

Intelligent Recommender System using Machine Learning to Reinforce Probation Students in OMAN Technical Education

Dr. Mohammed Faisal¹, Suad Abdullah Al-Riyami², Safia yasmeen²
Department of Information Technology, Nizwa College of Technology^{1&2}, OMAN
Department of Software Engineering, Faisal University², SAUDI ARABIA

Abstract

Machine learning is fascinating the entire industrial world to get new insights from different verticals. In extent to that recommender system is quiet popular to provide customized individual recommendation whenever there is information overflow, this prediction capabilities boosting the business in order to increase the revenue specially in the ecommerce platform. Past few years researcher's has adopted machine learning and recommender system proficiencies to accelerate educational data mining and learning analytics. In this article, researchers are contributing to find new and innovative insights using the said techniques. Currently OMAN is a fast growing country and rapidly trying to transform its existing education system into a modern and excellent education system. In this scenario, the proposed insights will help the academicians and scholars to improve the students' performance. Based on the analysis and observation percentage of scholars who is under probation has been rapidly increasing, the proposed model will use the basic concept of collaborative filtering recommender system using machine learning to predict the course wise grade before opting the individual course. The proposed model will predict how much grade he could get into the individual course and advisors can compare with grade, which is required to come out from the probation. In addition to that, to know the reason of the tutees probation, correlation analysis based on machine learning model has been implemented to find the most result effecting factors. The said insights will help the advisors and faculty members to recommend the students what actions required to clear his/her probation.

Keywords: *Recommender System, Collaborative filtering, users K-NN, Learning Analytics, Student performance and Educational Data Mining*

I. RELATED WORK

Systems that retrieve and filter the data through content and similar profiles are known as recommendation systems (RS) which make use of either or both collaborative filtering and content-based filtering. With the introduction of web based e-learning systems like Blackboard, Moodle, MOOCS, GitHub, etc., are having huge amount of educational data being produced and this enormous amount of dataset has been generated and stored in educational sectors. Collaborative filtering methods form a model from a user's past online cognitive behavior like browsing history by the users. This model is then used to predict objects that the user may have an interest in [1]. Content-based Filtering systems are based on problem attributes; and hybrid techniques attempt to combine both of these designs. The architecture of recommender systems and their evaluation on real-world problems is an active area of research.

Currently, many web analytical methods and tools are being used to analyze these educational data and generate different predictions and recommendations for students, teachers, professors in universities and schools. In last few years, there have been a lot of work done with this big data set and many recommender system (RS) algorithms have been proposed to predict the results, to recommend the possible outcome and generate the positive e commerce results.

The main idea for these recommender systems is to help the user to deal with the huge dataset available through the collaborative environments like social media and to personalize it as per the required objective [1]. The recommender system in education system and organizations relies on the historical data of the students and generates the recommendation based on the quantity of the data, cognitive skills results of the student, or teacher. Many social learning networking sites like Open Web Application Source, GitHub, EdX etc are useful in generating many collaborative, effective and open learning technological opportunities to online learners.

Recommender systems (RS) have become an important research area since the appearance of the first papers on collaborative filtering in the mid-1990s [2], [3]. RS is not only being helpful in educational area but also its getting better results in other commercial sectors like Amazon, Netflix, Facebook, EBay etc. S.M. Wang [5] framework the projections, for different e-learning stakeholders, of the deployment, aggregation, and combination of both learning and social analytics. It is stated that this practice can lead to more adaptive, interactive and functional social learning environments where datasets of information will be well connected and take the form of collective knowledge in order to serve participants needs. Studies in this direction are already present. As cited [7], interpretations for recommender systems were based on straightforward correlation statistics and predictive modeling. No much focus was on machine learning. Recommender systems can be classified depending on the type of method used for making recommendations [4], they can be broadly categorized as:

The Content-based recommending method to recommendation system has its roots in information retrieval [8], and information filtering for Analysis. These approaches recommend items that are similar in content to items the user has liked in the past, or matched to predefined attributes of the user. Decade back the content-based search started growing official in the form of data mining and data warehousing, which produced the huge datasets from libraries, official documents, corporate blogs etc.

Collaborative Filtering systems, a user is recommended items based on the past ratings of all users collectively. It is a technique of filtering the dataset based on the collaborative agents, viewpoints. Like many other machine-learning algorithms, the collaborative filtering is being used by many recommender systems in drawing the conclusions and inferences. In other word, it is an approach to explicitly provide the user with the information about certain product. Example if a student quest for certain Book online, the website can provide not only details about the book but will also show cause number of reviews and other books from same author or title will be recommended.

Hybrid recommender systems combine two or more recommendation approaches in different ways to profit from their paired approach advantages and overcome their limitations. For example, merging the CF system with CB system components can develop a multi-dimensional recommendation system considering both user condition and the product attributes, then generate the User * product analysis hybrid/Matrix too generate the recommendations. However, both collaborative and content-based methods have certain limitations but merging both can extended in various ways for generative more flexible and less intrusive recommender system

II. Proposed Model Description

The proposed model has been constructed to give contribution in the educational services for OMANI's Tech students, however the same model can be adopted by the any other nation educational system, especially where the students' performance are falling down drastically

[10]. The research article intend to predict the probationary students expected grade before registering the courses for the new semester, the proposed model will utilize the capabilities of recommender system by using collaborative filtering method[9].

To build a supervised learning based model, to build the training data set system utilize the past students results whoever has completed the course which probation scholars are going to register. On other hand test, data set will be the set of expected courses that the department can offer. The system will give the prediction of individual course, for instance student have the option to choose five courses from the eight courses. Prediction shows student will get 67 marks in five courses and 80 marks can obtain in rest three courses. In this scenario if student are restricted to use register five courses then the advisor can advise to register three courses which have been predicted high marks and add just two courses for which prediction shows the less marks instead of giving the all five courses where prediction shows the low marks.

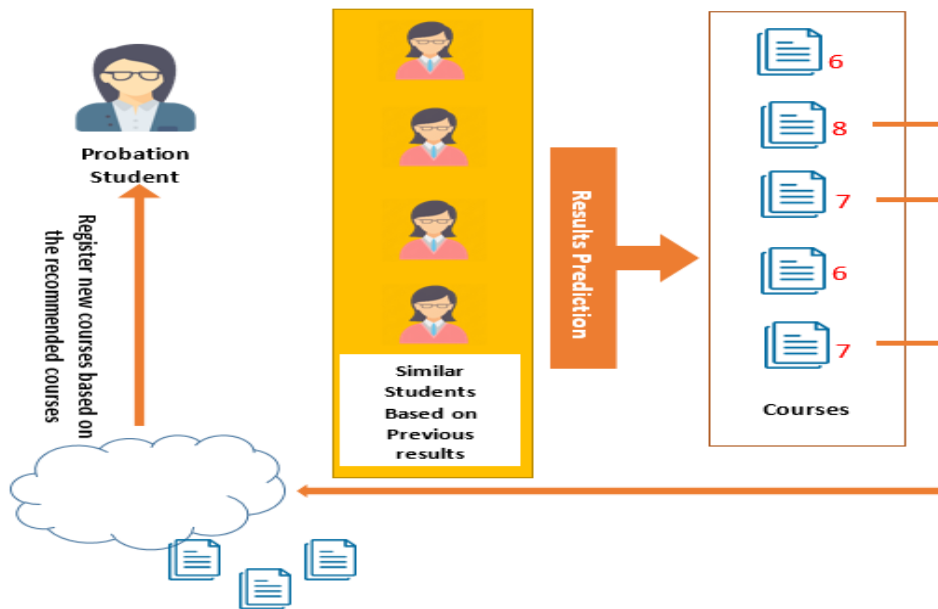


Figure 1: Collaborative Filtering Recommender System overview

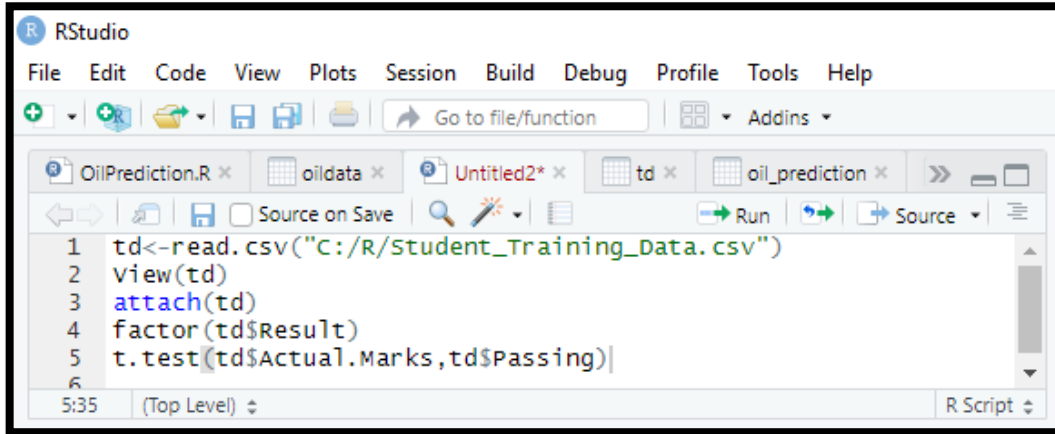
In term of probation cases as per the current practices in OMAN technical colleges, when the student getting 2.0 or less than 2.0 GPA in their current semester the will be labelled as first probation. To come out from the probation students need to score need to score more than equal to 2.0 CGPA (Cumulative GPA). Student will fall into second and third probation he/she unable to clear the probation in consecutive semester. Students will be dismissed from the college if they are unable to clear their third probation. In some exceptional cases, authority may give the extra chance to clear their probation.

III. METHODOLOGY ND PRACTICAL IMPLEMENTATION

In this research article, we have chosen SEMMA (Sample, Explore, Modify, Model and Access) data mining process model [9]. SEMMA model has been developed by SAS enterprise. It consists of five phases.

PHASE ONE: SAMPLE

In this phases we are doing data sampling to make sure the size is big enough. In this article, we have performed T test using R programming language. We have taken the sample data frame that is student result data.



```
1 td<-read.csv("C:/R/Student_Training_Data.csv")
2 View(td)
3 attach(td)
4 factor(td$Result)
5 t.test(td$Actual.Marks,td$Passing)|
```

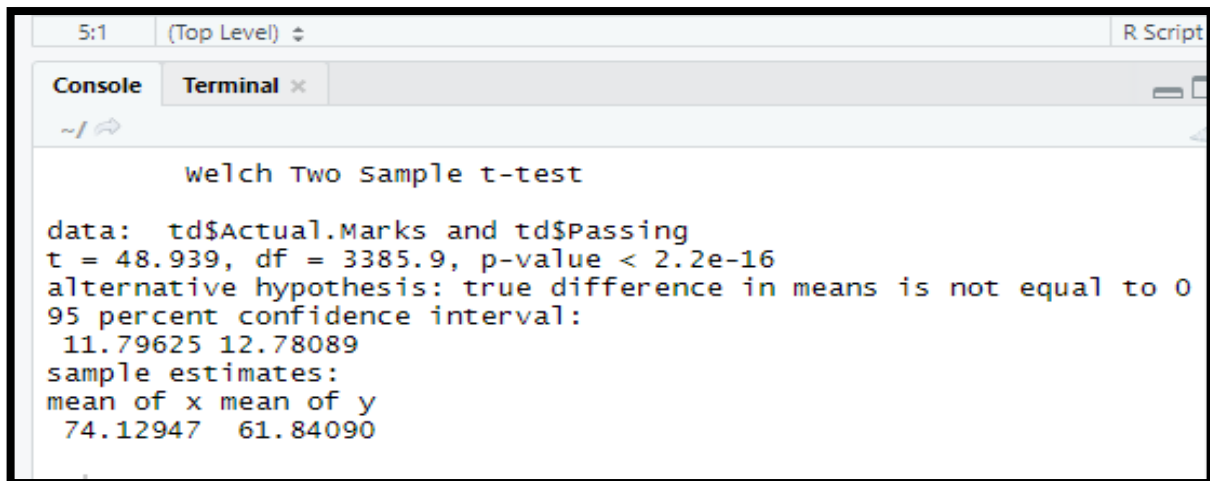
Figure 2: Code written to perform T test in R

Null hypothesis: Difference of x and y mean is equal to 0 (Zero)

Alternative hypothesis: Difference of x and y mean is **NOT** equal to 0 (Zero)

X= student actual marks (Continuous variable) Y= Student result (Categorical variable; Label: Pass, Fail)

To build a supervised learning based model, to build the training data set system utilize the past students results whoever has completed the course which probation scholars are going to register. On other hand test data set, utilize the probation student's result



```
welch Two sample t-test

data: td$Actual.Marks and td$Passing
t = 48.939, df = 3385.9, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 11.79625 12.78089
sample estimates:
mean of x mean of y
 74.12947  61.84090
```

Figure 3: Output as alternative hypothesis is true as mean of x and y is not 0

PHASE TWO:EXPLORE

Explore the proposed work chosen the collaborative filtering recommender system, which has developed specifically for individual recommendation wherever there is information over flow

in the e-commerce platform. In this research the basic concept, has been adapted and data tuned to supply in the constructed model as the follows:

Table 1: Data Description

Data Set	Collaborative Recommender system nomenclature	Actual Field Name	Data Description
Training Data Set	User ID	Student No	Student College/University ID to act as a user identification.
	Item ID	Couse No	Course code to act as an item identification
	Rating	Actual marks	Actual marks is the result of past students who already studied those courses, it considered as rating recommendation in the proposed system.
Test/Query Data Set	User ID	Student No	Same as the Training data set description
	Item ID	Couse No	

PHASE THREE: MODIFY

In this phase data preparation will be takes place including select, create and transforming the variables. Concisely we can say, we need to tune the data as per the requirement of the model, which we intend to develop in the next phase.

The sample data of the students results which has been taken from CIMS (Colleges Information Management System). It consist of following fields:

- Rec No: Record number in the CIMS
- Semester: Semester 1/2/3
- Student No: Student registration number
- Student: Name of student
- Department Name: IT/Engineering/Business
- Course No: Course Code
- Course Name: Name of the course
- Actual Marks: Total marks out of 100
- Passing grade: Passing grade
- Result: Pass/Fail

In the proposed recommender system we have dropped 07 fields and only kept the following 03 fields:

- Student No: Student registration number
- Course No: Course Code
- Actual Marks: Total marks out of 100

PHASE FOUR: MODEL

In this section process Proposed model has been iluustrated in the fig 3 and operators will be discussed to give the more clarity into the entire process. Where many rapidminer operators have been used to develop

the mode[11]. Each operators synopsis and purpose have been depicted in the Table 2. With reference to the Fig 3 model diagram the original dataset passed into the operator select attribute where we have dropped unwanted columns and kept only the required columns. Only three column kept for further model development those are student no., course no. and actual results. After selecting the required columns data has been passed to set role operator which had been renamed and utilized as user identification, student no. taken as a user identification role. In similar fashion course no. had taken to set role as item identification in the model. To extend the the further process data have been passed to set the predictor variable , where actual result had been set as label that is nothing but predictor variable.

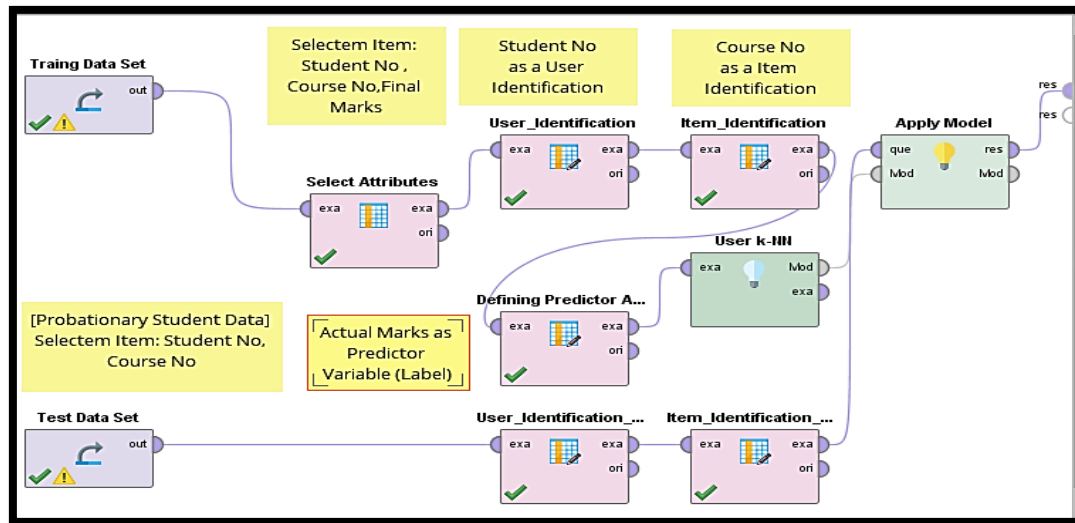


Figure 4: Collaborative Filtering Recommender system

After defining the roles the model will pass the data into the main operator called user k-nn which will facilitate the collaborative filtering based on the similar user. This operator will be used to train the model with the historical data as students previous years result. Apply model operators have been used to generate the result prediction. Apply model has been connected from two data sources that is training data and another the test data.

Table 2: Rapid Miner operators used in the Model

Name of operators	Synopsis	Purpose in Proposed Model
User K-nn	Unweighted k-nearest neighbor user-based collaborative filtering using cosine similarity.	User K-nn used in order to develop Collaborative Filtering based Machine Learning Model.
Set Role	This Operator is used to change the role of one or more Attributes.	Set role parameter has been changed as User Identification; Student Id and Item Identification as Course Code.
Select Attribute	This Operator selects a subset of Attributes of an ExampleSet and removes the other Attributes.	Select attribute have been used to drop the unwanted fields and picked only Student No., Course No. and Actual Result from the data set
Apply Model	This Operator applies a model on an ExampleSet.	Apply model has been used on training data and test data

In another hand if we will talk about preparation of test data before passing it into the apply model operators, it has been done in similar fashion except one major difference in the test data we are just passing the user identification and item identification except the predictor variable because the actual result need to predict by the model.

IV. PHASE FIVE: ASSES - RESULT AND DISCUSSION

This is last phase of the SEMMA methodology where result usefulness of the developed model will be discussed. This section has been emerged with result and discussion section of the article as its going to talk about the same.

The proposed system predicted result have been verified by splitting the data set into 70%-30% ratio as training and test data set. We have verified the predicted result by checking actual results of test data and we found the accuracy rate is around 90-95%. This research is an effort to produce a new insight by using the existing concept of recommender system which has been developed initially for e-commerce industries.

With reference to the Fig 1, we assumed that total remaining courses of a probation student is 5 and in the current semester we suppose to give 3 courses due to some constrain, as per the prediction expected marks out of is 67,82,75,67 and 79. In this prediction we can clearly two courses result prediction is just 67. So as per the recommendation courses to register is based on the high rate courses with the expected score 82,75 and 79. Remaining two courses which score is 67,67 can be register in the next semester. In such a way student can perform better and better in order to come out from the probation cases.

Row No.	Student No	Course No	prediction
1	16S1490	ITDB3208	74.144
2	16S1490	ITSE3211	80.937
3	16S1490	ITSE4206	69.522
4	16S1490	PHIL3201	73.170
5	16S163180	ENGL2100	78.924
6	16S163180	ITDB202	82.405
7	16S163180	ITIS104	80.489
8	16S163180	ITNT102	82.507
9	16S163180	ITSE101	75.346

Figure 5: Prediction Result

Prediction results illustrated in figure 5, for instance we can see the model has predicted the individual courses for each students, row 5 to 9 showing the prediction of five course of the same student. Academic advisor can choose four courses based on high score and can leave the least scored.

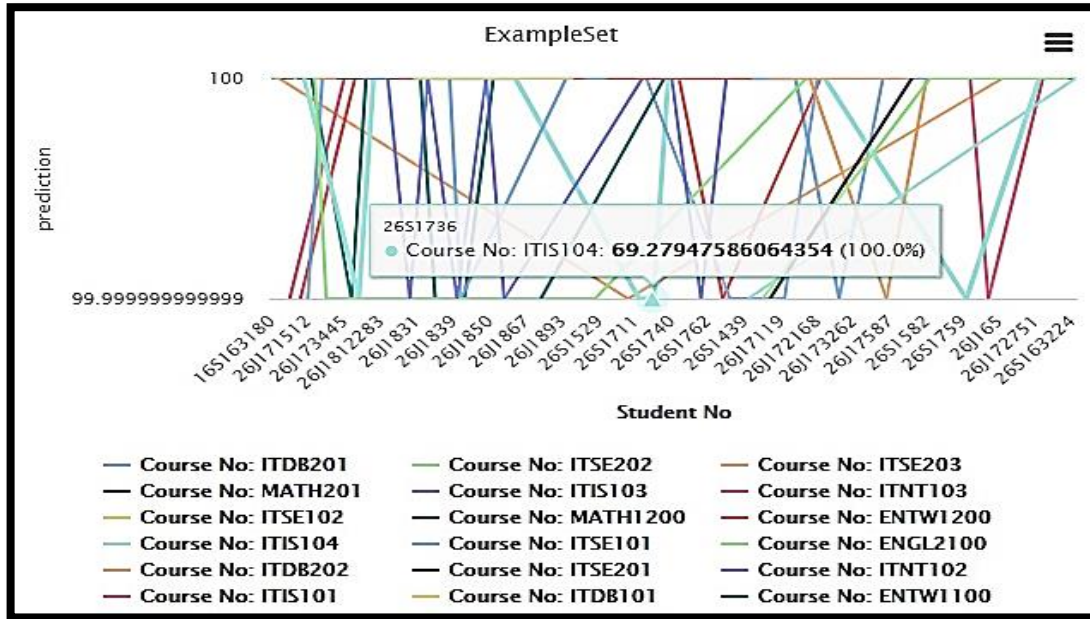


Figure 6: Result prediction graph

V. CONCLUSION AND FUTURE WORK

There are plethora of learning analytics insights and growing rapidly to contribute into the area of educational data mining. This research article contributed by introducing a new insight by utilizing the existing method called collaborative filtering and machine learning. The research intend to benefits the OMAN educational community (Can be utilized in any other educational institution) revitalize the students' performance in an innovative way, where students not even know how their grade are booting up. According to our experience just pressurizing the students to perform better and convincing them to come out from the probation is not the solution. With reference to this effort researchers are full of positive hope that, it will help the weak tutees to revive their grades.

This research can be extended to find more noble insights in the area of learning analytics. The other methods of recommender system can be utilized to construct much more similar insights to boost the student performance.

VI. ACKNOWLEDGEMENT

We would like to acknowledge our warm thanks to our key officials of Nizwa College of Technology our honorable Assistant Dean Academics Dr. Ibrahim Ali Saleh Al Aamri and our esteemed research head Dr. Khalifa Zayid Al Shaqsi for their approval, encouragement and support. We also like to extend our thanks to Head of Students Affairs department Ms Azza Al-Nabhani for proving the required data and for her kind support.

REFERENCE

- [1]. Prem Melville and Vikas Sindhwani, “Recommender Systems”, Encyclopedia of Machine Learning, 2010.
- [2]. U. Shardanand and P. Maes, “Social Information Filtering: Algorithms for Automating ‘Word of Mouth’,” Proc. Conf. Human Factors in Computing Systems, 1995.
- [3]. P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl, “GroupLens: An Open Architecture for Collaborative Filtering of Netnews,” Proc. 1994 Computer Supported Cooperative Work Conf., 1994.
- [4]. Cheung, K.-W., Kwok, J. T., Law, M. H., & Tsui, K.-C, Mining customer product ratings for personalized marketing. Decision Support Systems, . (2003).
- [5]. S.M. Wang, “A Module-Based Learning Analytics System for Facebook Supported Collaborative Creativity Learning,” in 2014 IEEE 14th International Conference on Advanced Learning Technologies, 2014.
- [6]. Prem Melville and Vikas Sindhwani, Recommender Systems, Encyclopedia of Machine Learning, 2010.
- [7]. Billsus, D., & Pazzani, M. J. Learning collaborative information filters. In Proceedings of the fifteenth international conference on machine learning (ICML),Madison,Wisconsin. San Francisco:Morgan Kaufmann, 1998.
- [8]. R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison-Wesley, 1999.
- [9]. Nguyen Thai,NgheLucas Drumond,Artus Krohn,GrimbergheLars Schmidt Thieme, "Recommender system for predicting student performance", Procedia Computer Science 1 (2010) 2811–2819.
- [10]. Iqbal, Zafar & Qadir, Junaid & Mian, Adnan & Kamiran, Faisal. Machine Learning Based Student Grade Prediction: A Case Study. (2017).
- [11]. Syed Tanveer Jishan, Raisul Islam Rashu, Naheena Haque and Rashedur M Rahman. “Improving accuracy of students’ final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique”, Decision Analytics (2015).