A Review of Analysis And Prediction Of Student Data Using Data Science

Aasawari Deshmukh, Mayuri Doye, Abhishek Mandeogade, Nikhil Nidhan,
Abhishek Khapekar, Dr. S.V. Solanki

Dept. of Information Technology
KDKCE, Nagpur, India, 440009

Abstract – Data science is a multidisciplinary blend of data inference, algorithm development and technology, used for analyzing and predicting the result. In this, data science is used for analyzing and data mining techniques are used for predicting the student performance which is present in large form of data in databases. Clustering and visualization are applied on data for cleaning the improper dataset. This is how the prediction algorithm can be used to identify the most important attributes in a student's data. Here improvement in student's achievement and success in a more efficient way using data mining. It will bring an impact and benefits on students' performance and college.

Keywords- Academic performance, Data Science, Data Mining, information, data, visualization, Prediction

INTRODUCTION

Basically, data science refers to the analysis and prediction of data collected from database of college. Student's academic result are affected by many factors like personal, social, and lack of knowledge about the technology and many more [5]

The main aim of every college is to provide higher academics marks and learn technologies regarding their fields. But nowadays, it's become very difficult to find the strength and weaknesses of the student because of large number of databases are present in institutes. And to solve this problem data science and data mining techniques are used on data sets of student's academic data for analysis and predicting their result. The best way to achieve the highest performance of student first analyses which factors affect the students' performance.

Data mining techniques may contain different algorithms like decision tree, k-nearest neighbor, regression etc. By using this technique, many types of knowledge can be discovered like association rules, clustering, classifications, and outlier detection [3]. There are many kinds of knowledge and information can be discovered from the data.

From the discovered knowledge, there is need to provide a college management with a helpful manner to overcome the problem of low grades of students [3], and to improve students' academic performance.

LITERATURE SURVEY

Bhavesh Patel in his paper [1] "Student Performance Analysis Using Data Mining Technique" Data mining techniques are implemented for standard procedure for analyzing student's performance from large volume of available data, extracting useful information and knowledge to support the decision-making processes. In this, various algorithm is applied on the data for predicting and performance of student among the selected algorithms. The analysis work is done by considering various types of algorithm like decision tree algorithm, rule-based algorithm.

Mohd Sharifuddin Ahmad [2] "Analyzing Students Records to Identify Patterns of Students' Performance" In this, the data mining technique used to identify the significant variables that affects and influences the performance of undergraduate students. Students' recent and past academic performance data are then used to study the academic pattern. Consequently, the ability to predict a student's performance could be useful in many ways to stakeholders of higher education institutions.

Impact Factor Value 4.046   **e-ISSN: 2456-3463**
National Conference on "Recent Advances in Engineering and Technology" SAMMANTRANA 19
Organized by Government College of Engineering, Nagpur
International Journal of Innovations in Engineering and Science, Vol 4 No.8, 2019
www.ijies.net

Mohammed M. Abu Tair [3] "Mining Educational Data to Improve Students' Performance: A Case Study". In this paper, the study of the educational data mining is used to improve graduate student's performance and overcome the problem of low grades of students and reason of their low grades. Here data mining techniques to discover association, classification, clustering and outlier detection rules.

Proyag Pal [4] in his paper "Challenges in Data Science: A Comprehensive Study on Application and Future Trends" Data Science refers to an area of work concerned with the collection, preparation, analysis, visualization, and preservation of large collections of information. Data science is much more than the analyzing of data. In data science prediction of the result is also done. Data scientists play an active role in the design and implementation work of four related areas such as data architecture, data acquisition, data analysis and data archiving.

Nawal Ali Yassein [5] in his research article "Predicting Student Academic Performance in KSA using Data Mining Techniques" In this research work it is found that there are any patterns in the available data (student and courses records) that are useful for predicting students' performance. But the prediction of actual performance of the students is the most challenging task. Data mining techniques are applied on the data for classifying and sort the data into proper form for analysis. In this classification and reduction features are used for reducing error rate.

Dorina Kabakchieva [6] "Predicting Student Performance by Using Data Mining Methods for Classification" in his paper Data mining methods are implemented at advanced universities today for analyzing available data and extracting information and knowledge to support decision-making and make a research on a university for high potential of data mining features.

METHOLOGY

Data Mining Techniques:

Data mining is a computational method of processing data to extract useful data from the large database. Data mining techniques are used to build a model to identify new knowledge information after extracting of useful data. There are several major data mining techniques have been developed and used like association, classification, clustering, prediction, and decision tree [5]. The following is the description of main techniques used in the analysis of student's performance:

Before applying the data mining techniques on the student's data set, there should be a methodology that governs the work. Figure 1 depicts the work methodology used in this paper. The methodology starts from the problem definition, then collecting data from college, then welcome to the data mining methods which are association, classification, clustering, and outlier detection, after that results are evaluated and finally the knowledge representation process [4].
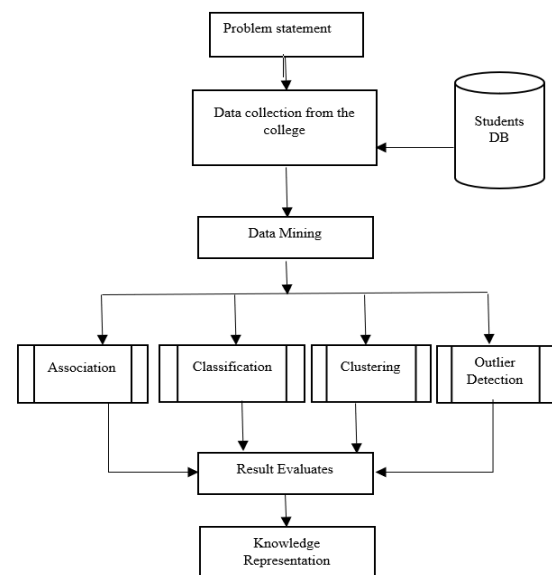

Fig. 1- **:** Data Mining Work flow

A. Classification

Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of a predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics [5].

B. Clustering

Clustering is a data mining technique which finds the collections of objects so that objects that belong to one cluster are more like each other than to objects belonging

Impact Factor Value 4.046   **e-ISSN: 2456-3463**
National Conference on "Recent Advances in Engineering and Technology" SAMMANTRANA 19
Organized by Government College of Engineering, Nagpur
International Journal of Innovations in Engineering and Science, Vol 4 No.8, 2019
www.ijies.net

to different cluster. In this the main objective of clustering is to find the high-quality clusters means that inter cluster distances are maximized and inter cluster distances are minimized. In clustering k-means algorithm is used for finding the best clusters from the data and make as a centroid [3].

C. Association Rules

Mining association rules searches for interesting relationships among items in a given data set. As part of association method, FP- Growth algorithm is applied to the data set. Mining association rules searches for interesting relationships among items in a given data set [5].

D. Outlier Detection

Outlier detection detects the data points which are different from the other rest of the points. In this outlier is used for analyzing the outliers in student's data for detecting students learning problem. There are two outlier's detection methods are present for discovering the outlier [3].

Table 1 shows the attributes from the student data set and its description.

Table 1: The Student's Dataset Description

| Attribute	Description
Student_ID	The student ID
Student_Name	The name of the student.
Gender	The gender of the student; male or female.
Date_of_Birth	The date of birth for the student.
Place_of_Birth	The Place of birth for the student.
Speciality	The specialty of the student.
Enrollment_Year	The year of enrollment in the college.
Graduation_Year	The year of graduation from the college.
City	The City for the student.
Location	The location in the city
Address	The address of the student.
Telephone_Number	The Telephone number for the student.
Matriculation_GPA	GPA for the student in the Matriculation.
Secondary_School_Type	The type of the Matriculation; Adabi or Ilmi
Matriculation_Obtained_Place	The place of obtained Matriculation; in Palestine or other country.
Matriculation_Year	The obtained year of the Matriculation.
College_GPA	GPA for the student in the College.
Grade	The grade for the student; Excellent, Very good, Good, or Average

PROPOSED RESEARCH MODEL
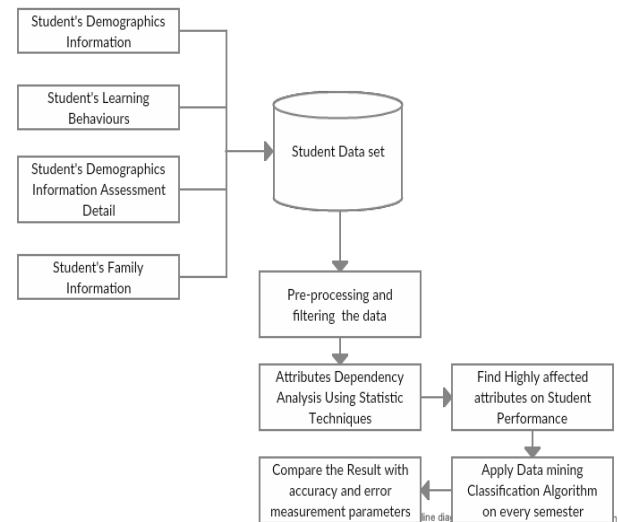
Fig.2: Research model

Step-wise procedure for Implementation Model:

Step 1: Collect the student's (Demographic Information, Academic Information, learning Behavioral Information) from the college.

Step 2: After collecting the data set pre-processing and filtering the data.

Step 3: Perform the data mining techniques (Classification, Clustering, Association rules) on student data set.

Step 4: After classifying the data it is show in the graphical form and gives the count of the students.

Step 5: Perform the data science on the student data set for analyzing their performance from the generated knowledge.

Step 6: After analyzing the performance, have to the predict that in which field or subject students are strong or weak by finding the most optimized model from knowledge[1].

OBJECTIVE OF RESEARCH MODEL

- The result from this study is used for identifying the factors affecting students' academic performance.

- In addition, the prediction is done for finding the student's weakness and strength in their field.

- In this way, students who do best can be pushed to achieve them an excellent level in every field of college.

- On the other hand, students who are low grade achievers could be assisted to gain better performance in their academics.

- This research is to ensure the quality of students is another part or progress in their performance in a positive direction [1].

CONCLUSION

This paper reviews the analysis and prediction of student's academic performance by applying various data mining techniques like classification, clustering, association rules, visualization. Here data mining technique is applied for discovering the knowledge from the large volume of database. In this the various algorithms are used for classifying and visualize the data. Data science is used for the analysis and prediction of student's academic performance to find out in which field or subject students are weak or strong.

REFERENCES

[1] *Bhavesh Patel, Chetan Gondaliya, "Student Performance Analysis Using Data Mining Technique", IJCSMC, Vol. 6, Issue. 5, May 2017, pg.64 – 71*

[2] *Alan Cheah Kah Hoe, Mohd Sharifuddin Ahmad, Tan Chin Hooi, Mohana Shanmugam, Saraswathy Shamini Gunasekaran, Zaihisma Che Cob, Ammuthavali Ramasamy," Analyzing Students Records to Identify Patterns of Students' Performance",*

[3] *Mohammed M Abuteir, Alaa Mustafa El-Halees, "Mining Educational Data to Improve Students' Performance: A Case Study", International journal of information and communication technology research, Volume 2 no.2, 2 february 2012*

[4] *Proyag Pal, Triparna Mukherjee, Dr. Asoke Nath," Challenges in Data Science: A Comprehensive Study on Application and Future Trends,"*

[5] *Nawal Ali, Yassei, Rasha, Gaffer, M Helali and Somia B Mohomad," Predicting Student Academic Performance in KSA using Data Mining Technique", Journal of Information Technology & Software Engineering, Yassein et al., J Inform Tech Softw Eng. 2017, 7:5, DOI: 10.4172/2165-7866.1000213*

[6] *Dorina Kabakchieva," Predicting Student Performance by Using Data Mining Methods for Classification", CYBERNETICS AND INFORMATION TECHNOLOGIES, Volume 13, No 1*

[7] *Chandra, E. and Nandini, K. (2010), 'Knowledge Mining from Student Data', European Journal of Scientific Research," vol. 47, no. 1, pp. 156-163.*

[8] *Mansur, M. O., Sap, M. and Noor, M. (2005) "Outlier Detection Technique in Data Mining: A Research Perspective", In Postgraduate Annual Research Seminar.*

[9] *Dekker, G., M. Pechenizkiy, J. Vleeshouwer s, "Predicting Students Drop Out: A Case Study. – In: Proceedings of 2nd International Conference on Educational Data Mining "(EDM'09), 1-3 July 2009, Cordoba, Spain, 41-50.*

[10] *Shannaq, B., Rafael, Y. and Alexandro, V. (2010), "Student Relationship in Higher Education Using Data Mining Techniques", Global Journal of Computer Science and Technology, vol. 10, no. 11, pp. 54-59.*