IAENG International Association of Engineers
Engineering Letters
Online ISSN: 1816093X, 18160948
Volume 29, Issue 2, Oct -2021

# Feature Aware Predictive Model For Education Data Mining

S Suresh Kumar
Research Scholar, Dept. of CSE
Visvesvaraya Technological University(VTU)-RRC
Bengaluru, Karnataka, India
sureshkatte89@gmail.com

Mallikarjuna Shastry P M
Professor, Dept. of CSE,
Reva Institute of Technology & Management (VTU),
Bengaluru, Karnataka, India
pmmshastry@yahoo.com

***Abstract*—Building predictive model to access and improve student performance has gained wide attention in recent time. Various data mining technique employing machine learning algorithm have been emphasized for building predictive model for improving student performance. Using such model teacher can interpret conceivable factor affecting student final score. Here we study using undergraduate dataset and perform prediction for different session. Here we study different ML technique on undergraduate dataset; the study shows poor result is achieved using standard ML model as feature used here highly non-linear in nature; in addressing ensemble learning mechanism is employed in recent work with good performance; However, poor selection of feature resulted in poor accuracy using standard ensemble learning model; In addressing aforementioned problem in this paper feature aware XGBoost (FA-XGB) model is presented. The FA-XGB employ an effective cross validation scheme to build ensemble classifier. Experiment outcome shows the FA-XGB based predictive achieves much better accuracy, precision, recall, F-measure, and specificity outcome in comparison with XGB and ensemble-based predictive model.**

***Keywords*— E-Learning, Feature Imbalance, Feature importance, Machine Learning.**

## I. Introduction

Education plays a huge role in the modern world. People need a good and proper way of education to survive in this competitive world [1]. Education helps an individual to obtain a good knowledge and a proper understanding of different subjects that are going to be applied in the real-time [2], [3]. Education can be obtained from the practical experiences outside of the classroom. In the 21st Century, as the Internet started expanding, the accessibility of the Internet led to an increase in demand for the e-learning platforms [4].

E-learning is the process where the combine use of hardware, software and educational theory is used to teach in the similar way of traditional learning. There are many techniques through which the traditional teaching can be converted to e-learning. The basic standard technique currently being used is the Learning Management System (LMS) where the curriculum is set using the videos, presentation, or online sessions.

Recently, e-learning platforms are attracting more and more students for better learning experiences, but many of the students encounter different challenges that cause hindrance in their learning. There are many benefits and challenges while learning from the e-learning platforms. An example of a challenge is some students find it difficult to adapt to an e-learning platform immediately after traditional classroom learning [5]. Due to this sudden change, they are not able to adapt the e-learning platforms. Students who are always studying in the classroom may have a different mindset and may not be able to change it and this leads to a low performance in their studies in e-learning platforms [6]. There are different challenges that could be because of technical issues, computer literacy, self-motivation, adaptability struggle, and time management.

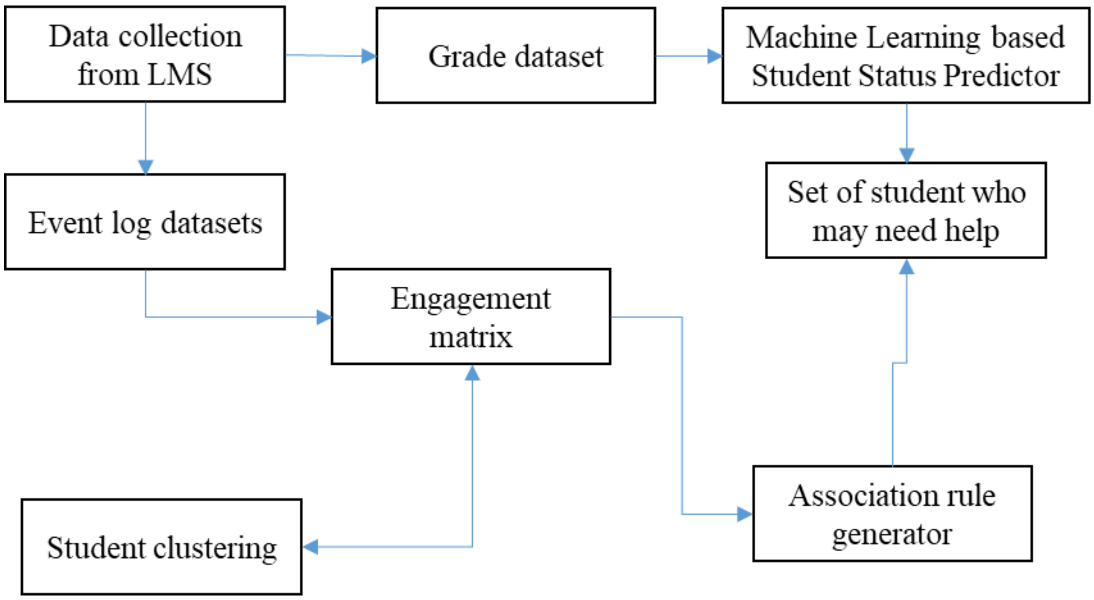


**Fig. 1**. Generic architecture of E-learning system.

Nowadays with the development of technology in e-learning systems, we can capture all the student activities at different stages using the mouse gestures, keystrokes done on the keyboard and clicks of the mouse [7]. The generic architecture of e-learning system is hon in Fig. 1. We can track these different stages using different logs such as chat logs, click stream logs, learning resource usage logs and other communication logs. These logs help to predict the current performance of the student and if any challenges faced can be rectified using the prediction. Many techniques are currently being used to measure the performance of the student in e-learning platforms as discussed in literature survey. From study it is identified ensemble-abased learning model are effective in achieving good prediction precision; however, failed to address feature imbalance issues.

This paper suggests a technique using the ensemble based Machine Learning to predict the performance of the student in the e-learning platform. Here we first model an XGBoost classification method; followed by modelling effective feature aware XGBoost model by employing ensemble learning mechanism. Further, the classification error is minimized through novel cross validation mechanism.

*Research significance of Feature Aware Predictive Model for EDM is described below.*

- The proposed work is efficient even when feature is imbalanced in nature by employing ensemble learning mechanism.

- Minimizes classification error through effective cross validation scheme.
- The proposed predictive model achieves very good classification performance in terms of accuracy, precision, recall, specificity and F-measure in comparison with existing predictive model [21], [22].

## II. Litterature Survey

Here survey of different prediction presented in recent times are studied. The development in the field of Machine Learning and with the use of AI techniques [8]-[12], Shabnam Mohamad Aslam et.al [13] tells the e-learning platform can be made more flexible. To deal with the problems of the e-learning platforms such as easy access of the e-learning platforms to the students, a customized learning platform can be used to increase the performance of the student. In a school the main objective is to conduct the examination, make aware the student their performance, and to introduce development of the different teaching methods in the field of education. In E-learning, examination is an issue for all the teachers. This research is done to check the outcomes of the e-learning systems of different organisations using the AI techniques such as Reinforced Learning Supervised Learning, Semi-Supervised Learning, and soft-computing technique [14]-[16].

Shristi Shakya Khanal et.al [17], they have given a summary using basic recommendation for the e-learning systems using four features: Knowledge-Based, Collaborative Filtering, Hybrid Systems and Content-Based. Using these features, they have generated a scientific technique through which it recommends an effective system. From this technique, they have come to an outcome that machine learning algorithms, techniques, datasets, validation and evaluation of the output is necessary.

Educational Process Mining (EPM) is a developing area in the Educational Data Mining (EDM) Alejandro Bogarin et.al [18], which aims to make knowledge which is not known to the people to be clear and to simplify the understanding of the educational process. The technique of EPM uses log data collected from various institutions so that it can be discovered, analysed and to provide a graphical representation of the whole educational process showing the process of how different things are currently being done in each institution. Their work summarizes the potential of EPM, and how different areas such as graph mining, intentional mining and sequential pattern mining are connected with each other with this mining technique. It also gives a brief description of the components of EPM and addresses the challenges faced by this mining technique.

Amita Dhankhar et.al [19], aimed at finding different ML method [20] used for enhancing prediction and also identify features used for validating performance of students. Mohammad Noor Injadat et.al [21] and [22] analysed undergraduate dataset using different ML technique and showed the need to build ensemble learning based predictive model for EDM. However, their ensemble model [23]-[25] could not achieve good accuracy requirement of EDM. In addressing research issues in next section this paper present feature aware prediction model for EDM.

## III. Feature Aware Prediction Model for Education Data Mininig

This section present feature importance aware prediction (FIAP) model for education data mining. A general education datamining (EDM) dataset can be represented using following equation

$$E = \{(a_1, b_1), (a_2, b_2), \dots, (a_m, b_m)\} \quad (1)$$

where $j = 1,2,3, \dots, m$, defines number of samples. The parameter $b_j \in \{-1,1\}$ represent $j^{th}$ sample outcome, and $a_j$ is a $n$ −dimension vector representing independent features observed of sample $j$. The EDM dataset is generally multi-dimensional in nature considering diverse feature set; however, with less number of sample $m$. Then to analyze and design a predictive model $\hat{G}$, for predicting the actual estimate of original function $G$ considering respective class label as described below

$$g: A \rightarrow B \quad (2)$$

where $A$ defines representing independent features observed of sample $j$ and $B$ depicts outcome of $j^{th}$ sample. Here we design a predictive model by minimizing certain objective function using XGBoost classification model.

### a) *XGBoost Prediction Algorithm:*

XGBoost algorithm is a well-known gradient tree boosting methodologies used by various standard models for solving various classification problems [26]. The idea behind using gradient tree boosting methodologies is to obtain outcomes by cumulating several tree classifiers. Therefore, for classifying student good or weak, the models trains dataset with $o$ samples with several classifier as described in below equation

$$\hat{Z}_j = G(Y_j) = \sum_{l=1}^{L} g_l(Y_j), \qquad g_l \in \alpha \quad (3)$$

where $Y_j$ depicts the $j^{th}$data within training data, $L$ depicts tree size used for classifying the EPM dataset, $\hat{Z}_j$ defines the classification outcomes of multi-label classification model with certain dimension, $l^{th}$ dimension describes the probability that it will be classified as being belonged to the $l^{th}$ class, and $\alpha$ defines set of decision tress as described below

$$\alpha = \{g(y) = x_{t(y)}\} \quad (4)$$

where every tree $g(y)$ agree with respect to leaf weight $x$ and structure parameter $t$. The objective of XGBoost classification model is to minimize the loss parameter

$$M(G) = \sum_{j} m(\hat{z}_j, z_j) + \sum_{l} \beta(g_l) \quad (5)$$

where

$$\beta(g_l) = \delta U + \mu \|x\|^2 \quad (6)$$

The first parameter $m(\hat{z}_j, z_j)$ in Eq. (5) defines the loss function among actual and classified outcomes. The second parameter $\beta(g_l)$ in Eq. (5) depicts the penalizing term; $U$

depicts leaves size within a tree, $\delta$ and $\mu$ depict the controlling parameter used for controlling computational complexity. In this work weighted loss function is considered for training data $y$ whose ID is described by $m$, the negative log probabilistic loss function is obtained using following equation

$$m(\hat{z}_j, z_j) = -\sum_k z(k) \log \hat{z}\,(m) = -\log \hat{z}\,(m) \tag{7}$$

where $z(k)$ depicts $k^{th}$ dimension of $z$, $\hat{z}(m)$ depicts the $k^{th}$ dimension of output $\hat{z}$. Further, the loss function is optimized in iterative manner for obtaining minimum loss. Thus, the optimized loss function under certain iteration $u$ can be described using following equation

$$M^j = \sum_{j=1}^{o} m\left(\hat{z}_j^{(u-1)} + g_u(y_j), z_j\right) + \beta(g_u) \tag{8}$$

The proposed methodology establish $g_u$ that can minimize the loss in greedy manner using following equation

$$M^u \cong \sum_{j=1}^{o} \left[m\left(\hat{z}_j^{(U-1)} + z_j\right) + h_j g_j(y_j) + \frac{1}{2} i_j g_u^2(y_j)\right] + \beta(g_u) \tag{9}$$

where $h_j$ depicts first order gradient of $m\left(\hat{z}_j^{(U-1)} + z_j\right)$ and $i_j$ depicts second order gradient of $m\left(\hat{z}_j^{(U-1)} + z_j\right)$; thus the tree $g_u$ can be established by minimizing Eq. (9) [28]. However, accuracy of classification using standard XGBoosting algorithm is severely affected due to selection of wrong feature. Thus, in next section this work present an ensemble-based learning model for XGBoost algorithm for achieving high prediction accuracy by optimizing the parameter through classification error minimization.

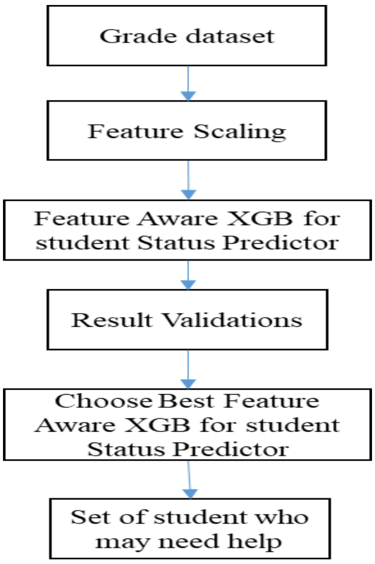


**Fig. 2**. Architecture of Feature Aware XGB-based EDM.

*b) Feature aware XGBoost algorithm:*

Here we employ a cross validation (CV) mechanism selecting useful feature sets in optimizing predictive model. Here we select predictive model that reduces validation error. In most of the standard EPM model have employed $K$-fold CV scheme for optimizing output. In $K$-fold CV, the EPM dataset is divided in random manner into $K$ subset of identical size; then $K-1$ subset are used for building predictive model and leftover subset are used for predicting errors in model. Finally, the $K$ combination of predicted error are average for obtaining CV errors. Later, a grid of $l$ suitable value are generated for establishing ideal optimizing parameter in minimizing CV errors. Finally, the model with minimum CV error are selected;

Here we model hybrid CV scheme by combining iterative cross validation (ICV) scheme and feature aware cross validation (FACV) scheme to build a predictive model that minimize prediction error considering feature importance. The FIAP-EDM model, employs a CV with two layer. In layer 1, features subsets are chosen as the main features. In layer 2, the main subset feature selected from layer 1 are used for building final predictive model.

Here we first model an **Iterative CV scheme** by constructing multiple sets of $K$ folds rather than constructing single $K$-fold sets; the single fold CV error is obtained using following equation

$$CV(\sigma) = \frac{1}{M}\sum_{k=1}^{K}\sum_{j\in G_{-k}} P\left(b_j, \hat{g}_{\sigma}^{-k(j)}(y_j,\sigma)\right) \tag{10}$$

The modified iterative CV error is obtained using following equation

$$CV(\sigma) = \frac{1}{SM}\sum_{s=1}^{S}\sum_{k=1}^{K}\sum_{j\in G_{-k}} P\left(b_j, \hat{g}_{\sigma}^{-k(j)}(y_j,\sigma)\right) \tag{11}$$

Then, the optimization parameter for selecting optimal value $\hat{\sigma}$ is obtained using following equation

$$\hat{\sigma} = \underset{\sigma\in\{\sigma_1,\dots,\sigma_l\}}{\arg\min}\ CV_s(\sigma) \tag{12}$$

In above equations, $P(\cdot)$ represent loss function, $\hat{g}_{\sigma}^{-k(j)}(\cdot)$ represent a function for estimating coefficients, and $M$ describes training data size.

In **Feature Aware CV scheme**, the predictive model is built by estimating multiple optimization parameters. Here multilayer of $K$-folds CV are created; the layer size is set according to optimization parameter considered; i.e., if parameter is optimized in layer 1, the parameter value is fixed and given to layer 2 for estimating additional optimization parameters.

**Predictive Model construction using hybrid cross validation scheme**.

**Phase 1. Feature selection:**

**Step 1.** First the EPDM dataset $E$ is roughly divided into $K$-folds of identical size. Now let's define parameter $E^{-k}$considering $k=1$ to $K$, with $k^{th}$ data tuples eliminated for outer training sample and $E^{k}$ with leftover data $k^{th}$ part is used for outer testing sample.

**1)** The below steps are iterated considering pre-configured size $S$. Then, the EPDM data $E^{-k}$ is randomly divided into $H$-folds of identical size. For $h = 1$ to $H$.

**A).** Outline $H$ different EPDM dataset $E^{-kh}$ with $h^{th}$ segment eliminated for inner training data and $E^{kh}$ with $h^{th}$ segment is used for inner testing data. For $l = 1$ representing grid size configured for optimizing parameter.

**I).** Construct predictive model $\hat{g}_{\sigma_l} = \hat{g}\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right)$.

**II).** Compute error utilizing loss function with respect to inner testing dataset by applying $\hat{g}_{\sigma_l}$ on inner testing dataset $E^{kh}$ as follows

$$\mathcal{E}_{\sigma_n} = \sum_{j \in E^{-kh}} P\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right) \tag{13}$$

**B).** Estimate $H$-folds CV errors for every $l$, thus, will have varied CV error considering varied tuples size $M_h$ in layer 1 for $k^{th}$ segment.

$$CV(\hat{g}; \sigma_l) = \frac{1}{M_h} \sum_{h=1}^{H} \sum_{j \in E^{-kh}} P\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right) \tag{14}$$

**C).** Here we establish the cross validation error of iterative CV process for different $l$ by iterating aforementioned steps $S$-times considering tuple size $M_h$ in layer 1 for $k^{th}$ segment as follows

$$CV_S(\hat{g}; \sigma_l) = \frac{1}{M_h S} \sum_{s=1}^{S} \sum_{h=1}^{H} \sum_{j \in E^{-kh}} P\left(b_j, \hat{g}(E^{-kh}; \sigma_l)\right) \tag{15}$$

**2.** Establish ideal value for optimizing parameter considering all probable $l$ through following equation

$$\hat{\sigma}_n = \underset{\sigma \in \{\sigma_1, \sigma_l\}}{\arg\min}\, CV_S(\hat{g}; \sigma_l) \tag{16}$$

**3.** The ideal value of optimization parameter are configured by minimizing objective function through gradient decent model; then the feature subset is established and final predictive model is selected through ranking model $r(\cdot)$ as follows

$$r(a) = \begin{cases} 0 \;\; if\ n_j\ is\ not\ chosen \\ 1 \;\; if\ n_j\ is\ chosen\ as\ final\ predictive\ model\ j = 1,2,3, \dots, n \end{cases} \tag{17}$$

The feature subset is obtained using following equation

$$F_s = \{r(n_1), r(n_1), \dots, r(n_n)\}, \tag{18}$$

where for different $K$-folds, the important feature with highest rank are established using following equation

$$F_{s_k} = \{r(n_1), r(n_1), \dots, r(n_n)\} \tag{19}$$

**Step 2.** We estimate how many times a feature is chosen for $K$ feature subsets with highest rank. Then, final feature subset is obtained using following equation

IAENG International Association of Engineers
Engineering Letters
Online ISSN: 1816093X, 18160948
Volume 29, Issue 2, Oct -2021

$$F_{s_{final}}=\{f_s(p_1), f_s(n_2), \ldots, f_s(n_n)\}, \tag{20}$$

where $f_s(\cdot)$ indicates whether $n^{th}$ feature is chosen or not as described below

$$F_s(a) = \begin{cases} 0 \;\; if\ q_j\ is\ chosen\ lesser\ than\ \frac{K}{2}\ times, j = 1,2,3, \ldots, n \\ 1 \;\; if\ q_j\ is\ chosen\ greater\ or\ equal\ to\ \frac{K}{2}\ times, j = 1,2,3, \ldots, n \end{cases} \tag{21}$$

**Step 3.** The above steps generates subset of $n'$ chosen features, where $n^{th}$ define number of feature chosen. The dataset used for training is subset through chosen features for constructing predictive model.

**Phase 2. Predictive model construction**

**Step 1.** Here we reduce the dataset $E$ to $E'$ by keeping feature selected in Phase 1, where $E' = (E; n')$.

**Step 2.** Similar to phase 1 the $K$-folds is considered. For $k = 1$ to $K$

**1).** Describe EPDM dataset $E'^{(-k)}$ with $k^{th}$ segment eliminated to perform training and $E'^{(k)}$ segment leftover is used as testing dataset. The below steps are done in iterative manner with predefined size $S$. For $l = 1$ to $L$, where $L$ defines the grid size for optimizing parameters.

**A).** Construct predictive model using following equation

$$\hat{g}_{\sigma_l} = \hat{g}\left(E'^{(-k)}; \sigma_l\right) \tag{22}$$

**B).** apply $\hat{g}_{\sigma_l}$ on inner testing dataset $E'^{(-k)}$ and estimate error utilizing below loss function equation for different $l$

$$\mathcal{E}_{\sigma_l} = P\left(b_j, \hat{g}\left(E'^{(-k)}; \sigma_l\right)\right) \tag{23}$$

**2).** Estimate the $K$-fold CV error for different $L$ of the optimizing parameter as follows

$$CV(\hat{g}; \sigma_l) = \frac{1}{M}\sum_{k=1}^{K}\sum_{j\in E'^{(-k)}} P\left(b_j, \hat{g}\left(E'^{(-k)}; \sigma_l\right)\right) \tag{24}$$

**3).** Obtain the cross validation error through iterative CV scheme using following equation

$$CV_S(\hat{g}; \sigma_l) = \frac{1}{KS}\sum_{s=1}^{S}\sum_{k=1}^{K}\sum_{j\in E'^{(-k)}} P\left(b_j, \hat{g}\left(E'^{(-k)}; \sigma_l\right)\right) \tag{25}$$

**Step 3.** Obtain the ideal value of optimizing parameter for different $l$ value using following equation

$$\hat{\sigma} = \underset{\sigma\in\{\sigma_1,\ldots,\sigma_l\}}{\arg\min}\ CV_S(\hat{g}; \sigma_l) \tag{26}$$

Step 4. The ideal value of optimizing parameter are configured as objective function which is minimized through gradient decent algorithm in obtaining final predictive model. In order to train the predictive model in layer 1, $H$-folds are constructed and iterated $S$ times for reducing randomness in predictive model considering different folds; thus, aid in reducing variance. The Layer 2 uses subset of feature set chosen for constructing final predictive model; thus aid in reducing the variance. The proposed predictive model through modified XGBoost algorithm aid in achieving higher accuracies in comparison with standard predictive model through machine learning models which is proved through experiment in next section.

**Table 1**. Example of prediction done using XGB and FIA-XGB model

| ID | Actual Class | G | W | Predicted Class_FIA-XGB | G | W | Predicted Class_XGB |
|---|---|---|---|---|---|---|---|
| 1 | **G** | 1 | 0 | **G** | 1 | 0 | **G** |
| 2 | **G** | 1 | 0 | **G** | 0 | 1 | **W** |
| 3 | **W** | 0 | 1 | **W** | 0 | 1 | **G** |
| 4 | **G** | 1 | 0 | **G** | 1 | 0 | **G** |
| 5 | **G** | 1 | 0 | **G** | 0 | 1 | **W** |
| 6 | **W** | 0 | 1 | **G** | 0 | 1 | **G** |
| 7 | **W** | 1 | 0 | **W** | 1 | 0 | **W** |
| 8 | **G** | 1 | 0 | **G** | 1 | 0 | **G** |
| 9 | **G** | 1 | 0 | **G** | 1 | 0 | **G** |
| 10 | **G** | 1 | 0 | **G** | 0 | 1 | **W** |
| 11 | **W** | 0 | 1 | **W** | 0 | 1 | **W** |
| 12 | **W** | 0 | 1 | **W** | 0 | 1 | **W** |

## IV. Result and Analysis

Here experiment is conducted to study the performance different predictive model on EPM dataset. Here we study different predictive model such as random forest (RF), Multi-layer perception (MLP), K-Nearest neighbor (KNN), Naïve Bayes (NB), Linear Regression (LR), support Vector Machine (SVM), Ensemble-based [21], [22], XGBoost (XGB), and proposed Feature Aware-XGB (FA-XGB) for analyzing on EPM dataset. All the model are implemented using python 3 framework. The accuracy, precision, specificity, recall, and F-measure are metrics used for validating different predictive models using following equation

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{27}$$

$$Precision = \frac{TP}{TP + FP} \tag{28}$$

$$Recall = \frac{TP}{TP + FN} \tag{29}$$

$$F - measure = \frac{2 * Precision * Recall}{Precision * Recall} \tag{30}$$

$$Specificity = \frac{TN}{TN + FP} \tag{31}$$

*a) Dataset description:*

Here experiment is done using dataset collected from under graduated student of University of Genoa [27]. The dataset of student from first-year which is collected through simulation environment of e-learning platform namely Digital Electronics Education and Design Suite. The EPM dataset is composed of different student sessions information such as Key stroke (KS), working time (WT), Mouse movement (MM), Mouse wheel (MW), Mouse left click

(MLC), Mouse right click (MRC), and Mouse wheel click (MWC). The e-learning environment provides wide variety of courses through integrative browser where it as student to complete different problem with varied complexities. In Table 1, the sample prediction outcome obtained using XGB and FIA-XGB is shown.

*b) Predictive model performance evaluation:*

Here we conducted experiment to validate the different predictive model performance. Here performing classification is considered to be a binary classification problem (i.e., prediction is done to classify whether student will fail or not in fourth coming session/exam). In here student who score more than 50 marks are considered as good and student who score below 50 marks are considered as weak/fail. The specificity and sensitivity performance achieved using different predictive model such as RF, MLP, KNN, NB, LR, SVM, Ensemble, XGB, and FA-XGB is graphically shown in Fig. 3. From Fig. 3 we can conclude that the NB-based predictive model achieves higher specificity outcome (i.e., 1); however at the cost of sensitivity (i.e., 0.857). On the other side the proposed FA-XGB-based predictive model achieves good balance between specificity (i.e., .945) and sensitivity (i.e., 1).

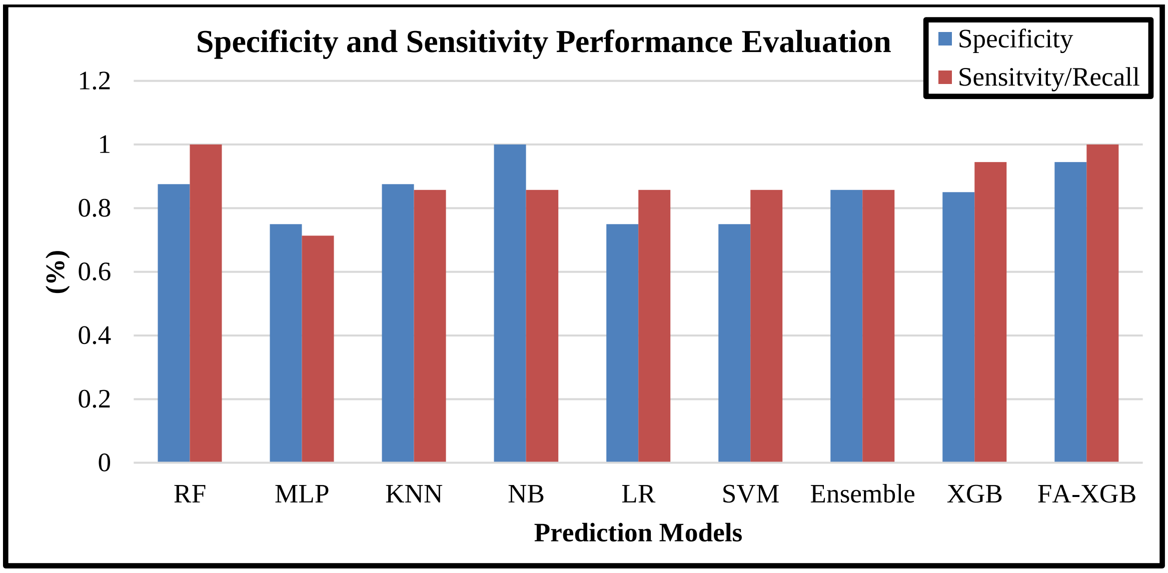


**Fig. 3**. Specificity and sensitivity performance of different predictive model for educational data mining.

Further, performance is validated considering different ROC metric such as specificity, recall, accuracy, precision, and F-measure using different predictive model as shown in Fig. 4. From Fig. 4 we can see the FA-XGB-based predictive model achieves much better performance in comparison with XGB and Ensemble-based predictive model.

IAENG International Association of Engineers
Engineering Letters
Online ISSN: 1816093X, 18160948
Volume 29, Issue 2, Oct -2021

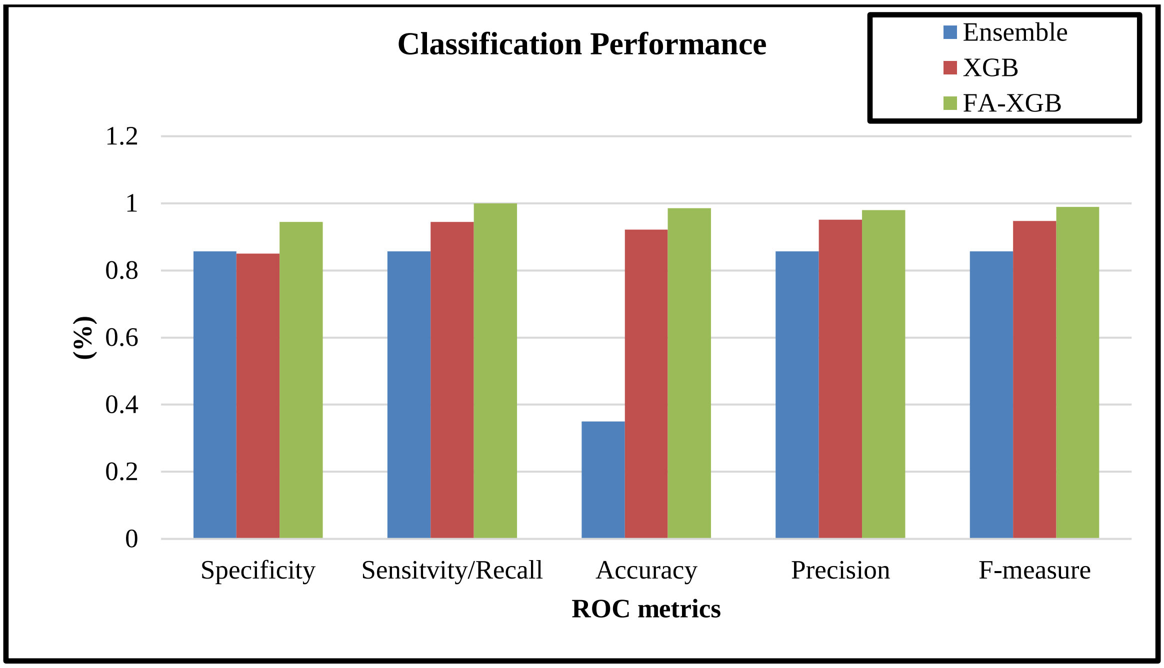


**Fig. 4**. Classification outcome achieved using different predictive model for educational data mining.

*c) Feature importance performance evaluation:*

The Fig. 3 shows graphical representation of feature importance parameter obtained using XGB and FA-XGB-based predictive model. From Fig 5, we can see the FA-XGB gives higher importance to feature in comparison with XGB. Further, the FA-XGB based predictive model gives importance in following order KS, WT, MM, MW, MLC, MRC, and MWC. On the other side, the XGB-based predictive model gives importance in following order WT, KS, MW, MM, MRC, MLC, and MWC. Further, it is noticed in both cases MWC is given very less importance. The Fig. 5 shows how selecting right feature aid in improving overall classification accuracy of proposed FA_XGB based predictive model.

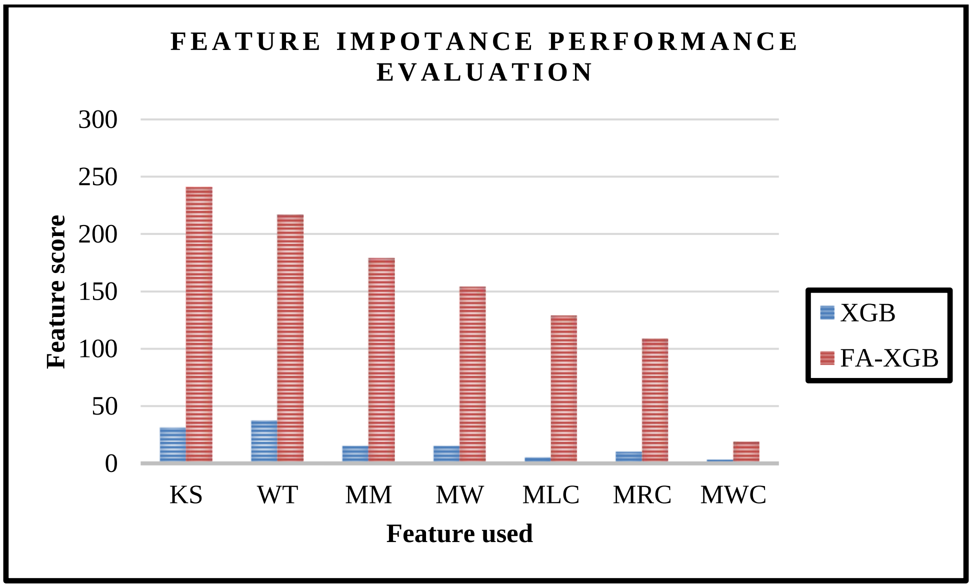


**Fig. 5**. Graphical representation of feature importance performance.

*d) Predictive model performance evaluation under different sessions:*

Here the performance is validated considering different ROC metric such as specificity, recall, accuracy, precision, and F-measure for different sessions such as session 2, session 3, session 4, session 5, and session 6 using different predictive model such as XGB and FA-XGB as shown in Fig. 6 to Fig. 10, respectively. From Fig. 6 to Fig. 10 we can see the FA-XGB-based predictive model achieves much better ROC performance in comparison with XGB-based predictive model.

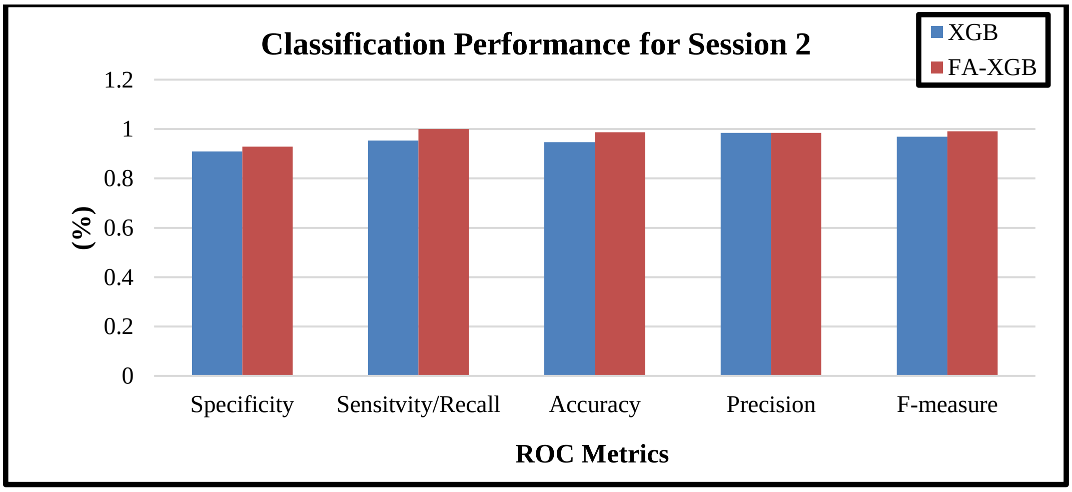


**Fig. 6**. Classification outcome achieved using different predictive model for educational data mining for Session 2.

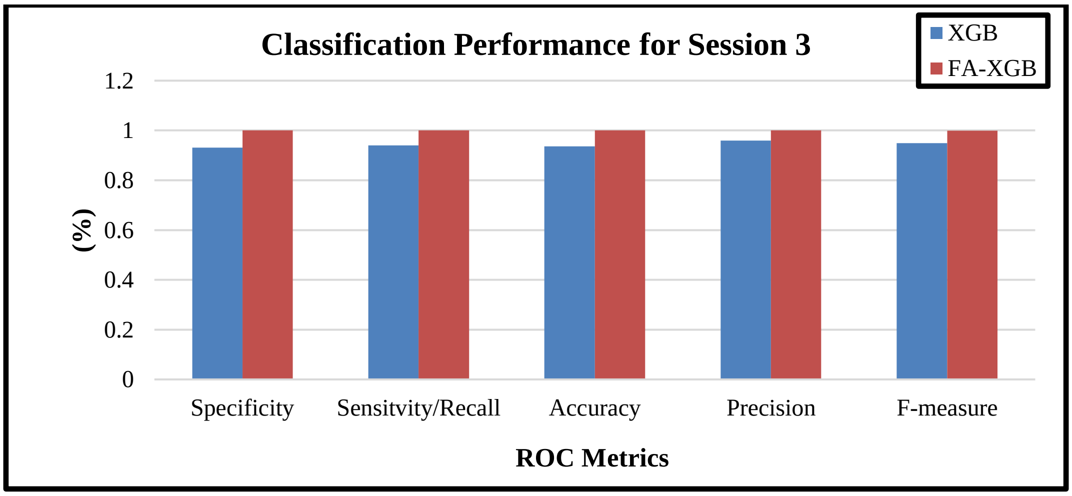


**Fig. 7**. Classification outcome achieved using different predictive model for educational data mining for Session 3.

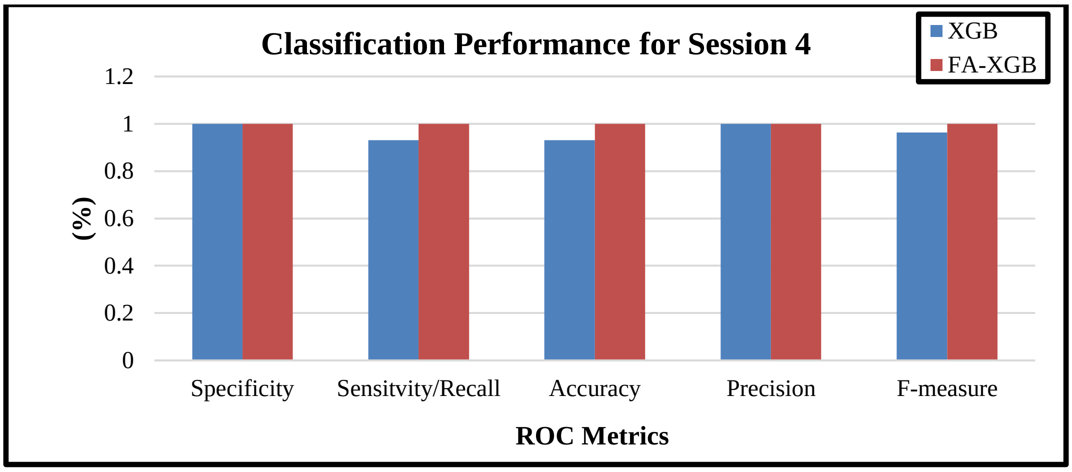


**Fig. 8**. Classification outcome achieved using different predictive model for educational data mining for Session 4.

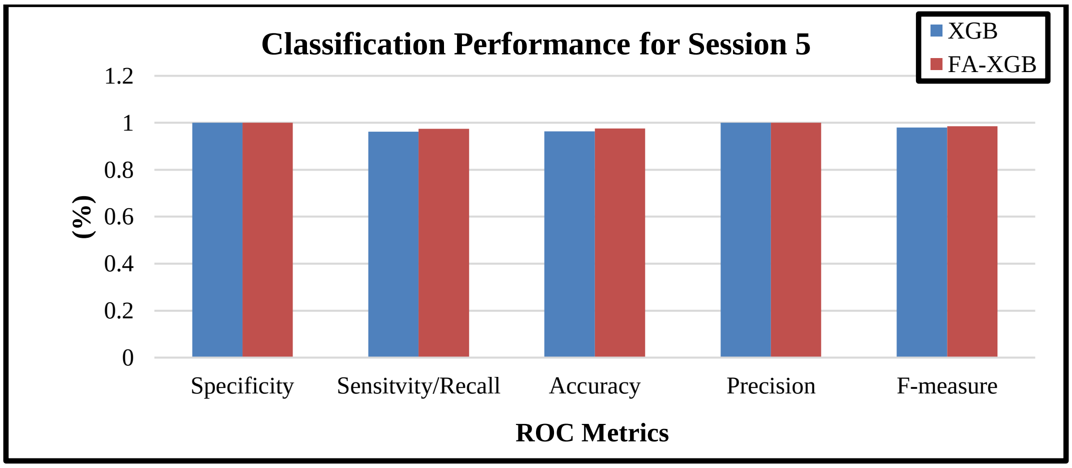


**Fig. 9**. Classification outcome achieved using different predictive model for educational data mining for Session 5.

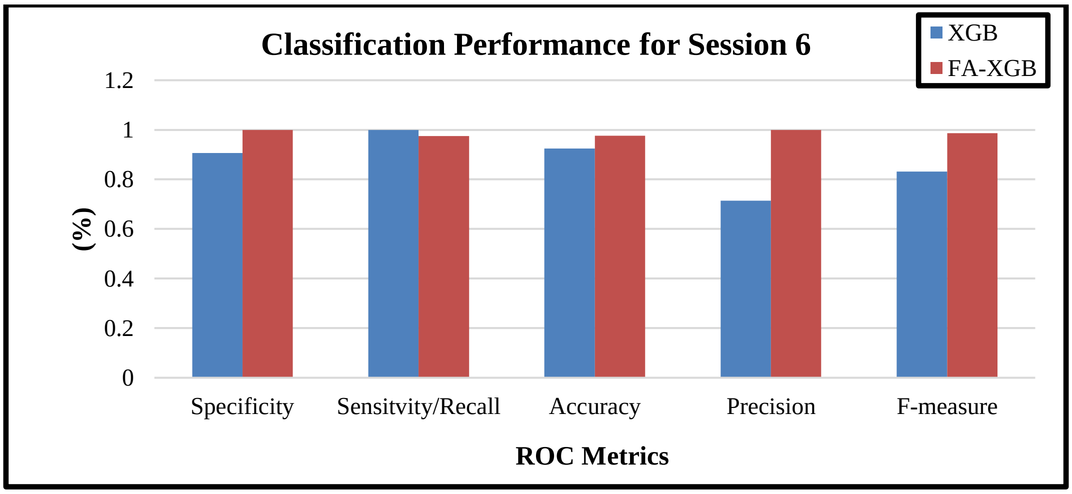


**Fig. 10**. Classification outcome achieved using different predictive model for educational data mining for Session 6.

IAENG International Association of Engineers
Engineering Letters
Online ISSN: 1816093X, 18160948
Volume 29, Issue 2, Oct -2021

*e) Feature importance performance evaluation for different sessions:*

The Fig. 11 to Fig. 15 shows graphical representation of feature importance parameter obtained using XGB and FA-XGB-based predictive model considering different sessions. From Fig 11, we can see the FA-XGB gives higher importance to WT feature and lower importance MRC feature; however, XGB gives higher importance to MW and lower importance to MRC with respect to session 2. In Fig. 12, we can see the both FA-XGB and XGB gives higher importance to MM feature and lower importance MWC feature with respect to session 3; however, higher weights is given by FA-XGB in comparison with XGB. In Fig 13, we can see the FA-XGB gives higher importance to KS feature and lower importance MLC feature; however, XGB gives higher importance to WT and lower importance to MW with respect to session 4. In Fig 14, we can see the FA-XGB gives higher importance to MW feature and lower importance MWC feature; however, XGB gives higher importance to MW and lower importance to MLC with respect to session 5; however, the XGB model completely neglects MRC features. In Fig 15, we can see the FA-XGB gives higher importance to MM feature and lower importance MWC feature; however, XGB gives higher importance to MM and lower importance to MWC with respect to session 6; however, the XGB model completely neglects MW features. From Fig. 9 to Fig. 13 we can interpret that selecting right feature aid in improving overall classification accuracy of proposed FA-XGB based predictive model in comparison with XGB-based predictive model.

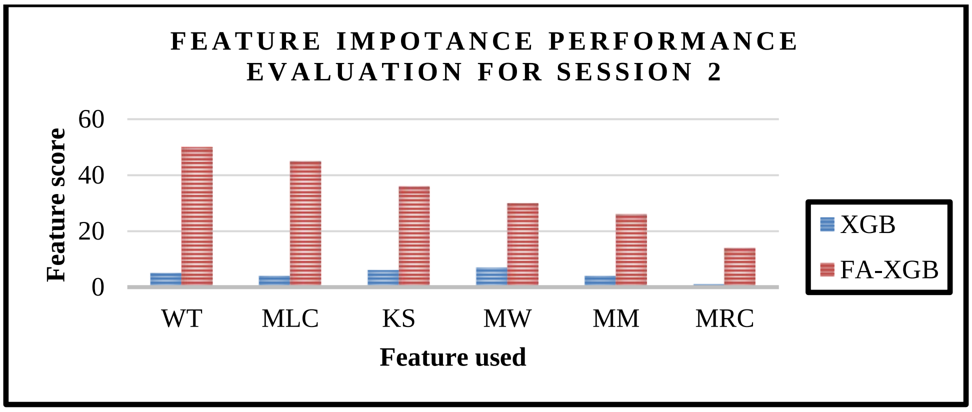


**Fig. 11**. Graphical representation of feature importance performance for Session 2.

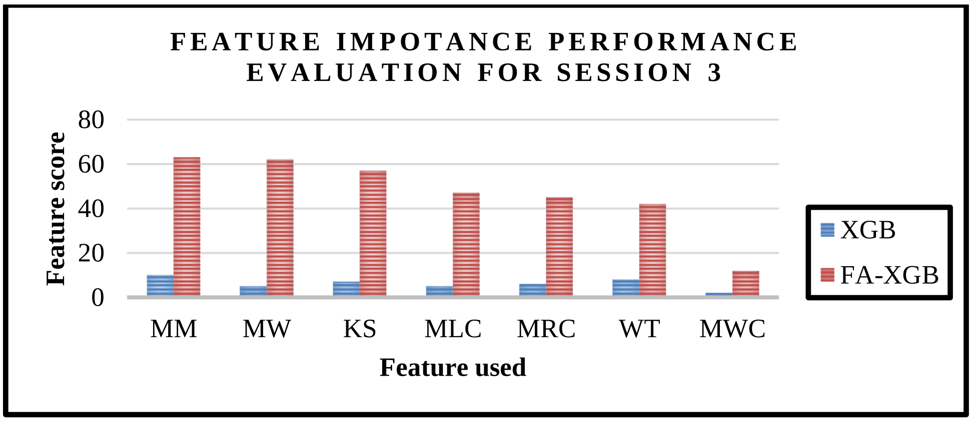


**Fig. 12**. Graphical representation of feature importance performance for Session 3.

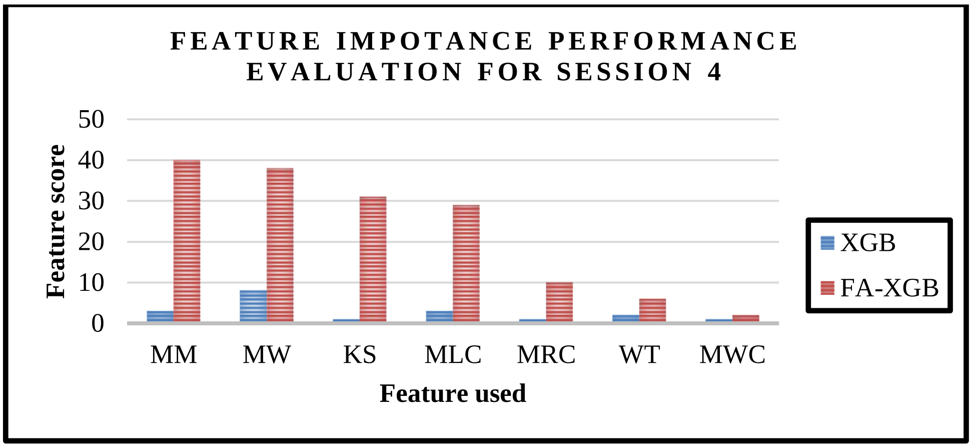


**Fig. 13**. Graphical representation of feature importance performance for Session 4.

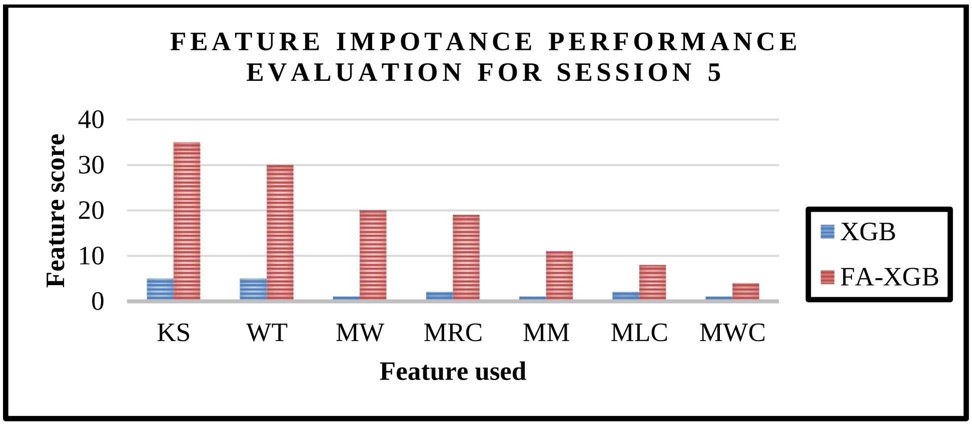


**Fig. 14**. Graphical representation of feature importance performance for Session 5.

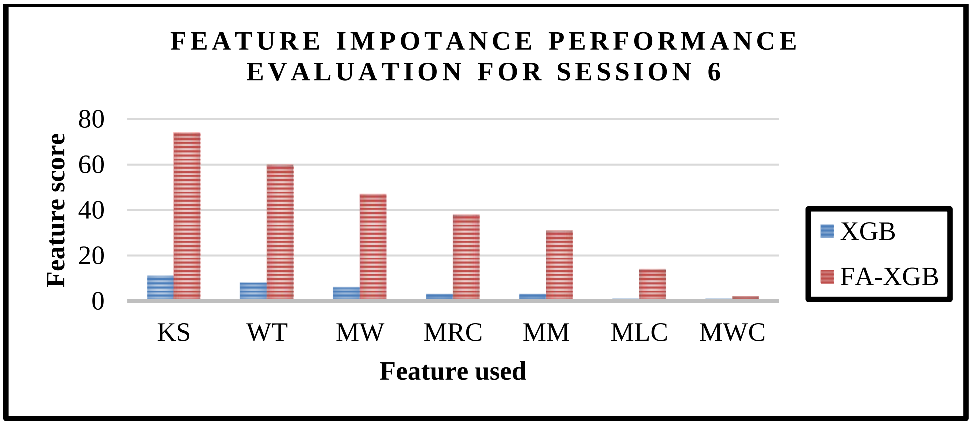


**Fig. 15**. Graphical representation of feature importance performance for Session 6.

## V. Conclusion

Educational data process mining has attained wide interest in recent time for aiming to personalize the content delivery and also enhance the performance of students. Researcher have focused on modelling predictive model employing machine learning algorithms to improve performance of students. However, such task are not easy to be accomplished; in particular during course of study. The EDM data generally exhibit high feature imbalance and non-linear in nature; Using ML fails to provide satisfactory result; thus ensemble learning mechanism is been employed with good performance. However, selection of right feature plays an important role in increasing classification accuracy. Thus, classification accuracy outcome obtained using existing ensemble learning is very poor; in addressing this paper presented feature aware XGB model; further, effective cross validation scheme is presented to reduce classification error for building ensemble-based predictive model for EDM. Experiment outcome shows the FA-XGB based predictive achieves much better accuracy, precision, specificity, recall, and F-measure in comparison with XGB and ensemble-based predictive model. Future work would consider further minimize the error during cross validation process and effective predictive model considering multi-label classification scenarios.

## References

[1] Bithari, T. B., Thapa, S., & K.C., H. (2020). Predicting Academic Performance of Engineering Students Using Ensemble Method. Technical Journal, 2(1), 89–98.

[2] Tatar, A.E.; Dü¸stegör, D. Prediction of academic performance at undergraduate graduation: Course grades or grade point average? Appl. Sci. 2020, 10, 4967.

[3] A. Moubayed, M. Injadat, A. Shami, H. Lutfiyya, Student engagement level in e-learning environment: Clustering using k-means, Am. J. Dist. Educ. (2020).

[4] J. Yang, J. Ma, S.K. Howard, Usage profiling from mobile applications: A case study of online activity for Australian primary schools, Knowl.-Based Syst. (2019).

[5] J. Hung, B.E. Shelton, J. Yang, X. Du, Improving predictive modeling for at-risk student identification: A multistage approach, IEEE Trans. Learn. Technol. 12 (2) (2019) 148–157, 2911072.

[6] Tariq, Arham; Niaz, Yasir; Amin, Ahmad, “Systematic Approach for Re-Sampling and Prediction of Low Sample Educational Datasets”, University of Bahrain Scientific Journals, DOI: ISSN: 2210-142X, Date: 2021-08-03

[7] Alshanqiti, A.; Namoun, A. Predicting student performance and its influential factors using hybrid regression and multi-label classification. IEEE Access 2020, 8, 203827–203844.

[8] Rastrollo-Guerrero, J.L.; Gómez-Pulido, J.A.; Durán-Domínguez, A. Analyzing and predicting students’ performance by means of machine learning: A review. Appl. Sci. 2020, 10, 1042.

[9] Ofori, F.; Maina, E.; Gitonga, R. Using machine learning algorithms to predict students’ performance and improve learning outcome: A literature based review. J. Inf. Technol. 2020, 4, 33–55.

[10] Jain A, Solanki S (2019) An efficient approach for multiclass student performance prediction based upon machine learning. In: 2019 International conference on communication and electronics systems (ICCES). IEEE, pp 1457–1462.

[11] B. Sekeroglu, K. Dimililer, K. Tuncal, Student performance prediction and classification using machine learning algorithms, in: Proceedings of the 2019 8th International

Conference on Educational and Information Technology, in: ICEIT 2019, Association for Computing Machinery, New York, NY, USA, 2019, pp. 7–11.

[12] I. Khan, A. Al Sadiri, A.R. Ahmad, N. Jabeur, Tracking student performance in introductory programming by means of machine learning, in: 2019 4th MEC International Conference on Big Data and Smart City, ICBDSC, 2019, pp. 1–6.

[13] Shabnam Mohamed Aslam, Abdul Khader Jilani, Jabeen Sultana, Laila Almutairi, "Feature Evaluation of Emerging E-Learning Systems Using Machine Learning: An Extensive Survey," IEEE Access, pp:69573-69587, 2021.

[14] Fujita H et al (2019) Neural-fuzzy with representative sets for prediction of student performance. Appl Intell 49(1):172–187.

[15] Y. Liu, Z. Li, C. Zhou, Y. Jiang, J. Sun, M. Wang, X. He, Generative adversarial active learning for unsupervised outlier detection, IEEE Trans. Knowl. Data Eng. (2019) 1.

[16] Karaci A (2019) Intelligent tutoring system model based on fuzzy logic and constraint-based student model. Neural Computing and Applications 31(8):3619–3628.

[17] Shristi Shakya Khanal, P.W.C Prasad, Abeer Alsadoon, Angelika Maag, "A systematic review: machine learning based recommendation systems for e-learning," Springer, Education and Information Technologies, pp: 2635-2664, 2020.

[18] Alejandro Bogarin, Rebeca Cerazo, Cristobal Romero, "A survey on educational process mining," Overview, WIREs Data Mining Knowl Discov, pp:1-17, 2017

[19] Amita Dhankhar, Kamna Solanki, Sandeep Dalal, Omdev, "Predicting Students Performance Using Educational Data Mining and Learning Analytics: A Systematic Literature Review," Springer, Innovative Data Communication Technologies and Application, pp: 127-140, 2021.

[20] Namoun A, Alshanqiti A. Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. Applied Sciences. 2021; 11(1):237.

[21] MohammadNoor Injadat, Abdallah Moubayed, Ali Bou Nassif, Abdallah Shami, Systematic ensemble model selection approach for educational data mining, Knowledge-Based Systems, Volume 200, 2020, 105992, ISSN 0950-7051.

[22] Injadat, M., Moubayed, A., Nassif, A.B. et al. Multi-split optimized bagging ensemble model selection for multi-class educational data mining. Appl Intell 50, 4506–4528 (2020).

[23] Injadat M, Moubayed A, Nassif AB, Shami A (2020) Systematic ensemble model selection approach for educational datamining. Knowledge-Based Systems 200:105992.

[24] Moubayed A, Aqeeli E, Shami A (2020) Ensemble-based feature selection and classification model for DNS typo-squatting detection. In: 33rd Canadian conference on electrical and computer engineering (CCECE'20). IEEE, pp 1–6.

[25] Sapiah Sakri and Ala Saleh Alluhaidan, "RHEM: A Robust Hybrid Ensemble Model for Students' Performance Assessment on Cloud Computing Course" International Journal of Advanced Computer Science and Applications(IJACSA), 11(11), 2020.

[26] Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System", KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Pages 785–794, 2016.

[27] M. Vahdat, L. Oneto, D. Anguita, M. Funk, M. Rauterberg, A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator, in: Design for Teaching and Learning in a Networked World, Springer International Publishing, Cham, 2015, pp. 352–366.