ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 5, May 2021

Mooc Student Dropout Prediction Using Machine Learning Algorithms

[1] Vivek Sasikumar Iyer, [2] Siddharth Chaudhury, [3] Dr.M.Shobana

[1][2] SRM University, CSE Dept., Chennai, India
[3] Assistant Professor, SRM University, CSE Dept., Chennai, India
Email: [1] vs3877@srmist.edu.in, [2] su7255@srmist.edu.in

Abstract---- **MOOC is an ultimate way to give educational content in higher studies settings by providing good-quality educational material to the studentts throughout the world. If we take in the differences between traditional learning paradigm and MOOCs, a trend focusing on predicting and explaining dropping out of students and low completion rates in MOOCs has emerged. The quantity of students taking these courses is very high, still, the completion rate is very low. Factors affecting the completion of a course by a student such as interest in the subject, reason for taking the course, if the teacher is able properly make students understand or not. Nevertheless, because of varying problems specifications and evaluation metrics, undertaking a high level evaluation of state-of-the-art machine learning architectures isn't easy. This paper provides a complete rundownn of MOOCs student's dropout probability with the help of machine learning techniques. Moreover, we have highlyghted a few answers being used to solve the dropout problem, provide an examinationof the challenge of prediction models, and give some important overview and suggestions that may pave the way to develop useful ML solutions for overcoming the MOOCs dropout problem. Essentially, we got features from every learning behaviors and formed multi-view behavior features. Also, examining these features, we proposed a new multi-view semi-supervised learning model to use a large number of raw data to help inadequate labeled data to improve prediction. We have used KDD Cup 2015 dataset, from the results we see that our proposed model attains better prediction of a student leaving a course than state-of-the-art methods. Long Short-Term Memory neural network (LSTM) prediction model makes use of time-control units, the unit has the ability to model early learning behaveiour with varying time intervals. Formulated from the LSTM model, we designed time-controlled gates for capturing a good long- and short term info and use the learning process info toget better forecast performance**

Keywords--- **MOOCs, Dropout Prediction, dropout problem, completion, Long Short-Term Memory, LSTM**

I. INTRODUCTION

The Massive Open Online Courses (MOOC) is an trendy classroom method that came into light under the rise of artificial intelligence and network IT platform tech. The invention of this method challenges old teaching methods. It has created an unparalleled network university, which revolutionises "big data, big changes, big arguments". From conventional univ learning to big data education, higher studies has overcome more resource of data. Hence, its unavoidable to execute talents trainings, learning analysis, curriculum research, teaching method and work based on big data. These tasks provide better benefits in mastering teaching progress, impact, and so on. Paralally, the advent of online course platforms such as MOOC is of huge opposetion for the usual education method. MOOCs brings a lot of benefits that traditional teaching methods don't, still many defects and issues need to be tackled. Additionally, students get access in the click of a finger to good education whatever be the place and academic background. These advantages have gathered a big quantity of learners to take these courses. MOOCs supply learners with quality, proper and reliable educational materials and learning opportunities over the globe, and make education at your convenience. Thus, the big-scale open network course will gain more popularity in coming times. By changing educational method, a lot of challenges need to be tackled in MOOCs. Be it the construction of the curriculum system and the network of curriculum knowledge systems, or the surveyllance of learner's behaveoural patterns, also adjusting the teaching process, are all major problems that MOOCs must tackle.

From last 5 years, MOOCs evolved very fast and is becoming moreover ambitious. Around 34,000 learners have undertaken such courses. Anyhow, only 7.5% learners finished their courses, and one of the reasons for the dropout was because of improper time management skills. Data mining technologies was used to survey students' learning state, thereby, the learning results could be realised and predicted beforehand. An undisputed truth is that MOOCs usually get a huge dropout rate . The efficient way to solve the issue would be to make accurate predictions based on learning behavior. Students usually emit varying traits in learning. Researchers make predictions taking such traits, and then perform punctual and efficient teaching interventions, which could help learners to stay in the course. The enrolment ratio is generally positively related with educational achievements, when we give some information to learners in the start of learning, it likely becomes a useful tool for improving learners' state. Thus, its very crucial to make an accurate prediction of the students' learning results and dropout rate. We analysed the different learning behaveiours of students to do some survey, which will predict if the student will finish a learning task, classifying various types of students early and point out potential dropouts. This experiment's result shows that the survey of the traits of the learner's behaviour will more efficiently predict if the student will complete the learning task or dropout.

II. RELATED WORK

MOOCs is an online education system developed by Stanford University in August 2011 and collaborated by a lot of universities. This programme supplies 1,800 courses teaching many subjects including chemistry, psychology, Science, maths, and computer, which provides learning opportunities for students from around the globe. Paralally, students can learn anytime, anywhere, without consideration of time and place constraints. There's a huge quantity of multidisciplinary courses in MOOCs, which is advantageous for gaining knowledge and improving your technique. But, differentiating from on-campus courses, the dropout rate of MOOCs has been increasing by a huge margin and the completion rate is very low. Since MOOCs records the operational info of every learning process of learners, it becomes a very efficient testing basis for MOOCs dropping out with data mining research, with the help of which we are able to deeply analyse the reasons affecting the dropout rate of learners. A connective view of the factors for this problem of dropout is the low amount of participation, but a learner's participation is very difficult to calculate. In most papers, the lack of participation of students is showed as the lack of interaction with MOOCs, eg, submission of tasks and tests, watching video lecture, and taking part in discussion forums. The researchers believe that the level of interconnectivity with the MOOCs classes affects the student's achievement proportionally and chances of dropping out of course. One more opinion is that MOOCs students lack the will to continuously learn and understand that even if they interrupt learning, it won't affect their learning results. Parallaly, there aren't proper ways to incurporate learning. Thus, less learners will be insistent on finishing the course. Generally, there are a large amount of people who register for the MOOCs, but there are lesser people who seriously think about taking part in the course, and those we finish the course are few. The authors decided that learning task, social interaction, learning legitimacy, evaluation techniques, and learning expectations are the main reasons that lead to high dropout rates. In light of the differrences between MOOCd and traditional learning, consideration of the needs of distance students, designers should take measures such as preliminary tests, big data study of learning patterns, doubt clearing sessions at different stages of the process of MOOCs to improve the quality of MOOCs teaching, and prevent dropout of students. The machine learning study of the MOOCs data is also a crucial part. Kloft and Amnueypornsakul predicted the dropout pattern by taking the clickstream data into the machine learning algorithm prediction model, extracting various feature vectors for predicting the student's behaviour. Amnueypornsakul made a sequence of learning behaveiours for students, such as browsing courses, taking exams, and submitting exams. They built 6 models, and results implied that for extensive users, the prediction accuracy rate is relatively high.

For users with less behaviour, the sequence is scanty and the accuracy rate is low. Sharkey continuously searched for predictive features and compared various methods. He found that machine learning models were above average in predicting dropouts, and predictions with variable were consistent with learner participation. A study of Yang showed that social reasons did have a say on dropouts. Sinha used data like click-through rate to model various sequences of learning events, trying to find sequence that are more appealing to learners. Taylor took in varying machine learning techniques for predicting dropout percentage, using logistic regression, support vector machines, deep belief networks, decision trees, and hidden Markov models. He screened the learner's 14-week learning pattern to train and test. The results show that people who are often involved in MOOCs learning can provide better prediction result, and the predictivity of features relating to submission of tasks and exams is high, the length of posting is more predictive than the number of postings, collaborative social related features, like wiki and forum, seem important in forecasting.

III. LITERATURE SURVEY

Because of the fame of the MOOCs as an alternative to get knowledge with their convenience of time and place, huge amount of learners take part in these courses and the MOOCs provider keeps updates of the activeities of the

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 5, May 2021

learners in log files. Due to its fame many researchers are interested in studying these data and get some hidden info from it.

Different features of the MOOCs such as task grade, certifyication, using linear regression and SVM, J. Qiu et al. conducted a research to separate the science which includes students from computer science, electronics engineer background, and non-science students which includes students from sociology, contemporary, arts and different backgrounds. The dataset they used was arranged conisdering the academic background of the learners, quantity of posts on the forums and differrent activities of learners in the MOOCs which was taken by the MOOCs provider. Researchers used various classification techniques to predicate the task grade and the certifyicate gaining of the student. The differrent models used are SVM, Linear regression, Factorization machine, Latent Dynamic factor graph are used on the dataset to get the views of the performance on grounds of behavioural features of dataset. Figure 1 shows the prediction of different methods applied to the dataset (a) shows the assignment grade prediction using different models and (b) shows the certificate earner prediction using different models. Again, atempts have been made by the researchers for finding the learning pattern of the students using regression model, Ordinary least square (OLS) estimates the interaction between the student and the forum activeities. The students with higher forum activeities have more chances to obtain the certification and gain better assignment grades than their counterparts. Table I compares the performance of different methods used in predicting the certifyication earner.

Category	*Method*	*AUC*	*Precision*	*Recall*	*F1-Score*
Science	LRC	92.13	83.33	46.51	59.70
	SVM	92.67	52.17	83.72	64.29
	FM	94.48	61.54	74.42	67.37
	LadFG	95.73	73.91	79.07	76.40
Non-Science	LRC	94.16	76.93	89.20	82.57
	SVM	93.94	76.96	88.60	82.37
	FM	94.87	80.22	86.23	83.07
	LadFG	95.54	79.76	89.01	84.10

R. C. Gallen and E. T. Caro reveals the things which inspires and demotivates a learner for completion of courses, authors use clustering techniques to show the relationship. The dataset used in the paper is questionnaire consisting of 26 question regarding MOOCs. Clustering techniques help to find the motivation and clear perception of the learner's reason for enrolling the course, the clustering techniques isolates students who are self-motivated for the course and who are not. The methodology used in the paper follows KDD (Knowledge Discovery in Databases) which consists of six stages namely problem analysis, data preparation, data mining, clustering evaluation, clustering interpretation and feature extraction.

Hypotheses has been made in the paper with regard to the motivation of learner for taking up the MOOCs, the motivation was a) for the interest in learning and b) for the purpose of doing better in careers. T-tests was applied on different interest variables of the dataset and it was decided that users with proficient internet skills are likely to enrol in MOOCs compared to users with bad internet skills.

S. Boyer et al. shows the capture of raw log files of student's activeities on MOOCs and then converting the data into student's trajectory and then application of the logistic regression model to predict the dropout of the students on weekly basis. To solve the real time usage of model, which predicts the stop out in the following courses, the authors used two approaches namely a) naïve approach and 2) Transductive learning approach, in naïve based approach they used the logistic regression model for the first course and then applied that model on second course.

Students behaveiour and performance are being studied and compared with two different MOOCs, its answers three questions as follows:

Q1. Which features are responsible for student dropout?

Q2. How interaction of students changes over time?

Q3. How the interaction and interest effects if students get replies in forums?

Authors used data from Edx courses to predict the dropout of the students using various classification models such as Logistic regression, SVM (Support Vector Machine), random forest and GBDT [9]. The dataset is classified as enrolment features, user features and course features, dropout prediction is a binary classification problem. These four models are used to train a part of the data and test on remainning part of the data and the result shows that GBDT has the highest accuracy on the result. Table 2 shows results of the different models used in the experiment. From above results we can conclude that GBDT has the highest accuracy of 0.88 which can be used by the teachers to get insight of the student's behavior who is likely to dropout from the course.

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 5, May 2021

Model	*Accuracy*
Logistic Regression	0.65
SVM	0.65
Random Forest	0.86
GBDT	0.88

IV. MOOC DROPOUT HEADACHE

In last few years, MOOCs are very growing fast, providing the potential to gain access to education for students all around the world. Anyhow, despite the potential benefits of MOOCs, the number of learners who didn't complete or withdrew a course had been very high. The high learners' dropout or withdrawal rate of MOOCs has been because of many factors that generally can be classified as either student-related factors or MOOC-related factors.

Two types of factors	
Student related	Lack of motivation
	Lack of time
	Insufficient background knowledge and skills
MOOC related	Course design
	Isolation and lack of interactivity
	Hidden costs

- **Student-related factors**

Lack of student's motivations – is one of the most huge factors in preventing students from completing a course. The inspiration for the students is determined by many factors which among others include the future economic benefit, development of personal and professsional identity, challenge and achievement, enjoyment and fun. It is therefore of great use to find out about motives that drive students to enrol a MOOCs. It was observed that 95% of learners see entertainment and enjoyment as the main reason for enrolling a MOOCs followed by a general interest in a topic as another reason selected by 87% of students. A small number of students (15%) chose to register the MOOC to help them decide if they want to take a higher education class, while a low percentage of students (10%) decided to enroll the MOOC because they could not afford to attend formal education.

Lack of time – the amount of time needed to finish a course is seen to be another factor whichh plays a huge role in avoiding students from completing the MOOCs requirements. A study conducted found out that watching online lectures and completing home tasks and tests require too much time of student's schedule, and this is said to be one of the reasons that cause students to drop out of MOOCs

Insufficient background knowledge and skills – One more reason which causes learners to drop out from a registered course is scarce background knowledge and lack of required skills. Mainly, difficulties with maths requireements are seen to be of issue that students were not able to complete a course. Since a lot of the interaction in MOOCs depend on text, it is a must for learners to have, beyond technical skills, strong skills in reading, writing, and typing. Not having these skills is mainly a cause for not completing or withdrawing a course.

- ***MOOC-related factors***

Course design – This is one the key reasons that cause students to dropout of MOOCs. Course design is made of three components, namely, course content, course structure, and information delivery technology. Within these three components, course content is the most important predictor of
MOOCs retention. Learners who finished the MOOCs stated that these courses provided the material they were interested in learning accompanied with real cases and examples and practical implementation. Suggestions about soft skills and the learning material also were given by theses course. On the contrary, issues related to the content of the MOOCs such as the courses were too complex or technical, the language used was too complicated, the course had too many modules, etc., were Some of the problems reported by learners who couldn't complete the MOOCs.

Feeling of no interaction and lack of interactivity in MOOCs - is another factor which is shown to have a direct effect on learners' dropout in MOOCs. Results of a study conducted by authors about MOOCs' dropout showed that there isn't any interaction of students in a discussion or brainstorming providing thus low interaction and poor feedback between the lecturer of the course and students. Learners also stated in the study that teamwork and communication between students were also not

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 5, May 2021

Present and this creates the feeling of isolation.

Hidden costs – this is another reason that may cause many learners withdraw rate of MOOCs. These costs represent a sum of currency which sometimes is needed to be paid by students to get their certificates or to purchase pricey textbooks recommended by lecturers of the courses. Identifying and exploring the factors that have a direct effect on student's dropout or withdrawal of MOOCs would enable researchers, lectures, and educational technologists to investigate and propose new strategies and techniques that will help students to persist longer and complete MOOCs successfully.

V. IMPLEMENTATION

Machine Learning process diagram

With the use of dataset and feature engineering we split the dataset into training data and new data. With the help of learning algorithm we train the model and we also score the model which takes us one step closer to evaluate the model.

DATA PRE-PROCESSING: According to the problem statement ,'x' is taken as input with selective no of timestamps(inputs) and 'y' is taken as output the data might be taken in pairs of vectors and will be stored sequentially. It also includes splitting the dataset into training dataset and testing dataset so that after training the model with training dataset then we use the testing data to predict the further output.

LSTM ARE SENSITIVE TO THE SCALE OF THE DATA WE USE MINMAX SCALE

df.fit.transform()

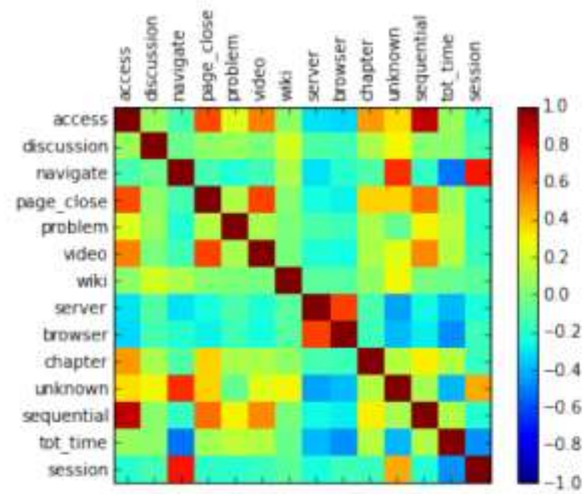
We are using a LSTM model(Long short term memory) use tenserflow and keras for lstm (sequential,dense and stacked)

1.model.add(to add layer)

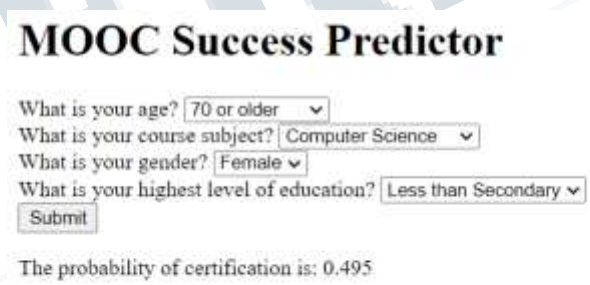
2.model.compile()

3.model.summary()

4.modelfit(x_train,y_train,validation data=x_test,epoch)



The above graph is the heat map which corelates all the features used in this model Ie visualisation which is one of the modules



The above image represents the module of calculating the probability of dropout after inputting age, course etc. We have used Flask server to implement the module and used tensorflow also.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 5, May 2021

The above diagram tells us to recommend courses based on what keywords we type which is one of our modules. We used KDDCUP 2015 dataset for implementation and various tensorflow imported modules to continue with the project

VI. DATA DESCRIPTION

The dataset used for this work has been sourced from the KDD Cup Challenge 2015. The data is entirely relation and event based, with no contextual information. The details of the initial dataset are as follows:
Objectmodule data - <course id, module id, category, children, start> contains data for each course having several modules.

Course id = ID for a particular course

Module id = each course can have multiple modules, with a module ID

Category = category of a module can be – chapter, course info, about, sequential, vertical, discussion, outlink, static tab, peergrading, course, combined, open, ended, html, video, dictation, problem

Children – module ID's for all children for a particular module

Start – start date for a module

The object module contains several duplicate entries, that is the same set of <course id, module id, category> are repeated in the file. We hence processed the object file to obtain the 'objectunique' file, which contains a single entry for the <course id, module id, category> triplet.

Log data - <enrollment id, time, source, event, object> contains data for an event on the page of a particular module.

Enrollment id = enrollment ID of the student, the enrollment ID for the same student is different for differrent courses.

Time = time of event

source = server, browser

event – access, discussion, navigate, page close, problem,video, wiki

object – ID of the module

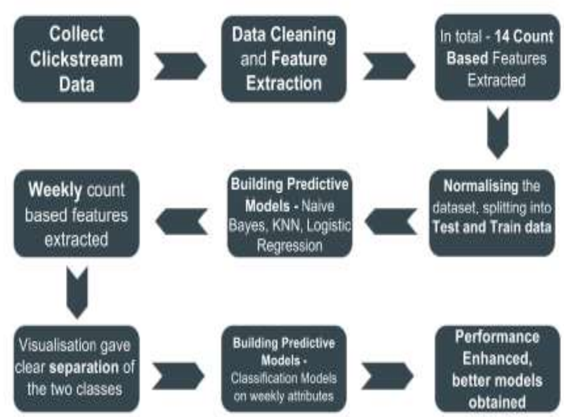
The log data contains several objects which have no listing in the object data file.

truth train data - <enrollment id, result> whether the student completed the course or not.

Enrollment id = enrollment ID of the student.

Result = 1 indicates that the student dropped out, and 0 indicates that the student completed the course.

VII. WORKFLOW OF DIAGRAM



VIII. FEATURE EXTRACTION

We obtained 14 features from the data sets listed above, after processing the data we obtained –
<enrollment id, access, discussion, navigate, page close, problem, video, wiki, server, browser, chapter, unknown, sequential, tot time, session, result> From each entry of enrollment ID in the log data we obtained the category of the object listed in the log record. The only non-zero features obtainned from this
Grouping were number of choapter accesses and number of sequential accesses, other modules in the log data, but unavailable in the module data, were categorised as unknown. The log data was itself used to form the heart of the clickstream attributes, each enrollment ID was grouped to obtain, the number of times a user has accessed a page of any module, the number of discussions participated in, total page navigations, count of page closes, number of problem, video and wiki accessed. We also include the number of server and browser requests issued, by grouping the log data. A session is the period of time when the user is active.

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 5, May 2021

We look at fifty minute time windows, from the start of a user's log record. Any click activity that the user performs is considered to be part of one session. If a user is inactive for fifty minutes, the session expires, and any further logs for the user are part of the next session. The total time is the total length of all the sessions put together. It needs to be understood that a session may be less than or more than fifty minutes, but a session expires when the gap between two consecutive click records for the user is greater than fifty minutes. Finally the result column was added, indicating whether a user dropped out – indicated by a 1 or not – indicated by a 0.

All the above mentioned 14 features were then normalised, limiting them to a range of 0 to 1.

We hence obtained a cleaned data set, for analysis, containing 120542 entries and 16 columns. This was followed by the creation of the train and test data sets. We randomly sampled the created features data set into two parts, following the Pareto Principle, which is the 80-20 rule, with eighty percent records in the training data and the remaining twenty percent in the test data. The resulting training data contained 96434 rows, and the testing data contained 24108 rows.

IX. VISUALISATION

We performed various visualizations techniques to get a brief idea of the data and the relation between various features. We started with constructing the correlation matrix and the heat map of all the extracted features.

We see that access and sequential are highly correlated, also session and navigate look to have a high correlation. However, later when forming the model, we establish that removal of either of the features for the development of the model does not give any improvement in performance. The attributes were hence left as it is.

X. CLASSIFICATION MODEL

Dropout prediction is a binary classification problem. Specifically, each <students,courses>pair is a sample, its features is $X(k) = (x(k)\ 1\ ,\ x(k)\ 2\ ,\ x(k)\ 3\ ,\ ...,\ x(k)\ n\ )$,wherein $x(k)\ i$, $1 <= i <= n$ is one of the features of this sample, it can be discrete or continuous variable. If we have known whether student *Ui* is drop-out in the course *Cj* , then this sample $Sk =< Ui, Cj >$ will be labeled as $y(k) = 1$, otherwise,

$y(k) = 0$. More specifically, we collect 39 courses data, each course contains 40 days users behavior log.We finally get more

than 20, 000 <students,courses>samples. For each sample we calculate UserFeature, CourseFeature and Enrollment Feature using the first 30 days data, and get 112 features for each sample, then we can calculate each sample *Sk* label *yk* using the next 10 days data. We divide all samples into two sets, one is training set, containing 120542 samples, the other is test set, containing 80362 samples. Commonly used supervised classifiers such as Logistic Regression, Nave Bayes classifier, SVM, decision-making tree and so on can be applied to this problem, The space of classifier selection model is large, we only train and tune four models, SVM, Logistics Regression, Random Forest and Gradient Boosting Decision Tree(GBDT).We adopt 5-fold cross-validation[10] for all classification models. The Best result we achieve is 88% accuracy with GBDT model.

XI. EXPERIMENT SETUP AND RESULT

The problem is solved as binary classification problem so outcome can be certified or not certified. Various Machine learning models are applied on the dataset to predict whether the student will be able to get certification or not. Various Machine Learning models used are Random Forest(RF), Decision Tree(DT), Support Vector Machine (SVM) and Naïve Bayes. To briefly summarize some important aspects for highlights: MOOCs witnesses a large number of students enrolling for the courses every year but gradually the students start dropping out of the courses due to various reasons and eventually at the end very few students are able to obtain the certification. Various reasons for dropout can be less interaction in forums, quality of course, losing interest over weeks and others. Above stated Machine Learning models are applied on dataset with the help of Data Analytics Software Weka [16]. Then comparison is made between models and one with highest accuracy is chosen. Table 4 gives a brief descryiption about the architecture, type and algorithm of the models used in the experiment.

Repeated K- fold cross validation was applied during the modelling, the elements from dataset were randomly partitioned into 10 equal size subsets. For each round of cross validation, 9-fold subsets are used as the train set and single subset is used for testing purpose.

Models	Decision Tree	Random Forest	SVM	Naïve Bayes
TP Rate	0.993	0.93	0.983	0.956
FP Rate	0.045	0.851	0.093	0.013
Precision	0.993	0.935	0.983	0.971
Recall	0.993	0.93	0.983	0.956
F-Measure	0.993	0.901	0.983	0.96
MCC	0.952	0.271	0.879	0.774
ROC Area	0.999	0.998	0.945	0.996
PRC Area	0.999	0.988	0.976	0.995

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 8, Issue 5, May 2021

Results shows that Decision Tree Models work best for the small part of the dataset, on classifying whether a student will dropout or not with a accuracy of 0.9931, whereas other models Random Forest, SVM and Naïve Bayes give lower though compelling results, with an accuracy of 0.9298, 0.9826 and 0.9560 respectively. This is due nonlinear form of relationship between the dataset features and target. The stated nonlinear classifier models are powerful models. These models are capable to handle nonlinear distinguéishable problems by transfer feature to high dimensional space. So non-linear models show higher accuracy on then KDDCUP dataset, depending on the size of dataset and complexity models like Decision Tree, Random Forest will give best prediction.

XII. CONCLUSION AND FUTURE SCOPE

In this paper overview of models to predict dropouts of students in MOOCs and various factors affecting it has been comprehensively reviewed. To briefly summarize some important aspects for highlights:
Machine learning techniques can be used to predict the dropout of the students which can help the instructors to model their course accordingly. Dataset from MOOCs can be used to train on different classification models and by examining their accuracy, models can be selected to use it in real time prediction of dropouts of students in the course. This paper uses small dataset with instances 5200, larger dataset can be used to see the result in even bigger picture.

REFERENCES

[1] Chen, R. (2012). Institutional characterristics and college student dropout risks: A multilevel event history analysis. *Research in Higher education*, *53*(5), 487-505.

[2] Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016, April). Combining click-stream data with NLP tools to better understandd MOOC completion. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 6-14).

[3] Fei, M., & Yeung, D. Y. (2015, November). Temporal models for predicting student dropout in massive open online courses. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (pp. 256-263). IEEE.

[4] Leon Urrutia, M., Fielding, S., & White, S. (2016). Professional developent through MOOCs in higher education institutions: Challenges and opportunities for PhD students working as mentors. *Journal of Interactive Media in Education*, *1*, 1-11.

[5] Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, C., Zhang, Q., & Xue, Y. (2016, February). Modeling and predicting learning behavior in MOOCs. In *Proceedings of the ninth ACM international conference on web search and data mining* (pp. 93-102).

[6] Ulriksen, L., Madsen, L. M., & Holmegaard, H. T. (2015). Why do students in stem higher education programmes dropopt out?–explanations offered from research. In *Understanding student participation and choice in science and technology education* (pp. 203-217). Springer, Dordrecht.