified into a specific category. In this analysis, the class label is used as a binary classification method. The 4 values are present in the 2x2 matrix in the contingency table as indicated in Table 3.

*Table 3.Contingency Table*

| | ACTUAL CLASS | |
|---|---|---|
| PREDICTED CLASS | TP = 166 | FP = 9 |
| | FN = 9 | TN = 146 |

Here, "True Positive" (TP) refers to the set of assured occurrences predicted to be positive, "False Positive" (FP) relates to the amount of uncertain instances estimated to be positive, "True Negative" (TN) corresponds to the quantity of uncertain instances estimated to be negative, and "False Negative" (FN) refers to the quantity of assured occurrences assessed to be negative.

It is quite possible to describe recall in terms of one of the other classes. The proportion of the true positive to the range of real confirmed samples is known as the recall of class label, which is also known as sensitivity. It's easy to think of it as the classifier's power to bring all of the confirmed samples together.

Sensitivity = TP/ Range of real samples

The proportion of the true negative to the number of real negative elements is specified as the recall of a negative sample, also known as "specificity." It can be thought of as the classifier's ability to pick up all of the negative samples.

Specificity = TN/ Range of negative samples

**Comparison and Analysis**

ML approaches are used in the majority of dropout forecasts. The values in Table 4 demonstrate that the prediction accuracy of the selected features by our technique is significantly better than that of the benchmark method, indicating that our feature extraction method is effective.

*Table 4. Comparison with RF and RF-ANN*

| S.No | Metrics/ Learning Alg | Accuracy |
|---|---|---|
| 11 | RF | 78.73 |
| 22 | RF-ANN | 94.86 |

In Figure 7, which compares the accuracy ratings of the RF and RF-ANN methods, the hybrid method outperforms the baseline method by 82%.

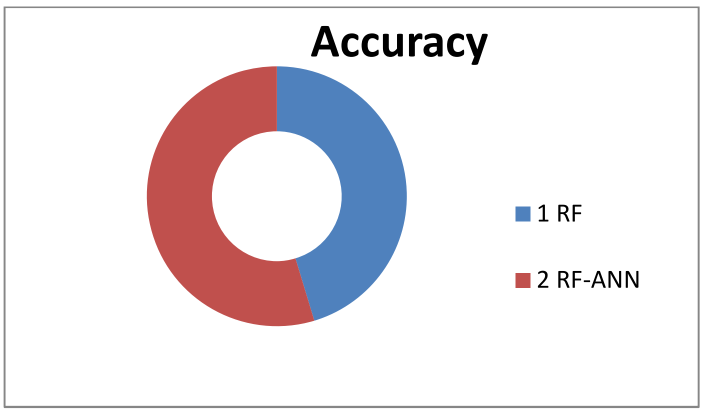


*Figure 7 Compare RF and RF-ANN accuracy.*

In general, techniques based on multiple supervised classification models make predictions, with SVM and LR being the best performance measures for the classifiers. As a result, we trained these two models in this sample and compared the evaluation measure[24] to the proposed model as represents in Table5.

*Table5.Comparison between Classification models*

| SS.No | Metrics/ Learning Alg | | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| 11 | **Baseline models** | ***LR*** | 81.82 | 85.14 | 81.42 | 83.24 |
| 22 | | ***SVM (LinSVC)*** | 81.21 | 86.28 | 79.89 | 82.96 |
| 33 | **ANN** | ***MLP*** | 94.86 | 94.86 | 94.86 | 94.86 |

In this experiment, we compared the different learning algorithms with their accuracy range of 75-95%. The best value was 94.86% present in the MLP-ANN approach and it is illustrated in Figure8.

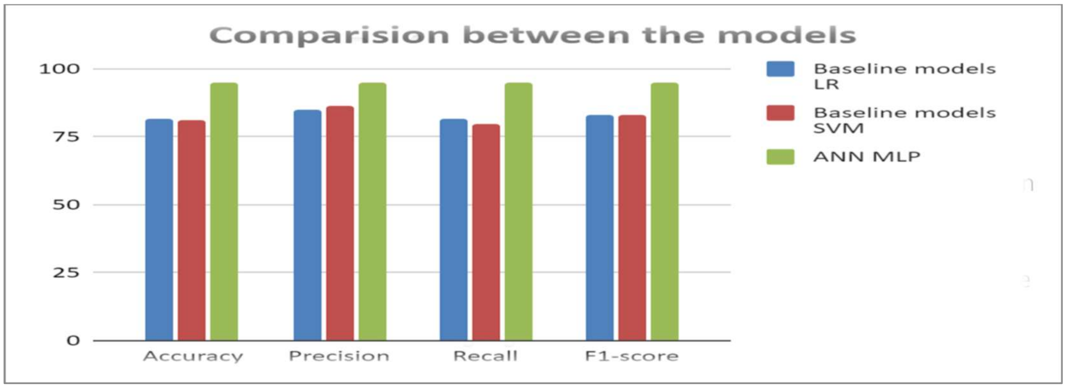


*Figure 8.Learning Analytics of the Model*

Precision values began at 85 in LR, gradually increased to 86 in SVM, and eventually achieved 94 in ANN. In ANN, the recall value peaked at 95, then dropped to 81 in LR, and consequently to 79 in SVM. The values of the F1-score were 83, 82, and 94 in LR, SVM, and ANN respectively.

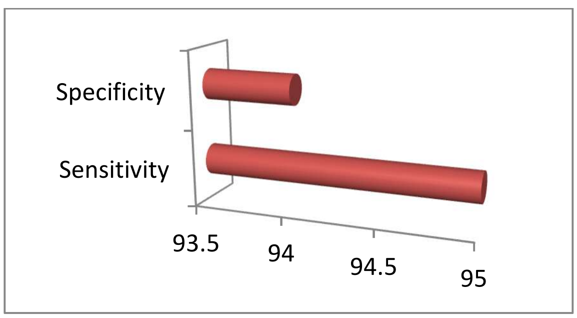


*Figure 9.Sensitivity and Specificity of the proposed model*

In Figure 9, the proposed model's sensitivity and specificity values are 95 percent and 94 percent, respectively.

The results demonstrate that some input selection methods and machine learning models are more effective than others at forecasting the dropout issue in MOOCs.

Considering the outcomes

- The RF-ANN model created in this work is primarily used in a vast amount of MOOC data and emphasizes resolving the benchmark concept's dropout prediction model.
- The KDD CUP 2015 dataset contains learner activity data that is used by the RF-ANN model to anticipate attrition for new initiatives. This model has a 94% accuracy rate.

## 5. CONCLUSION & FUTURE SCOPE

The various approaches were introduced to predict learner dropout in online programs by researchers. This work's significant accomplishment could be best summed up as follows: To begin with, we define and extract a number of interpretive behavior elements from raw learning behavior data using random forest approach. Second, for dropout prediction, our work offers a Multi layer perceptron in ANN. It solves the difficulties of ANN framework setup, recurrent learning, and behavioral inconsistency. It effectively uses the individual feature score based on the random forest structure as a warm-up for the entire algorithm. Finally, our effort tests the efficacy of the proposed method just on the standard KDD 2015 dataset, and we find that it outperforms baseline approaches across various criteria. When compared to the most commonly used machine learning algorithms, such as Logistic Regression and Support Vector Machine, the accuracy numbers are 78 and 85 percent, respectively. In the future, pay closer attention to the feature importance of various types of datasets with different geographical aspects.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

## REFERENCES

[1] M. CONACHE, R. DIMA, and A. MUTU, "A Comparative Analysis of MOOC (Massive Open Online Course) Platforms," *Inform. Econ.*, vol. 20, no. 2/2016, pp. 4–14, 2016, doi: 10.12948/issn14531305/20.2.2016.01.

[2] C. Alario-Hoyos, M. Pérez-Sanagustín, D. Cormier, and C. Delgado-Kloos, "Proposal for a conceptual framework for educators to describe and design MOOCs," *J. Univers. Comput. Sci.*, vol. 20, no. 1, pp. 6–23, 2014, doi: 10.3217/jucs-020-01-0006.

[3] S. S. Sahasrabudhe and R. Majumdar, "MOOC for skill development in 3D animation: Comparing learning perceptions of first time and experienced online learner," *Proc. - IEEE 16th Int. Conf. Adv. Learn. Technol. ICALT 2016*, pp. 6–7, 2016, doi: 10.1109/ICALT.2016.97.

[4] B. T. Ming Wong*, "Pedagogic Orientations of MOOC Platforms: Influence on Course Delivery," *Asian Assoc. Open Univ. J.*, vol. 10, no. 2, pp. 49–66, 2015, doi: 10.1108/aaouj-10-02-2015-b005.

[5] S. Nithya and S. Umarani, "A Novel Analysis of MOOC ' s Prediction," vol. 17, no. 10.

[6] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," *Comput. Human Behav.*, vol. 98, no. April, pp. 166–173, 2019, doi: 10.1016/j.chb.2019.04.015.

[7] J. Shobana and M. Murali, "Abstractive Review Summarization based on Improved Attention Mechanism with Pointer Generator Network Model," *Webology*, vol. 22, no. 1, pp. 77–91, 2021, doi: 10.14704/WEB/V18I1/WEB18028.

[8] S. Nithya, "MOOC Dropout Prediction using FIAR-ANN Model based on Learner Behavioral Features," vol. 13, no. 9, pp. 607–617, 2022.

[9] S. N. Abdul Rabu, N. H. Ismail, N. I. Osman, and S. K. Mohamad, "Motivation, Engagement, Enjoyment, and Learning Achievement Toward Gamified Classroom Via Learning Management System To Enhance Learning Attitude," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 19, pp. 5531–5544, 2022.

[10] A. Rivas, A. González-Briones, G. Hernández, J. Prieto, and P. Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments," *Neurocomputing*, vol. 423, no. xxxx, pp. 713–720, 2021, doi: 10.1016/j.neucom.2020.02.125.

[11] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Comput. Electr. Eng.*, vol. 66, pp. 541–556, 2018, doi: 10.1016/j.compeleceng.2017.03.005.

[12] B. Hong, Z. Wei, and Y. Yang, "Discovering learning behavior patterns to predict dropout in MOOC," *ICCSE 2017 - 12th Int. Conf. Comput. Sci. Educ.*, no. Iccse, pp. 700–704, 2017, doi: 10.1109/ICCSE.2017.8085583.

[13] L. Qiu, Y. Liu, and Y. Liu, "An Integrated Framework with Feature Selection for Dropout Prediction in Massive Open Online Courses," *IEEE Access*, vol. 6, pp. 71474–71484, 2018, doi: 10.1109/ACCESS.2018.2881275.

[14] M. Adnan *et al.*, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," *IEEE Access*, pp. 7519–7539, 2021, doi: 10.1109/ACCESS.2021.3049446.

[15] D. J. Lemay and T. Doleck, "Predicting completion of massive open online course (MOOC) assignments from video viewing behavior," *Interact. Learn. Environ.*, vol. 0, no. 0, pp. 1–12, 2020, doi: 10.1080/10494820.2020.1746673.

[16] R. Chuentawat and Y. Kan-Ngan, "The comparison of PM2.5 forecasting methods in the form of multivariate and univariate time series based on support vector machine and genetic algorithm," *ECTI-CON 2018 - 15th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol.*, pp. 572–575, 2019, doi: 10.1109/ECTICon.2018.8619867.

[17] S. Umarani, T. R. Chaithanya, and M. Divya, "Deployment of P-Cycle in Optical Networks : A Data Mining Approach Deployment of P-Cycle in Optical Networks : A Data Mining Approach," no. September 2016, 2019.

[18] "KDD CUP 2015 Dataset." https://data-mining.philippe-fournier-viger.com/the-kddcup-2015-dataset-download-link/.

[19] I. H. Jo, T. Yu, H. Lee, and Y. Kim, "Relations between student online learning behavior and academic achievement in higher education: A learning analytics approach," *Lect. Notes Educ. Technol.*, no. 9783662441879, pp. 275–287, 2015, doi: 10.1007/978-3-662-44188-6_38.

[20] C. A. Boulton, C. Kent, and H. T. P. Williams, "Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar' university," *Comput. Educ.*, vol. 126, no. September 2017, pp. 129–142, 2018, doi: 10.1016/j.compedu.2018.06.031.

[21] S. Nithya. and Dr.S.Umarani, "Comparative Analysis of the Learning on KDD Cup 2015 Dataset," *Webology*, vol. 19, no. 1, pp. 705–717, 2022, doi: 10.14704/web/v19i1/web19050.

[22] D. Lecoeuche, A. Aussem, and M. Gasse, "On the use of binary stochastic autoencoders for multi-label classification under the zero-one loss," *Procedia Comput. Sci.*, vol. 144, pp. 71–80, 2018, doi: 10.1016/j.procs.2018.10.506.

[23] S. Ardchir, M. A. Talhaoui, H. Jihal, and M. Azzouazi, "Predicting MOOC Dropout Based on Learner ' s Activity Features Predicting MOOC Dropout Based on Learner ' s Activity Features," no. April, 2019.

[24] Y. Wen, Y. Tian, B. Wen, Q. Zhou, G. Cai, and S. Liu, "Consideration of the local correlation of learning behaviors to predict dropouts from MOOCs," *Tsinghua Sci. Technol.*, vol. 25, no. 3, pp. 336–347, 2020, doi: 10.26599/TST.2019.9010013.