

# APPLYING DATA MINING CLASSIFICATION APPROACHES FOR PREDICTION OF STUDENT EXAM RESULTS

<sup>1</sup>Abhilasha Dangi, <sup>2</sup>Dr. Sumit Srivastava

<sup>1</sup>Research Scholar, Deptt. Of Computer science  
Mewar University, Chittorgarh, Rajasthan, India,

<sup>2</sup>Associate Professor, Department of Computer Engineering,  
Manipal University Jaipur District Bagaru, Jaipur, Rajasthan, India.

**Abstract :** Educational data mining deals with designing techniques for extracting information from data of academic arena. We used academic data mining in this paper to enhance the output of undergraduates and solve the issue of low test scores of undergraduates. This forecasting will assist the teaching members to identify the students who are weak and helping them to develop better test results. From CCT College, the student details have been taken for analysis, and data mining tools were used to extract reliable information. The numerous academic data mining techniques that guide the administrators to recognize and enhance student success have been demonstrated in this article. After data pre - processing the dataset, we adapted data mining algorithms to explore classification standards.

**IndexTerms** - Data Mining Tools, Educational Data Mining (EDM), Classification techniques, Forecasting.

## I. INTRODUCTION

EDM is a recent area in the study of information extraction from Databases that concentrates on mining useful trends and exploring valuable information from academic information management. The information can be individual or educational. It can also be obtained from e-learning systems that have a massive amount of data used by several institutions. [9][2].

A fascinating area of study, which primarily focuses on evaluating and interpreting students, is the key goal of higher education institutions to make informed choices or to increase the success of students. The information acquired can be used to provide educational leaders in higher education institutions with supportive and positive guidance to strengthen their decision-making process, increase the academic achievement of students and minimize the rate of failure, understand exactly the actions of students, support teachers, strengthen learning as well as many other advantages [1]. Classification is the most powerful data mining method used to identify and forecast values. EDM is no exception to this statement, so in this research article it's used to analyze the data of gathered students via a questionnaire as well as provide categorizations based on the information obtained to forecast and rank the success of students in their forthcoming semester.

From the knowledge acquired, we need to give a supportive and positive advice to college management to solve the issue of low grades of graduate students and to boost the academic success of students. In data retrieval and data processing, there are many common classification approaches and strategies used. Any process or strategy has its benefits and drawbacks. Therefore, this study shows various classification strategies to validate and check the findings of different classifiers. The best outcome in terms of prediction accuracy and quality may eventually be chosen.

The rest of portion of this study is summarized as follows Section 2 describes similar works in data mining techniques. Section 3 details the data collection and the processes of planning and analysis done. Section 4 mentions our experimental studies about analyzing the data mining methods. Finally, in Section 5, we end this text with a summary and future work.

## II. Review of Literature

The literature review suggests that certain topics may be of concern to multiple scholars during the last few decades. Many experiments have been carried out just to explain the value of "data analysis" strategies in education, revealing that this is a new phenomenon for the method of acquiring reliable and precise knowledge about the learning behaviour and usefulness in the learning [2][3]. Ayesha, Mustafa, Sattar and Khan [28] describes Use of the k-means clustering algorithm to forecast the academic habits of students. For both teachers and students, the knowledge produced just after introduction of data mining techniques can be beneficial. Data Mining "was also used to evaluate the course and topic of existing research problem in the area of education strategies, and to assess student success." [4]. Cortez and Silva [10] Study on the educational institutions in Portugal has been carried out and the findings have shown that successful and reliable forecasts can be made. This is built by implementation resources that help enhance school curriculum administration and the efficacy of education, and this is a very substantial benefit. Varghese, Tommy and Jacob [6] The "K means" method has been used in their analysis to group 8000 students based on 5 factors (input median of the test results / examinations in the Institution mean results, paper mean score, conference notes and frequency study notes). The "K means" method was used in their analysis. A good association among presence and educational outcomes was demonstrated by the study. Within this new research purpose, multiple inquiries have been undertaken A Naïve Bayes methodology, for instance, was used by Bhardwaj to forecast student success depending on 13 variables [5]. The findings have been used to establish a paradigm that is used to preprocess the students at risk of complications and hence to trigger a curriculum of instruction and counseling. Gulati and Sharma [7] claim that Information by 'Data Mining' research will enhance the preparation, student success and operational management of the educational sector. The assessment research was directed by Ayesha Mustafa[8], depending upon the nature of education and test review at the start and end of the programs. Noaman and Al-Twijri [11] A new report concerning the Universities of Saudi Arabia's admission criteria has been released. They uses models and a framework that suits the population as well as the factors that characterise it, with methods they

have created. Input entry to the volume of observations in previous college, entry notices, and even features that define the requirements of the university were taken into consideration. Some reports have shown the effect of the Data Mining method of Moodle[12]. The numerous data mining strategies which can be implemented to facilitate learning outcomes on online media are defined by Sun[9]. To have a model of academic achievement, Aslam and Ashraf[13] used cluster analysis. Grafsgaard, Wiggins, Boyer, Wiebe, and Lester [14] A device that identifies facial movements based on the students' annoyance or comprehension in the class has been developed. To recognise unspoken actions and correlate them with the information learned, they have used techniques. A database for using human behaviour predictive model is also defined by Seong Jae Lee[15].

### III Prepare Your Paper Before Styling MATERIALS AND METHODS

#### 3.1 Data Set Description and Preprocessing

The educational application is commonly supported in Rajasthan as a government by the programs planned and established by the University of Rajasthan (UOR). The University does have approximately 10,000 students. with its constituted institutions established. Registered students through the application. 152 entries from the Center of Converging Technologies (CCT), an independent entity funded by AICTE / UGC operating the specified data collection, are included. The entrants consist of a person's private details sent to the CCT for the B.Tech / M.Tech dual degree programs. There are 42 numerical / nominal features in the Data Collection Fig. [1] Equivalent to father / mother sector, caste, sub-caste, 10th-12th proportion, term 1, term 2, term 3, etc.

Column1	Column2	Column3	Column4	Column5	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14	Column15	class	column16	column17	column18
CCT1001	1 DS	House Wife		1	1	85.2	1	78.6	TRUE	English	Genral	Hindi	0 Indian	1 good	Hosteler		
CCT1002	0 DS	House Wife		2	1	84	1	73.4	TRUE	English	SC	Hindi	0 Indian	1 average	average		
CCT1003	0 DS	House Wife		1	2	75	2	62.62	FALSE	Hindi	SC	Hindi	0 Indian	1 poor	average		
CCT1004	0 Hosteler	House Wife		1	2	82	2	60.92	FALSE	Hindi	General	Hindi	0 Indian	2 poor	average		
CCT1005	0 DS	Teacher		1	2	78	2	70	TRUE	Hindi	General	Hindi	0 Indian	1 average	average		
CCT1006	0 DS	House Wife		1	2	75	2	56.92	TRUE	Hindi	sc	Hindi	0 Indian	1 poor	average		
CCT1007	1 ds	House Wife		1	2	7	2	78.4	FALSE	Hindi	Genral	Hindi	0 Indian	2 good	average		
CCT1008	1 Hosteler	Teacher		1	2	70	2	67.89	FALSE	English	General	Hindi	0 Indian	2 average	average		
CCT1009	0 DS	House Wife		1	1	75	1	60.21	FALSE	English	Genral	Hindi	0 Indian	1 poor	average		
CCT1010	1 DS	Accountant		1	1	80	1	67.2	FALSE	English	Genral	Hindi	0 Indian	1 average	average		
CCT1011	0 DS	House Wife		1	1	82.06	1	87.38	TRUE	English	Genral	Hindi	0 Indian	1 super	good		
CCT1012	0 DS	Teacher		1	1	78	1	77.5	TRUE	English	Genral	Hindi	0 Indian	1 good	average		
CCT1013	0 DS	House Wife		1	1	78.89	1	76.8	FALSE	English	Genral	Hindi	0 Indian	1 good	average		
CCT1014	1 Hosteler	Teacher		1	2	85.83	2	86.15	TRUE	Hindi	OBC	Hindi	0 Indian	2 good	good		
CCT1015	0 ds	House Wife		1	2	82	2	70.2	TRUE	Hindi	Genral	Hindi	0 Indian	2 average	average		
CCT1016	1 DS	House Wife		1	2	76	2	94	FALSE	Hindi	Genral	Hindi	0 Indian	1 super	good		
CCT1017	0 ds	Interior designer		1	2	75	2	69.6	FALSE	Hindi	Genral	Hindi	0 Indian	2 average	average		
CCT1018	0 DS	House Wife		1	2	77	2	60.92	TRUE	Hindi	Genral	Hindi	0 Indian	1 poor	average		
CCT1019	1 DS	Teacher		1	2	64.53	2	65.08	FALSE	Hindi	Genral	Hindi	0 Indian	1 average	average		
CCT1020	1 ds	House Wife		1	1	81	1	62	TRUE	English	Genral	Hindi	0 Indian	2 poor	average		
CCT1021	0 DS	House Wife		1	1	76	1	80.46	TRUE	English	Genral	Hindi	0 Indian	1 good	good		
CCT1022	1 ds	Interior designer		1	1	84.8	1	74.2	TRUE	English	Genral	Hindi	0 Indian	2 good	average		

Figure 1: Student Dataset

#### 3.2. DATA MINING METHODOLOGIES

Classification is a quite basic method of data mining which is often used. To grasp the description, understanding of dataset is compulsory.

There are two steps of the process for classification:

- Creation of a teaching model
- Analyzing the model through knowledge research

Neural Network based algorithms: In this process, a model is generated that gives a data modeling format. All features linked to the subset are rerouted into a graph at the moment of subset classification [10].

K nearest neighbors: It is a distribution - free system which relies on the use of distance estimation. All existing instances can be preserved in it and can be categorised according to the distance measure if a new instance is inserted [11].

##### 3.2.1 MULTILAYER PERCEPTRON NETWORK CLASSIFIER

It is a kind of neural network fee-forward network. It is made of 3 node layers, i.e. the input layer, the secret layer and the output layer. From a set of input data, it produces a set of outcomes. A Multilayer perceptron made up of multiple insertion layers that are related as a guided graph among hidden layer and output layer to one another [10]. It is a deep learning strategy that could be used in speech recognition, image identification and language processing.

##### 3.2.2 K- NEAREST NEIGHBOR CLASSIFIER

The nearest neighbour is determined in this method and per the k-value that specifies the amount of nearest neighbours to be counted and thus defines the class of a specimen data point. [1]. often, using more than one neighbour to decide the class of a given criterion depending on the justification for calling: KNN. As sets of data has to be in the storage at runtime, this methodology is refers to as a memory-based strategy.

#### IV. RESULTS AND DISCUSSION

Python is a platform that incorporates a wide variety of mining algorithms and is generally used in data mining problems. The Student.arff file was generated with both the data gathered above. Classification methods use the collection of data to estimate and interpret the system and effects for the usefulness of the proposed planning process. MLP and KNN classifiers are used for research, a 5 fold cross validation is chosen.

**Table 1. Prediction Accuracy Recall and F1 Score of five folds on MLP Classifier**

	Folds	Accuracy	Precision	Recall	F1 Score
0	1	0.72692	0.73664	0.72692	0.7215
1	2	0.71923	0.71844	0.71923	0.71584
2	3	0.76538	0.74926	0.76538	0.73482
3	4	0.69615	0.72156	0.79615	0.70341
4	5	0.72519	0.73322	0.72519	0.71729

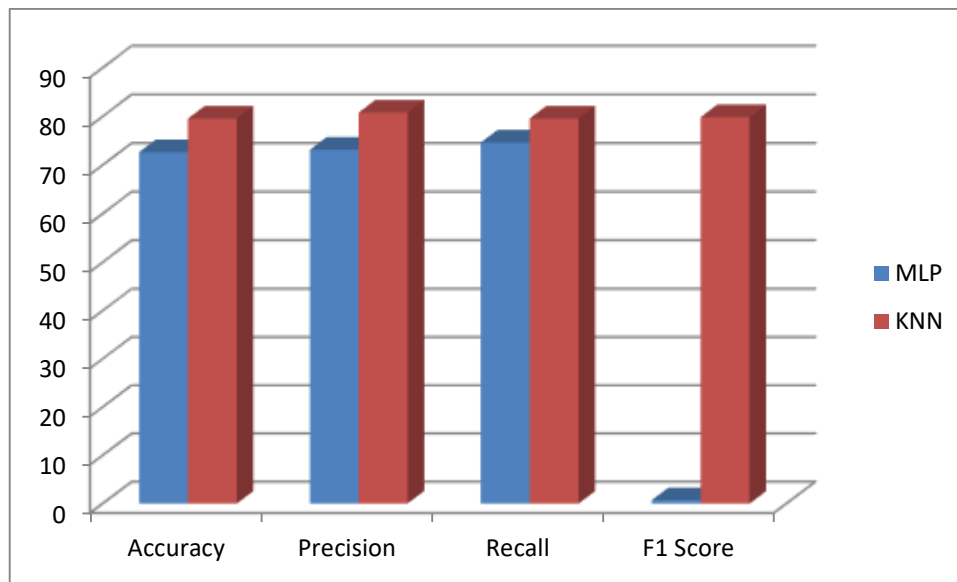
Table-1 demonstrates predictive performance measures of the MLP classifier for a particular data set. Table-2 describes performance measures of the KNN classifier for a particular set of data. In Table-3 the average of the 5 runs of the proposed algorithm was estimated and analyzed.

**Table 2 : Prediction Accuracy Recall and F1 Score of five folds on KNN Classifier**

	Folds	Accuracy	Precision	Recall	F1 Score
0	1	0.82692	0.83465	0.82692	0.82146
1	2	0.81153	0.81174	0.81153	0.80423
2	3	0.72692	0.74387	0.72692	0.72608
3	4	0.79615	0.82156	0.79615	0.80341
4	5	0.81744	0.82897	0.81744	0.81002

**Table 3 : Average Percentage of 5 Runs**

	MLP	KNN
Accuracy	72.6574	79.5792
Precision	73.1824	80.8158
Recall	74.6574	79.5792
F1 Score	0.718572	79.8572



**Figure 2. Comparison of Average Accuracy**

Figure.2 shows the Prediction accuracy of the MLP and KNN methods for data mining. The findings demonstrate that KNN delivers the best of 79.5792% , 80.8158 percent precision ,79.5792% recall and 79.8572% F1 Score respectively.

## V. CONCLUSION

This dissertation deals with the assessment of student academic achievement using multiple classifiers. This research helps the institution to consider the academic success of students in advance and can then work on weak students to improve their academic results. For the student record, predictive precision above 72% was demonstrated by all classification methods analysed in the study. It is reported that KNN 's efficiency is higher than the MLP methodology used in the research. A small data collection was used to do the analysis and it can be applied to a large volume of data. In addition, to evaluate the educational data collection to help determine what the future job will be, a combination of different classification approaches will be used.

## VI. ACKNOWLEDGMENT

We acknowledge the director of Centre for Converging Technology to provide us the data-set for performing the pilot project for classifying the data.

## REFERENCES

- [1] T. N. PHYU, "SURVEY OF CLASSIFICATION TECHNIQUES IN DATA MINING," INT. MULTICONFERENCE ENG. COMPUT. SCI., VOL. I, PP. 18–20, 2009.
- [2] M. Ramaswami and R. Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining," Int. J. Comput. Sci. Issues, vol. 7, no. 1, pp. 10–18, 2010.
- [3] M. Ramaswami and R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining," J. Comput., vol. 1, no. 1, pp. 7–11, 2009.
- [4] A. Permata Alfiani and F. Ayu Wulandari, "Mapping Student's Performance Based on Data Mining Approach (A Case Study)," Ital. Oral Surg., vol. 3, pp. 173–177, 2015.
- [5] V. Kumar, "An Empirical Study of the Applications of Data Mining Techniques in Higher Education," Int. J. Adv. Comput. Sci. Appl., vol. 2, no. 3, pp. 80–84, 2011.
- [6] J. Varghese, Bindiya M; Tomy, Jose; Unnikrishnan, A; Poulose, "Clustering Student Data to Characterize Performance Patterns," Int. J. Adv. Comput. Sci. Appl., pp. 138–140, 2010.
- [7] P. Gulati and A. Sharma, "Educational data mining for improving educational quality," Int. J. Comput. Sci. Inf. Technol. Secur., vol. 2, no. 3, pp. 648–650, 2012.
- [8] S. Ayesha, T. Mustafa, A. Raza Sattar, and M. I. Khan, "Data mining model for higher education system," Eur. J. Sci. Res., vol. 43, no. 1, pp. 24–29, 2010.
- [9] V. P. Breşfelean, "Analysis and predictions on students' behavior using decision trees in weka environment," in Proceedings of the International Conference on Information Technology Interfaces, ITI, 2007, pp. 51–56.
- [10] P. Cortez and A. Silva, "Using Data Mining To Predict Secondary School Student Performance," 5th Annu. Futur. Bus. Technol. Conf., vol. 2003, no. 2000, pp. pp. 5–12, 2008.
- [11] M. I. Al-Twijri and A. Y. Noaman, "A New Data Mining Model Adopted for Higher Institutions," Procedia Comput. Sci., vol. 65, no. Iccmit, pp. 836–844, 2015.
- [12] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," Comput. Educ., vol. 51, no. 1, pp. 368–384, 2008.
- [13] S. Aslam and I. Ashraf, "Data Mining Algorithms and their applications in Education Data Mining," Int. J., vol. 7782, pp. 50–56, 2014.
- [14] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Predicting Learning and Affect from Multimodal Data Streams in TaskOriented Tutorial Dialogue," Proc. 7th Int. Conf. Educ. Data Min., no. Edm, pp. 122–129, 2014.
- [15] S. J. Lee, Y. Liu, and Z. Popović, "Learning Individual Behavior in an Educational Game: A Data-Driven Approach," Proc. 7th Int. Conf. Educ. Data Min., no. Edm, pp. 114–121, 2014.