The Application of Gaussian Mixture Models for the Identification of At-Risk Learners in Massive Open Online Courses

Raghad Alshabandar
Department of Computer Science
Liverpool John Moores University
Liverpool, United Kingdom
R.N.AlShabandar@2013.ljmu.ac.uk

Abir Jaafar Hussain
Department of Computer Science
Liverpool John Moores University
Liverpool, United Kingdom
A.Hussain@ljmu.ac.uk

Robert Keight
Department of Computer Science
Liverpool John Moores University
Liverpool, United Kingdom
R.Keight@2015.ljmu.ac.uk

Andy Laws
Department of Computer Science
Liverpool John Moores University
Liverpool, United Kingdom
A.Laws@ljmu.ac.uk

Dr Thar Baker
Department of Computer Science
Liverpool John Moores University
Liverpool, United Kingdom
T.M.Shamsa@ljmu.ac.uk

***Abstract*— With high learner withdrawal rates in the setting of MOOC platforms, the early identification of at risk student groups has become increasingly important. Although many prior studies consider the dropout issue in form of a sequence classification problem, such works address only a limited set of behavioral dynamics, typically recorded as sequance of weekly interval, neglecting important contextual factors such as assignment deadlines that may be important components of student latent engagement. In this paper we therefore aim to investigate the use of Gaussian Mixture Models for the incorporation such important dynamics, providing an analytical assessment of the influence of latent engagement on students and their subsequent risk of leaving the course. Additionally, linear regression and , k- nearest neighbors classifiers were used to provide a performance comparison. The features used in the study were constructed from student behavioral records, capturing activity over time, which were subsequently organized into six time intervals, corresponding to assignment submission dates. Results obtained from the classification procedure yielded an F1-Measure of 0.835 for the Gaussian Mixture Model, indicating that such an approach holds promise for the identification of at risk students within the MOOC setting.**

I. Introduction

With an increasing interest in open educational resources, Massive open online course (MOOCs) have become an area of continuing growth within both industry and academic settings[1]. In particular, MOOCs offers access to high-quality learning material for people around the world, for a nominal cost. However, despite the lowering of barriers to high-quality education, the ability of students to enrol and withdrew from courses freely often results in high rates of attrition[2][3]. As such, during 2012, the University of Duke offered a bioelectricity course, attracting around 12,175 registered participants, of which only 315 learners continued to undertake the final exam. At the end of 2012, It was reported that 93% of participants withdrew [2].

Identifying at-risk students with a high probability of premature withdrawal from courses has become of crucial importance, especially including the feedback of adequate information to the remote instructors, such that courses may be adapted to improve engagement [4]. Many studies have been conducted by researchers within which dropout prediction models have been proposed; such studies consider the learning behaviour record across various time intervals as key factor to infer student withdrawal [5][6]. Various behavioural features can be driven from behavioural records such as watching video, undertaking assignments, accessing the home page, and reading PDF documents [5][6]. For example, in the case of learner access to the home page of a course module, in current weeks they will continue to interact with the course in the next week. Otherwise, if the student fail to frequently click within the home page of the respective module, the probability of the student entering an at-risk status is increased [6].

Categorizing the latent engagement pattern of learners with respect to the impact on their continuation within course activities remains a challenge [6][7]. Few studies have been undertaken to investigate the latent engagement state as a sequential classification problem[7]. A notable limitation of the few existing studies is that behavioural features are distributed weekly. As a consequence, prediction models over time depend on a weekly basis without accounting for the assignment submission date as factor of significant influence for student withdrawal. Within this approach,the estimation procedures in characterize at-risk student could be inaccurate due to the failure to account for context-sensitive factors including the submission date of assignments.

The first contribution of this paper is to examine the link between the assignment submission date and dropout rate. In this work, Virtual Learning Environment (VLE) database has been used. The student's activity represented as a daily behavioural record, the behavioural features has been extracted according to assignments submission. With this method students records per each type of behavioural feature distributed across seven-time slices lead to, multi-view of behaviour features.

The second contribution of this paper is the use of mixture model to predict at-risk student. Mixture model is an effective probabilistic model that can be applied to infer the student latent engagement state over time.Mixture model provides advantages over traditional machine learning with the capability of automatically identify unlabeled data.

This reminder of this paper is organised as follows. Section 2 will give detailed information about prior works In section 3, shows the methodology, including data descriptions, feature extraction give overview of Gaussian Finite Mixture model, experiment setup and experiment results .The conclusion and future works are described in Section 4.

II. Literature Review

Student withdrawal is one of the issues of concern in MOOCs, prompting a number of proposed works in the literature to investigate the problem of student dropout in online environments. To understand why students withdraw from online courses, some researchers have focused on machine learning and statistical methods while others have considered interview transcripts and surveys [8]. In this section we will summarize prior works towards the analysis of learner withdrawal in MOOCs.

In [5][6] [9], the authors consider the student dropout issue in the form of a time series classification problem. In particular, clickstream data has been considered from which a series of features are extracted, such as the number of lecture videos viewed, the number of threads posted in online forums, and the number of attempted quizzes. The author in [9] employs logistic regression to predict student drop out events. The authors split the course into fifteen-time slices based on the weekly interval. The results show a best predictive performance of AUC 0.95, obtained in the middle of the course ,with the lowest AUC of 0.77 obtained at the end of course [9]. Hidden Markov Model (HMM) has been used in to predict the student retention over time[5]. In this work, the author proposed both a composite and individual HMM. The study reported satisfactory performance for the composite HMM, obtaining an AUC value of 0.71, for which multiple behavioural features were considered as a source of input. Subsequently, the individual HMM provided insight into patterns of student activity, for instance, participants who never check the course progress frequently are more likely to withdraw following the fourth week of the course [5]. In a further study, at-risk students within online course settings have been investigated by [6]. To understand the student motivation in relation to a particular activity, hidden latent engagement was analyzed through application of a Nonlinear State Space Model (NSSM) [6]. NSSM model was compared empirically with various other models, namely logistic regression (LG), simultaneously smoothed logistic regression (LR-SIM) and long-short-term memory (LSTM). Experimental results indicated that NSSM showed an AUC value of 0.9 at the beginning of course. In contrast, the lowest AUC appeared at the end of the course with an AUC of 0.7. The results showed that latent engagement patterns under analysis are time varying in nature[6].

The Recurrent Neural Network (RNN) in conjunction with Short-Term (LSTM) has been applied by [10]to predict student dropout in two MOOCs platform Coursera and edX.The results show that LSTM is the best classifier that capable to discover the nonlinear latent representation of the model.The researchers in reference [11]propose a multi-view semi-supervised learning model to address the issue of dropout prediction problem.With this approach the unlabeled data was driven from student behaviour record that could improve prediction performance of inadequate label.In this study, six behavioural features were considered, these features including undertook assignments, watching videos, access other objects, post forum, and closing web page.Four type of classifier used to train each feature separately.The finding reveals that access other objects attribute is most effective features into withdrawal rate.The an average value of F-measure around 83%-84% for all classifiers[11]

Data from discussion threads are used by [2] to measure the impact of social networking on student withdrawal rates. Survival models were defined in the study as predictive models using logistic regression to infer the likelihood of learners' survival during the course. The results reported that social factors significantly impact the dropout rate and could be used in the conception of MOOCs more suitable for social engagement, in turn decreasing the attrition rate[2].

III. Methodology

A. Data Description

In this study, a database has been obtained from the Open University, UK. The Open University delivers various online courses for undergraduate and postgraduate students. During 2013-2014, the Open University released a dashboard known as the Open University Learning Analytic Dataset (OULAD) [12]. Two kinds of features have been considered in the database, namely demographic and behavioural. We consider ("BBB") course that launched in two semesters, mainly in October during 2013.

The database is structured according to a relational schema, where all tables are joined to form a central composite table. The central table is designated "studentInfo", containing information relating to student demographic features such as gender, age, and geographic area.

In particular, the database contains fields related to student performance and assignments, in addition to student interaction with online courses. In terms of behavioural features, a Virtual Learning Environment (VLE) system was used to capture student interaction within the online course setting. Each VLE is represented as an activity type, indicating the type of learning resources that students are required to engage with each module. There are various types of learning resources such as reading pdf files, access to home and subpage, and undertake quiz [12].

The table “StudentVle” includes information regarding to student activities in particular modules. A series of student activities are collected on a daily basis and recorded in this table. The database captures daily information relating to student behaviour within an online course, in addition to the number of clicks that correspond to students interaction with learning material on each day. The students are identified within both the VLE and studentInfo tables through unique numbers, providing consistent access to records [12].

1) Feature Extraction

From the student VLE data capture, a daily feature extraction procedure has been undertaken. The VLE features were extracted according to the assignment submission dates. The course is split into seven time slices, where each time slice mapping is oriented around the final date of the assignment submission. However, the first time slice captures student VLE information relating to learning activity prior to the course start, since students are permitted to enrol prior to official course commencement. Our analysis of VLE features is undertaken to examine the association between student performances with the measured beavoural features with respect to assignments submission dates. For each student there are a number of VLE learning activities at specific time *t*. We aggregate the VLE activity type for each student per time interval into single values. Hence, at each time interval, the students VLE information records include the VLE activity type in addition to two extended features, namely the number of sessions (o^t) and total number of clicks(c^t).The procdure of feature extraction is desribed in algorthim 1

ALGORTHIM1. FEATURE EXTRACTION PROCEDURE

```
Split course into 7 interval  t where t ∈ [1,7]
  For each student  $S_i$, i= 1…n, do
    For each course activity type $X_j$,, i = 1…m, do
      For each session $O_p$per activity type $X_j$, p = 1…w, do
            $O^t = \sum_{p=1}^{W} O_p$
            for each sum_click $C_e per$  activity type $X_{j,e}$ = 1…c, do
        $C^t = \sum_{e=1}^{c} C_e$
          End
        End
End
```

2) Features

Static Behavioural features: It is complete set of behavioural features**.**Let the tensor $X \in \mathbb{R}^{T \times n \times m}$, in which $X_{t,i,j}$ represents the *j*th activity of the *i*th student at time *t*. S is a set of students denoted as an n-dimensional vector [S_1 ...., *Sn*], where n is the number of students. Furthermore, M is defined as an m-dimensional vector that represents VLE learning activity types, M = [M_1 ...., M_m], where *m* is the number of learning activities that the *i*th student is assigned.

Dynamic Behavioural Features : It is set of behavioural features that varying over time .Let *T* be a sequence of disjoint time intervals, where $t \in [1,6]$. To represent all student activities at time *t*, we define the the type of students activity records as the vector $X_{t,i,j} = [X_{t,i,1}, X_{t,i,2,} \dots, X_{t,i,m,}]$. Here the *j*th denoted learning activity is undertaken at time *t* by student S_i, such that *j* = 1, … , *m*; where *m* is given as the number of learning activities.

Demographic Features: The demographic attributes, given as $G \in \mathbb{R}^{n \times L}$, in which $G_{i,k}$ represents the *k*th demographic feature for the *i*th student, where the set of demographic features assigned to each student are considered constant over the course duration. The demographic features for the *i*th student may therefore be given by the *L* dimensional vector $G_{i,k} = [G_{i,1}, G_{i,2}, \dots, G_{i,L}]$, where *k* = 1, …, *L*.

3) Target Class

The students should participate in five CMA assignments and six TMA assignments.The assignment submission dates were measured by counting the number of days since the respective course start. Overall, in both semesters, assignments should be submitted within the same period.

To present the set of assigments submitted by student S_i at time *t*, we define the vector $A_{t,i,a} = [A_{t,i,1}, A_{t,i,2,} \dots, A_{t,i,a}]$. Here, $A_{t,j,a}$ denoted the a^{th} assignments undertaken by student S_j at time t. Additionally, $t \in [1,6]$ indexes the course time intervals within each of which students are allocated a single assignment in addition to the last interval where a final exam is assigned.

The target class at-risk student is derived from students assignment which can be represented as $Y^t(i)$vector where $Y^t(i) \in \{0,1\}$, if student S_i undertook a^{th}assignment at time t , $Y^t(i)$ denoting 0 ,and 1 otherwise.

B. Gaussian Finite Mixture Model

Mixture model is a probabilistic model that infers groups of observation within a population without prior knowledge of subgroup memberships [13]. Mixture model has been widely applied in various domains such as statistical inference, machine learning, clustering, classification, and hidden variable modelling [13].The estimation of the parameter is based on cluster analysis, where the components represent a probability distribution across cluster memberships.

Different approaches have been used in the literature to determine the number of clusters; such approaches can be classified into three categories, namely stochastic and deterministic [14]. In stochastic approaches, Markov Chain Monte Carlo (MCMC) method is employed. The deterministic approaches can be categorized into two main categories. In the first category, Bayesian criteria are employed such as the Bayesian Information Criterion (BIC) and Laplace Empirical Criterion (LEC) [15]. In the second category, coding theory is considered for selecting the number of clusters, for instance using Minimum Message Length (MML) and Akaike's Information Criterion (AIC)[16].

Gaussian finite mixture model is a popular type of mixture model. The key features of this approach is the capacity to model complex data by mixing the properties of a density function of sub populations into finite mixtures of components. In the finite Gaussian mixture the Bayesian Information Criterion (BIC) and Integrated Complete-Data likelihood (ICL) criterion are used to determine the number of clusters [13].

Let X=$\{X_1, \ldots. X_n\}$ a sample of n univariate observations. The probability of X_i can be derived from the probability density function (PDF) as follows [13]

$$\text{P}\ [X_i] = \int_a^b p(x)\ dx \qquad (1)$$

In mixture models we assume observations are denoted by $X_n = (X_1, \ldots, X_n)$, where each observation belongs to g components. The empirical estimate of the PDF of X_i can be computed as [17]

$$f\ (X_i) = \sum_{g=1}^{G} \mathcal{T}_g\, f_g(X_i; \theta_g) \qquad (2)$$

Where G is number of components and $\mathcal{T}_g$ is mixing weight of observation X_i associated with components of the g_{th} $(\sum \mathcal{T}_g = 1; \mathcal{T}_g > 0)$. $f_g(X_i; \theta_g)$ is the density of g_{th} component with estimated parameter θ_g in mixture model.

If the observation data follows a normal distribution, the Gaussian density function is considered to characterise the finite mixture model (FMD). In this case, within each cluster, the data is centred by the mean μ_g and the covariance$\sum_g$. The density of observation X_i takes the following form [18].

$$f\ (X_i) = \sum_{g=1}^{G} \mathcal{T}_g\ \theta_g\ (X_i|\mu_g, \sum_g) \qquad (3)$$

The covariance $\sum_g$ is used to specify the Geometric characteristics {shape, volume, orientation} of each cluster. Reference [18] applies constraints on the covariance $\sum_g$ to represent the various models of elliptical clusters. The authors proposed the eigenvalue decomposition framework. The eigenvalue decomposition can be describe as follows [18].

$$\sum_g = \lambda_g\ D_g A_g D_g^T \qquad (4)$$

Where D_g is an orthogonal matrix and A_g is a diagonal matrix. The D_g, A_g parameters control the shape and orientation of g_{th} components in mixture model while λ_g is constant which governs the volume of the g_{th} components.

1) *Mixture Discriminant Analysis (MDA)*

The Mixture Discriminant Analysis (MDA) is a predictive model used for supervised classification problem based on mixture Model. The model aims to assign the observation data belonging to the unknown class to one of the true classes. The density of each class in MDA model follows a finite Gaussian mixture distribution. The MDA can be described according to [17], formally defined as:

$$f\,(X_c) = \sum_{g=1}^{G_c} \mathcal{T}_{gc}\ \theta\ (X|\mu_{gc}, \sum_{gc}) \qquad (5)$$

Where $\mathcal{T}_{gc}$ is the mixing weight of class c associated with the g_{th} component, such that $(\sum \mathcal{T}_{gc} = 1; \mathcal{T}_{gc} > 0)$. Accordingly, $\mu_{gc}, \sum_{gc}$ represent the mean and covariance of components g for class c respectively.

The authors in [19] develop a MDA model which assumes the number of components associated with each class is known and the covariance matrix within each class is similar. In another study, Eigenvalue Decomposition Discriminant Analysis (EDDA) has been proposed by [20], assuming that each class belongs to a single Gaussian component.

The Expectation-Maximization algorithm is typically used to estimate the model parameters in EDDA [20]; The EM algorithm consists of two steps, namely Expectation (E) step and Maximization (M) step [19] [21]. During the E step, the conditional probability that an observation x_i associates with the g_{th} component is computed. In the subsequent M step, further parameter estimates are computed to maximize the expected log-likelihood obtained during the E step. The estimated parameters are then used to initiate further E-M steps iteratively until convergence. The ML procedure therefore continues until all observations are assigned to a cluster corresponding to the highest posterior probability [19] [21].

C. *Experiment setup*

The method applied in this study follows a binary classification problem. Accordingly, an EDDA model is designed and evaluated for the prediction of student risk within course environments, as considered within the hard categories of “at risk”, or “not at risk”. The supervised classification algorithms did not take into consideration the impact of unlabeled data on one of class [22]. EDDA capable autonomously discover unobserved latent engagement and assign these unlabeled data to one of the classes. The mixture model is powerful inference framework can approximate represented high dimensional data as a linear combination of multiple Gaussian components [22][23].

A total of 22 behavioural features are extracted in this study, denoted as the n-dimensional vector $X^t(i)$, producing a sample indexed over each student S_i , per each completed time interval *t*, where t ∈ [1,6]. Subsequently, we split data into 60% for use with model training and 40% for test evaluation. We consider the training dataset as the complete form of the variables, consisting of a set observations, denoted$X^t(i)$, and a set of latent variables. The latent variables can be represented by $L^t(i)$, whose unknown labels can be described as $Z^t(i) = [Z_{t,i,1}, Z_{t,i,2,} \dots, Z_{t,i,m,}]$, such that $Z^t(i) \in \{0,1\}$. The class label for the i^{th} observation, $X^t(i)$, is given as $Y^t(i) \in \{0,1\}$. For example, if the i^{th} student submits the a^{th} assignment at the current time interval *t*, and his previous latent state at time *t*-1 was active, then $Y^t(i)$ = 0, else 1.

To includes all information about the learners past behaviour.At each interval, we combine the student behaviour feature at time *t* with student learning behaviour at the pervious time *t-1*. The Algorithm (2) describe the learning procedure of EDDA model per each interval.

```
Algorithm (2) Learning Procedure by mixture model

Given the incomplete training dataset $\{(X_{1,}Y_{1,}), \dots (X_{N,}Y_{C,})\}$
For each $X_i$,i=1 … N do
 Initialize model parameters $\theta$
 Initialize the hidden engagement state Z
 Repeat
   Procedure E-Step
      Use the estimate parameters: $\widehat{\mathcal{T}_g}$ ,$\widehat{\mu_g}$,$\widehat{\Sigma_g}$
        Compute the initial expected value of latent engagement
      $\widehat{Z_{ig}}$by using Eqn. 6
        End procedure
   Procedure M-Step
  Update expected value of latent engagement  $Z_{ig}$ via Eqn. 7
   and Eqn. 8 respectively
   End procedure
Until converge
End For
```

$$\overline{Z_{ig}} = \frac{\widehat{\mathcal{T}}_g\theta(L_i|\widehat{\mu_g},\widehat{\Sigma_g})}{\sum_{g=1}^{G}\widehat{\mathcal{T}}_g\theta(L_i|\widehat{\mu_g},\widehat{\Sigma_g})} \qquad (6)$$

$$L(\theta_{g,}\mathcal{T}_g,\Sigma_{g,}|L_{ig})=\sum_{i=1}^{n}\log\left[\sum_{g=1}^{G}\mathcal{T}_{g\theta\,(L_i|\mu_g,\Sigma_g)}\right] \qquad (7)$$

$$Z_{ig} \begin{cases} 1\ if\ L_i \text{ belong to group g} \\ 0\ otherwise \end{cases} \qquad (8)$$

To gain maximum like hood for initial estimates parameters ($\widehat{\mathcal{T}_g}$,$\widehat{\mu_g}$,$\widehat{\Sigma_g}$).The function L($\theta_{g,}\mathcal{T}_g,\Sigma_{g,}|L_{ig}$) is applied. This estimates are calculated consider latent variableL_i.

The expected value of latent variable $Z_{ig}$are update until latent variables L_i assign to component g that matching the highest probability.

D. Result Evaluation

Model responses from the experimental procedure were obtained for the classifiers under study, namely the EDDA, KNN, and LR models, designed within the context of a discrete binary outcome. Ground truth labels were defined as "Withdrawal" and "No withdrawal" respectively, with results listed in Table 1. The results indicate that the number of estimated parameters vary significantly over time. However, a similar number of latent variables were found to be present among intervals 2, 3, 4. A possible interpretation is that the student's hidden state engagement could vary at the beginning and end of the course while in the middle of the course stable.

The value of BIC also differs across each time interval. The BIC select EEE (Equal volume, shape and orientation) model with one component per each class for intervals 1, 2, 3, and 4 respectively. The highest value achieved at interval 1 with a value of -26709.2 as a result, the strongest and best model obtain at interval 1 while weakness EDDA acquired by interval 3. The optimised value for interval 5 and interval 6 are similar. BIC chose the EEV (same size and shape but different orientation).The values of BIC in such models are -721515 and -30341.42 respectively.

To evaluate the performance of EDDA over the set of test data, both ROC analysis and confusion matrix values were computed, forming the basis for comparing model responses to ground truth labels, over each model. Various performance summary metrics are considered, namely accuracy, sensitivity, specificity, the F1 measure, and the AUC. Table 2 lists the results of the EDDA, k-nearest neighbors (KNN), and logistic regression (LG) classifiers over each of the six time intervals.

In particular the metrics sensitivity, specificity, and F1 - Measure are suitable summary computations in the presence of class imbalanced data, as is the case in the current study. In general, unbalanced data can result in misclassification through biased selection of the majority class. The F-score, family of metrics in particular consider both precision and recall to expose the presence of class-relative error types, which are not apparent from computing the accuracy alone.

The empirical results show that the highest F1-Measure is acquired by the LG classifier follwed by EDDA, with values of 0.85 0.83 across interval 4 respectively. In intervals 6, 2, 3 and 1, the EDDA model achived the highest F1-Measure, obtaining average values of 0.75, 0.70, 0.68 and 0.40. Conversely, the KNN classifier obtained a moderate to good range of accuracy with F1-Measure values of 0.60-0.85 across all intervals. The lowest F1-Measure is seen at interval 1 resulting from the kNN classifier, yielding a value of 0.35

Considering the sensitivity outcomes, the highest value is achieved by the EDDA classifier at interval 4 with a value of

0.957. Although the highest sensitivity is achieved by the EDDA model compared with the other classifiers over intervals 3, 6, and 2, with average values of 0.94-0.92, the EDDA model acquired low sensitivity over interval 5. The LG model achieved the highest sensitivity at intervals 1 and 5, yielding values of 0.86 and 0.66, respectively.

All classifier models obtained viable specificity values over the intervals 2, 3, 4 and 6, ranging over values 0.94 to 0.91. However, the KNN and LR models present slightly lower specificity with a value of 0.90. The lowest range of specificity is obtained by LR at interval 1 with a true negative percentage of 70%.

ROC is used in this study to select a decision threshold value for the trade-off between true and false positive rates across each time interval. Figure 1 lists the ROC curves respectively. Overall, the range of AUC values falls within 0.97-0.90 for intervals 2, 3, 4 and 6 respectively. Conversely, intervals 1 and 5 acquired the lowest AUC values, yielding 0.85-0.73 respectively.

E. Discussion

The simulation result from the experiment is demonstrated that latent engagement has a significant impact on student withdrawal within MOOCS environment. The estimation of the hidden engagement state could help to identify at-risk students early. Hidden engagement state could not be inferred by using shallow machine learning method; As this, model could lead to misclassification. The output result of KNN and LG could be inaccurate.

The empirical results reveal that best sensitivity acquired by EDDA model in the middle and the end of course while EDDA achieves low sensitivity by interval 5. Due to the number of unobserved variables approximate 58% higher by interval 5 than intervals 2, 3, 4. This implication of increasing the number of hidden state could impact the true positive rate within a particular interval. As a result, uncovered variable could lead to misclassification problem.

In spite, Overall results show that the incorporate hidden variable within the classifier could enhance the sensitivity and specificity model compared with LR linear model.The EDDA showed an advantage over LR and Knn as EDDA model, classify the target by using Bayes rule to compute the joint probability of observation features with hidden feature while LR and KNN predict class label directly.

TABLE 1. EDDA MODEL RESULT

Interval	**No. Training examples.**	**No. estimated parameter**	**Log. likelihood**	**BIC**	**EDDA subtype**
Interval 1	1582	1178	-9384.088	-26709.2	EEE
Interval 2	2107	1173	-61417.71	-121812-4	EEE
Interval 3	2058	1173	-24363.37	-57676.14	EEE
Interval 4	2380	1173	-125269.6	-259659	EEE
Interval 5	2037	2024	-353046.9	-721515	EEV
Interval 6	2079	2024	-7439.392	-30341.42	EEV

TABLE 2. Classification Performances for EDDA Model

Interval	**Classifies**	**Acc.**	**F1**	**Sens.**	**Spec.**	**AUC**
Interval 1	EDDA	0.807	0.405	0.674	0.821	0.769
	KNN	0.785	0.351	0.626	0.801	0.760
	LR	0.718	0.362	0.865	0.702	0.831
Interval 2	EDDA	0.932	0.709	0.928	0.932	0.977
	KNN	0.938	0.697	0.875	0.943	0.963
	LR	0.935	0.690	0.895	0.938	0.972
Interval 3	EDDA	0.923	0.686	0.941	0.929	0.949
	KNN	0.937	0.685	0.826	0.947	0.937
	LG	0.929	0.671	0.895	0.931	0.954
Interval 4	EDDA	0.926	0.835	0.957	0.919	0.899
	KNN	0.922	0.830	0.895	0.929	0.929
	LG	0.930	0.854	0.952	0.922	0.943
Interval 5	EDDA	0.620	0.465	0.568	0.830	0.730
	KNN	0.829	0.567	0.560	0.896	0.791
	LG	0.859	0.631	0.661	0.903	0.855
Interval 6	EDDA	0.920	0.758	0.935	0.918	0.954
	KNN	0.911	0.621	0.550	0.966	0.951
	LG	0.893	0.491	0.385	0.971	0.949

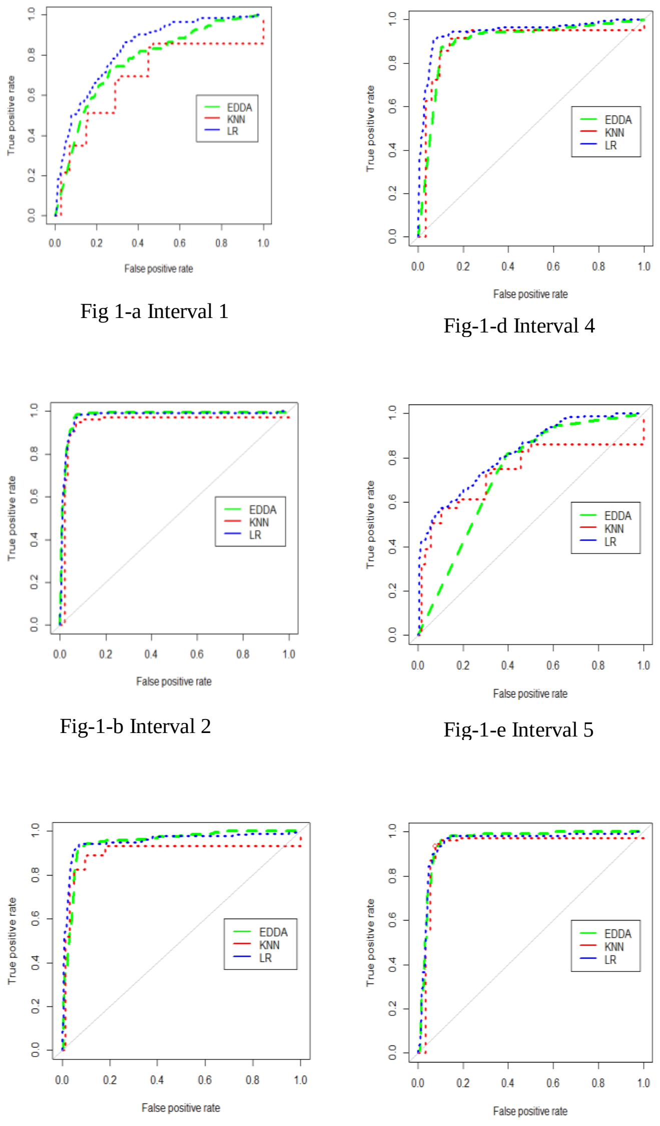

Fig 1-a Interval 1

Fig-1-d Interval 4

Fig-1-b Interval 2

Fig-1-e Interval 5

Fig-1-c Interval 3

Fig-1-f Interval 6

Fig .1. ROC Curve across 6 Time Intervals

IV. Conclusion

A study has been conducted to predict at-risk learners within MOOCs environment. There is 22 VLE activity type. For each activity type, two behavioral features have been extracted: the number of sessions (o^t) and the total number of clicks (c^t).we aggregate daily extracted features for each student per each activity type into single value based assignments cutoff date.

Our objective is to investigate the student latent engagement state and examine how could influence at-risk students for that purpose; the mixture model is utilized to identify at-risk learners. The results reveal that EDDA model capable of inferring latent characteristics of the students. The number of latent variables increase about 75% at the end of the course as a consequence, the latent engagement varying over time, however, is still stable in the middle of the course.

The simulation results of our experiment shown that EDDA by interval 4 achieves the height performance In intervals 6,2, 3 and 1 in contrast, the lowest performance of EDDA acquired by interval 5. The finding suggests that latent engagement has impact on student decision to persist or withdrew from the course. The task difficulty can be considered as main factor that could affect student latent engagement state.

References

[1] J. Qiu *et al.*, “Modeling and Predicting Learning Behavior in MOOCs,” *Proc. Ninth ACM Int. Conf. Web Search Data Min.*, pp. 93–102, 2016.

[2] D. Yang and C. P. Rose, “‘ Turn on , Tune in , Drop out ’: Anticipating student dropouts in Massive Open Online Courses,” pp. 1–8.

[3] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, “Predicting MOOC Dropout over Weeks Using Machine Learning Methods,” *Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.*, pp. 60–65, 2014.

[4] J. He, J. Bailey, and B. I. P. Rubinstein, “Identifying At-Risk Students in Massive Open Online Courses,” *Proc. 29th AAAI Conf. Artif. Intell.*, pp. 1749–1755, 2015.

[5] G. Balakrishnan and D. Coetzee, “Predicting student retention in massive open online courses using hidden markov models,” *Electr. Eng.*, 2013.

[6] F. Wang and L. Chen, “A Nonlinear State Space Model for Identifying At-Risk Students in Open Online Courses,” *Proc. 9th Int. Conf. Educ. Data Min.*, pp. 527–532, 2016.

[7] A. S. Lan, C. G. Brinton, T. Yang, and M. Chiang, “Behavior-Based Latent Variable Model for Learner Engagement,” pp. 64–71, 2016.

[8] S. M. Glynn, P. Brickman, N. Armstrong, and G. Taasoobshirazi, “Science motivation questionnaire II: Validation with science majors and nonscience majors,” *J. Res. Sci. Teach.*, vol. 48, no. 10, pp. 1159–1176, 2011.

[9] C. Taylor, K. Veeramachaneni, and U.-M. O’Reilly, “Likely to stop? Predicting Stopout in Massive Open Online Courses,” 2014.

[10] M. Fei and D.-Y. Yeung, “Temporal Models for Predicting Student Dropout in Massive Open Online Courses,” *2015 IEEE Int. Conf. Data Min. Work.*, pp. 256–263, 2015.

[11] W. Li, M. Gao, H. Li, Q. Xiong, J. Wen, and Z. Wu, “Dropout prediction in MOOCs using behavior

features and multi-view semi-supervised learning," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016–Octob, pp. 3130–3137, 2016.

[12] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics Dataset," *Proc. Data Lit. Learn. Anal. Work. LAK 2016 Conf.*, 2016.

[13] X. Bold and X. This, "Chapter 2 Finite Mixture Models," no. x, pp. 9–19.

[14] N. Bouguila and D. Ziou, "[2007][10] High-Dimensional Unsupervised Selection and Estimation of a Finite Generalized Dirichlet Mixture Model Based on Minimum Message Length.pdf," vol. 29, no. 10, pp. 1716–1731, 2007.

[15] C. Fraley and A. E. Raftery, "Bayesian regularization for normal mixture estimation and model-based clustering," *J. Classif.*, vol. 24, no. 2, pp. 155–181, 2007.

[16] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 24, no. 3, pp. 381–396, 2002.

[17] M. Fop, T. B. Murphy, and A. E. Raftery, "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models," vol. XX, no. August, pp. 1–29, 2016.

[18] N. Russell, L. Cribbin, and T. B. Murphy, "upclass : An R Package for Updating Model-Based Classification Rules," *Cran.R-Project.Org*, no. May 2014, 2006.

[19] TREVOR HASTIE and ROBBERT TIBSHIRANI, "Discriminant Analysis by Gaussian Mixtures," 1996.

[20] H. Bensmail *et al.*, "Regularized Gaussian Discriminant Analysis Through Eigenvalue Decomposition Stable URL : http://www.jstor.org/stable/2291604 Linked references are available on JSTOR for this article : Regularized Gaussian Discriminant Analysis Through Eigenvalue Decompos," vol. 91, no. 436, pp. 1743–1748, 1996.

[21] C. Fraley and a E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *J. Am. Stat. Assoc.*, vol. 97, no. 458, pp. 611–631, 2002.

[22] C. Bouveyron, "Adaptive Mixture Discriminant Analysis for Supervised Learning with Unobserved Classes," *J. Classif.*, vol. 31, no. 1, pp. 49–84, 2014.

[23] W. W. Moe and P. S. Fader, "Capturing evolving visit behavior in clickstream data," *J. Interact. Mark.*, vol. 18, no. 1, pp. 5–19, 2004.