

Extension of Data Science Techniques in Education Prognosis

Barkha Bhardwaj^{1*}, Shivam Tiwari², Bramah Hazela³, Namrata Dhanda⁴

¹²³⁴Department of Computer Science & Engineering, Amity School of Engineering & Technology
Amity University, Uttar Pradesh, India

Abstract: *In today's world, the most critical problem being faced by the students while looking for the courses like which course is better among the available course for them and how can they say that this particular course is beneficial for them? Due to this dilemma, students usually go for a course in which either they regret later or drop the course in between. As a result, they face problem during their courses. Most of the time, student go according to the suggestions given by the peer, teacher, friends, family that ultimately creates confusion in the process of selecting the course and they make a random decision that affects their motivation to complete the course. So, there is a need for such a system that would provide them a better prediction about a course that would actually be beneficial for them and should also maintain their motivation to complete it. So, in this paper we're discussing data science and data mining techniques like Regression, Decision Tree, Classification, Clustering, etc. for the better prediction in higher education.*

Keywords: *Random decision, prediction, data mining, regression, classification, clustering.*

I. Introduction

Presently, nowadays the information mining device is extremely best in class for updating advanced education. There are numerous nations which are utilizing information mining systems in the field of training. In the instructive field, information mining assumes an essential job, presently in the advanced world information digging is helpful for us and this procedure is utilized in each field. By utilizing information mining strategies we will help the understudies who are keen on advanced education. The information mining procedures are considered as a benefit in the instructive field and the extent of this system is that we anticipate the best field for the understudies who are directly for the understudies. Since understudies don't realize which course is best for us and in this kind of circumstance, they select one course But they don't think about this course and they don't have any information this is the best field for him or not? They face numerous issues in advanced education. Furthermore, they can't make any progress in their life. So to deal with this circumstance, we can utilize information mining systems and strategies. Furthermore, by utilizing this kind of methods we can support the understudies.

By the information mining strategies, we can anticipate the best course for the understudy and after this forecast, the understudy can make progress. Here we utilize numerous procedures for anticipating the course. In this theme of the forecast, numerous kinds of research have been accumulated And this exploration depends on numerous characteristics, understudy information, similar to understudies' execution, intrigued subject of understudies with regards to which subjects understudies gain the most noteworthy imprint. In this, we will anticipate the course for understudies who are best for understudies. Here we will gather the past information of the understudies and uses numerous characterization strategies for the forecast. Here we utilize any kind of information mining procedures, similar to Regression, Decision Tree, Classification, and Clustering. By utilizing these kinds of strategies we will foresee the advanced education and by utilizing procedures, we can help the understudies for course choice. Which course is better for the understudies?

Also, in this report, I will talk about numerous kinds of research which are done on the past by utilizing numerous procedures. What's more, I will likewise examine the consequences of numerous past sorts of research and reference all in this paper. In the given figure 1, we are appearing administering information science procedures to conquer the forecast issues effectively. Apart from the study of data science algorithm, we need to scrutinize the machine learning methodologies, trends and also perspective of each learning mechanism [7].

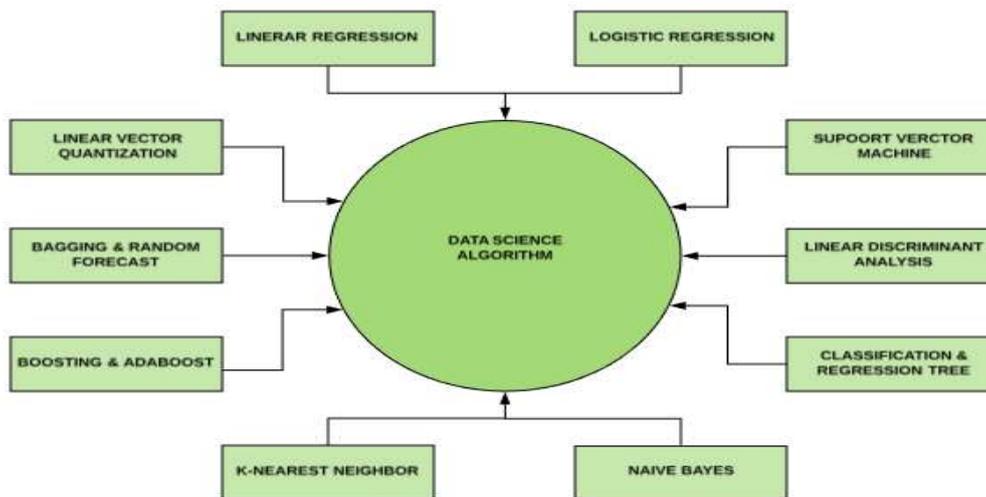


Figure 1. Data science algorithms

II. Literature Review And Discussion

In this paper we discussed researcher's paper. Firstly By researcher Nasiri and Minaei tells for higher education, research depends on two types of issues: GPA and LMS. Both types of algorithm are using data mining techniques. These data mining techniques were used for indicating the weak points and by using the association rules it solved problems because both algorithms were different from each other and gave the different-different results [1].

Bin the paper by Tovar and Soto. It is based on improvement of the predictive model. If a student is having difficulty in passing the course then instructors help the student. They also find capable, brilliant students having knowledge and but still failing to perform up to their capabilities. This paper is based on prediction, wherein they are trying to predict the student's performance using the data mining techniques [2].

By Knauf et al., In this model they collected the data of the students, like student progress report, performance of student and students personal information, weakness of student and after the collection of students' data, they used it to build a prediction model using data mining techniques. On the basis of this model, students can opt for a particular course and can successfully complete it [3].

By Ningning, this paper is based on data mining and data warehousing techniques [4]. According to the author, there are many students who drop a particular course because they lose interest in the course later, so author proposed a model to identify students having problem and students not having problem with the particular course. This model provides the guidelines to help the student facing any problem with the course just like an online instructor.

For higher education author used clustering algorithm. In this, selection of course process depends on the teacher. In this process faculty provides the guidelines and give instructions to the students for achieving success. And on the basis of teacher point of view, student selects courses. In this process the prediction is based on teacher's guidance to help students select suitable course and it's was proved to be successful. By this process students confidence increased, many changes in student's behaviour were observed and students gave good performance in their selected course [5].

In this paper many types of algorithm used for predicting on higher education. The algorithm is as follows- ID3, K-means, Random forest and Naive Bayes etc. In this paper, the author collected the data of students and which involved data of 10th class and 12th class data .Then using different mining techniques, we obtained different results with different accuracy. The accuracy on data set of size 200, ID3 algorithm gave accuracy of 75%, K-means gave accuracy of 54%, Naive Bayes gave accuracy of 59% and Random forest accuracy gave accuracy of 80%. In this result highest accuracy was of Random forest. So Random forest was proved to be the best algorithm among all above algorithm for predicting the course. By using this type of algorithm students can select the correct course on the basis of their skills and by this, students can make right decision for the bright future. The author applied many of techniques for the prediction of course. He used decision tree to predict the course for student. The author recorded of students' performance and created a data set is 150 rows and the accuracy of prediction on course was found to be 98.26%. So we can apply this method and predict the course for the students. The author discussed about classification method and regression method to predict the performance of the students and on the basis of students performance and marks scored in

different subjects predicting on the course to go for in life ahead. By using decision tree, clustering, regression method the results obtained were quite accurate. In this the author discussed about, students selecting the course on the basis of their opinion. In this process, selection of courses depends on the student marks, grade, and field of interest and so on. By using many types of method student can select one particular course. Author used neural networks to predict about things, like student performance, grades, and etc. It was proved to greatly help in predicting about the course student should select for in higher studies [10].

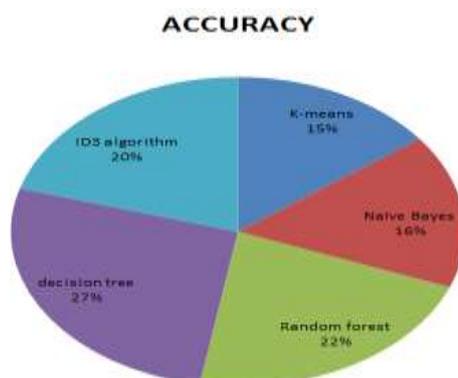


Figure 2 Accuracy of data science algorithm

III. Motivation

In this investigation, we learned different parts of information science methods that are utilizing in advanced education forecast framework [11]. The investigation produces different thoughts, properties that can be utilized to foresee the advanced education expectation framework. A choice tree is one of the inquisitive calculations of information science methods that are utilized by such a significant number of examine dependent on Educational information mining. By keeping all the part of investigates that have been done I figured out how to deal with the issue in fact and how to tackle the issue by utilizing a few calculations. I am additionally recommending information mining device WEKA through the investigation of information mining systems [8]. In the arrangement of study, there are different outcomes for different calculations that give the plan to discover a superior calculation that gives better precision result with the relative investigation of the calculations. This examination will help in further investigates and will dependably rouse me a great deal. Utilizing numerous calculation of information science we can tackle the issue [9]. Utilizing grouping technique will anticipate the best course. It is essential for us. There are numerous sorts of research which depend on the order strategy and giving suitable outcome and it is the best procedure for us. By these arrangement strategies, we figured out how to anticipate specific things. There are such a large number of calculations and which one is best for forecast and relapse strategy is additionally the best calculation. By this calculation, we can get different sorts of results. Step by step instructions to separate the normal esteem and by utilizing this strategy we take care of numerous issues in all respects effectively. This strategy is crucial to us. By utilization of this strategy, we can choose a superior field for the understudies'. Here in this paper information science techniques are useful for us. By this system of information mining, we can show signs of improvement result. We can get much specialized data. Information mining strategies are assuming an extremely vital job thusly. Step by step instructions to take care of the issue, and how we might show signs of improvement result and how to foresee future outcomes, information mining strategies tackled every one of the issues.

SL NO	LEARNING TECHNIQUE	USED DATASET	USED ALGORITHM	EXAMPLE
1	Supervised	Continuous and categorical	Regression, Classification	House price prediction and medical imaging etc.
2	Unsupervised	No target data	Clustering and Association	Customer segmentation and market basket analysis etc.
3	Semi-supervised	categorical	Classification and Clustering	Test classification and line-finding on GPS data etc.
4	Reinforcement	Categorical and no target data	Classification and Control	Optimized marketing and self-driving car etc.

Table 1 data science learning methodologies

In this paper Table 2, we're also going to discuss some prestige data science algorithm with some features like accuracy, training time, linearity, number of parameter required etc. based on the learning mechanism used by the data scientist. All these are based on types of learning mentioned in the above table.

SL NO.	DATA SCIENCE ALGORITHM	ACCURACY	TRAINING TIME	LINEARITY	PARAMETER
1.	logistic regression	AVERAGE	FAST	YES	5
2.	decision forest	EXCELLENT	MODARATE	NO	6
3.	decision jungle	EXCELLENT	MODARATE	NO	6
4.	boosted decision tree	EXCELLENT	MODARATE	NO	6
5.	neural network	EXCELLENT	SLOW	NO	9
6.	averaged perceptron	GOOD	MODARATE	YES	4
7.	support vector machine	AVERAGE	MODARATE	YES	5
8.	locally deep support vector machine	GOOD	SLOW	NO	8
9.	Bayes' point machine	AVERAGE	SLOW	NO	3

Table 2 data science algorithm details

IV. Conclusion

In this paper, we talked about numerous sorts of information science strategies and techniques for training forecast. This kind of counsel will help the understudies in foreseeing the courses. By utilizing these methodologies, understudies can ready to choose the better field and effectively get the method for progress. In the meantime, the understudy will give a superior act and fantastic criticism in their intrigued area. This is helpful and essential for those understudies, who can't ready to choose the best course. They were choosing incorrectly courses since they are not inspired by the area picked by them. By utilizing any of these approaches, they can pick a superior field. In this way, these methodologies will give a superior situation to the individuals who can't ready to take the correct choice for their future. These sorts of methodologies are vital for us and it'll assume a critical job in the forecast of an instructive area. By these systems, most likely we can improve the understudy's execution. We can expand the achievement rate of the understudies and improve the subsequent level of the understudies as needs are. Utilizing this framework, we can likewise diminish the rate of the disappointment of understudies and limited implosion cases. In this paper, after an investigation of the considerable number of inquiries about and procedures, we can say that this framework is exceptionally fundamental for us. We will likewise ready to utilize information digging procedures for the further improvement of the whole framework. This framework is imperative for anticipating the most appropriate courses for the understudies. In future degree, we'll endeavour to make sense of numerous other distinctive frameworks and diverse kinds of strategies that can give a strong establishment to the understudies.

References

- [1]. M. Nasiri and B. Minaei, "Predicting GPA and Educational Data Mining," IEEE Third International Conference on E-Learning and E-Teaching (ICELET), pp. 53-58, 2012
- [2]. R. Knauf, Y. Sakurai, K. Takada and S. Tsuruta, "Prediction Of Higher Education by Data Mining," IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 488-492, 2010.
- [3]. E. Tovar and O. Soto, "Predict the Performance of Students," IEEE Frontiers in Education Conference (FIE), pp. F3J-1, 2010.
- [4]. G. Ningning, "prediction of higher education," International Forum on Information Technology and Applications, vol. 1, pp. 436-439, 2010.
- [5]. F. Wu, "Apply Data Mining to Students," IEEE Education Technology and Computer Science (ETCS), vol. 1, pp. 606-609, 2010.
- [6]. Kardan, A. Ahmad., Sadegh, Hamid, Ghidary, Shiry.Saed., and Sani, Reza Fani., Mohammad. (2013) "Prediction of student course selection in online higher education institutions" 65: 1-11.
- [7]. M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, pp. 255-260, 2015.
- [8]. E. Osmanbegović, M. Suljić, and H. Agić, "DETERMINING DOMINANT FACTOR FOR STUDENTS PERFORMANCE PREDICTION BY USING DATA MINING CLASSIFICATION ALGORITHMS," Tranzicija, vol. 16, pp. 147-158, 2015.
- [9]. A. M. Shahiri and W. Husain, "A review on predicting student's performance using data mining techniques," Procedia Computer Science, vol. 72, pp. 414-422, 2015.
- [10]. M. Zaffar, M. A. Hashmani, and K. Savita, "Performance analysis of feature selection algorithm for educational data mining," in Big Data and Analytics (ICBDA), 2017 IEEE Conference on, 2017, pp. 7-12.
- [11]. S. Hussain, N. A. Dahan, F. M. Ba-Alwi, and N. RIBATA, "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA," Indonesian Journal of Electrical Engineering and Computer Science, vol. 9, 2018.