Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 9, Issue. 9, September 2020, pg.38 – 42

Student Performance Prediction using Machine Learning

Mrs. Aksheya Suresh; Bala Subramaniyan S; Eswar Kumar R; Gokulkumar N
[1]Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Thandalam, Chennai-602105, INDIA
[2,3,4]UG Students, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Thandalam, Chennai-602105, INDIA
DOI: 10.47760/IJCSMC.2020.v09i09.004

Abstract: This paper focuses on improving student performance prediction, based on their performance characteristics. Due to various distractions, students may divert from their actual track. This might even lead to a course drop out. Predicting students' performance will help in self-analysis. The dataset used consists of data about students' performance from the academic and other classroom activities during the course time. Educational data mining algorithms are used to predict student performance which is a module in automated intelligent education systems. EDM is a methodology which is used to mine valuable information and patterns or forms from a massive educational database. Subsequently, the student's performance is predicted from the obtained useful information and patterns. The aim of our study is to discover the performance of students using some classification techniques and discovering the best one which yields optimal results. The result of this study is extremely significant and hence provides a greater insight for evaluating the student performance and underlines the significance of data mining in education. It also shows that how students attributes affect the student performance.

I. INTRODUCTION

Discovering Knowledge from huge Databases is known as Data mining. It discovers hidden information from diverse data sources pertaining to diverse fields. Several techniques can be employed in different fields of data mining together with weather forecasting, oil research, business, medical, marketing and EDM etc. To extract and analyze the knowledge present in educational data sources, a sub domain of data mining has also been developed termed as Educational data mining (EDM). Data mining, statistics and machine learning are applied on EDM data to derive knowledge from educational environments. Currently, it is in demand and gaining more attention because of increase in the educational data of c learning systems, and even progressing traditional education. Alarmed with evolving techniques for discovering the distinctive types of data present in scholastic environments ,It seeks to extract meaningful information in order to advance and appreciate learning processes from vast amounts of raw data. Probing traditional records of database can offer answer to Problems such as "find the students who failed the examinations", whereas EDM offers answers to additional problems like "predicting the students who are more likely to pass".

II. EXPLORATORY DATA ANALYSIS (EDA)

To study and observe the behaviour of data, attributes ,relationship between attributes and target variable are graphically visualized ,so that pattern of data is keenly studied and to explore the dependency and weightage of attribute so as to extract reliable features to develop a reliant model with robust features. The primary plots that are visualized:

- Individual attributes plot against its frequency [to observe the data].
- Attributes against target variable('FinalGrade')[to study the dependency and weightage]

III. LITERATURE SURVEY

Educational data mining in short EDM is widespread nowadays due to increase in e-resources, usage of online tools for education and Internet. Lots of research is taking place to make best of education tools and technologies.

[1]. The usage of EDM techniques to predict or analyze the students' performance and improve the students who are falling below satisfactory grades, an Artificial neural network classifier model was built which can be beneficial for both students and teachers to discover knowledge from huge data present in educational sector. Sentiment analysis was carried on to understand the different way of students learning and their plan of study in order to improve teaching.

[2]. Student's behavioral features were considered with other features and a model was proposed based on data mining techniques which yielded 22.1% high accuracy after removing behavioral features. Further, by employing ensemble methods there was found 25.8%increase in accuracy.

[3]. Academic data set consisting of 473 instances, and found that 70% accuracy was yielded by Bayesian classifier. The naive Bayes classifier, ANN, KNN and j48 were used to categorize student's dropouts. 87% and 79.7% accuracy was yielded by K-nearest neighbors and decision trees applying 10-fold cross-validation.

[4]. SVM's separates the classes in high-dimension space by constructing hyper plane. Data mining techniques like Logistic regression and Multi-classifiers produced outstanding results on data related to health.

[5]. Combining this concept and speeding up the time of training explains their widespread use in EDM. Decision Tree was used to predict classes pass or fail on a dataset of 15150 instances, 85.92% accuracy was attained.

IV. PROBLEM STATEMENT AND OBJECTIVE

1. Problem Statement

Predicting grade of students for Maths subject based on data collected from two portugues schools. This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in maths subject.

2. Objectives

The main objectives are:

- The massive growth in the educational sector needs to create awareness about handling the huge volume of student data.
- To provide the personal details of the student
- Nowadays educational data mining technique plays a vital role in predicting academic performance.
- The educational data mining is a technique is used to extract information from these volumes of data.

V. PROPOSED SYSTEM

1. Data Preprocessing

Both datasets are processed to check 1. Null Values 2. Duplicates & 3. Invalid values but fortunately there are no such irregularities in the dataset, implying that data is already clean and processed. Since both datasets have same set of attributes and has similar kind of data, the both datasets are merged vertically, so as to make the dataset effective and increase the dataset, this process supposedly key aspect of data preprocessing.

2. Modification of Dataset

A new column called 'FinalGrade' is asserted which reflects 'Grade3' and a broader level view of 'Grade3'. The column is inferred broadly as five categories under given conditions as follows: ● 'Excellent'[(data.G3 >= 18) & (data.G3 <= 20)] ● 'Good' [(data.G3 >= 15) & (data.G3 <= 17)] ● 'Satisfactory' [(data.G3 >= 11) & (data.G3 <= 14)] ● 'Poor' [(data.G3 >= 6) & (data.G3 <= 10)] ● 'Failure' [(data.G3 >= 0) & (data.G3 <= 5)]

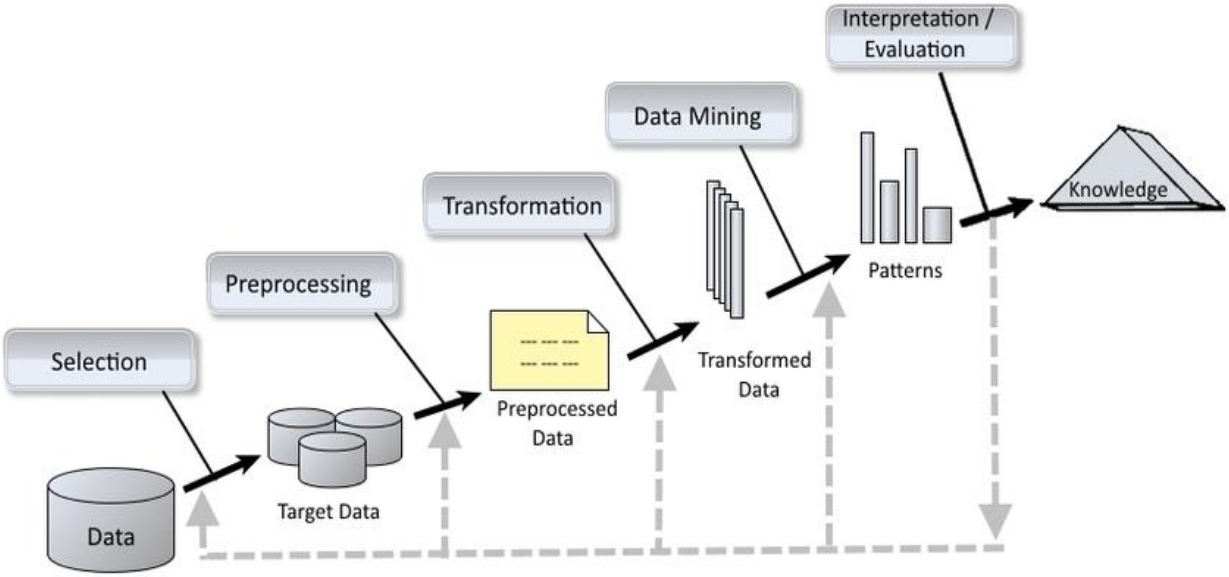
3. Encoding

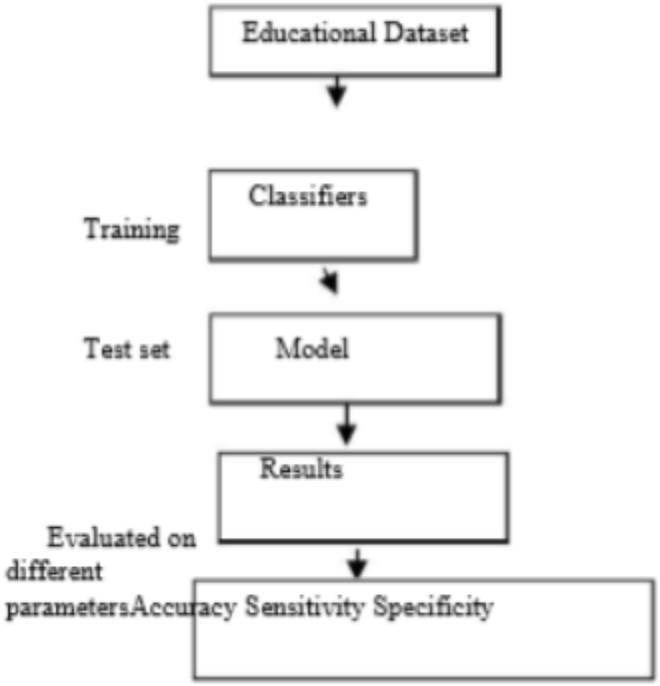
There several attributes which are non-numeric and categorical, so they must be encoded, because model cannot deal with non-numeric attributes. But as observed, the dataset has only few categories(max of 5) for every column, so best assumption would be one-hot-encoding which was our primary choice as categories are extracted as features which intensifies the effect of feature.

4. Exploratory Data Analysis(EDA)

To study and observe the behaviour of data, attributes ,relationship between attributes and target variable are graphically visualized ,so that pattern of data is keenly studied and to explore the dependency and weightage of attribute so as to extract reliable features to develop a reliant model with robust features.

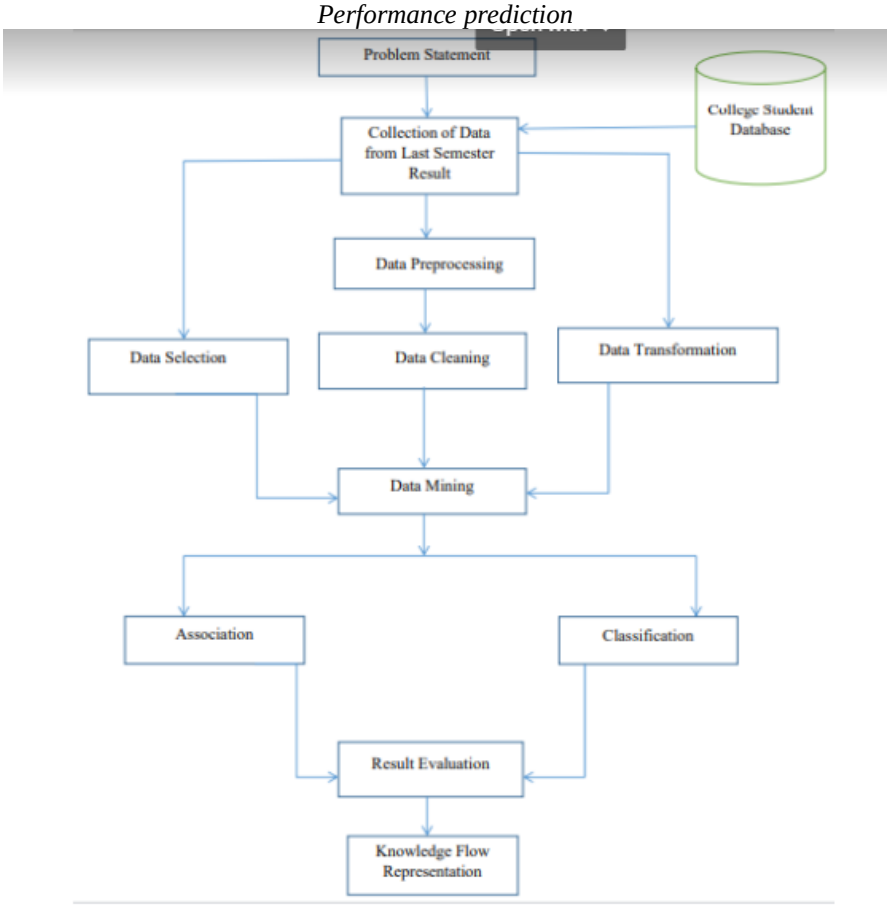
VI. ARCHITECTURE DIAGRAM





VII. FLOW CHART REPRESENTATION

Performance prediction



VIII. CONCLUSION

Predicting students performance is mostly useful to help the educators and learners improving their learning and teaching process. This paper has reviewed previous studies on predicting students performance with various analytical methods. Most of the researchers have used calculate average grade and internal assessment as data sets. While for prediction techniques, the classification method is frequently used in educational data mining area. Under the classification techniques, Neural Network and Decision Tree are the two methods highly used by the researchers for predicting students performance. In conclusion, the meta-analysis on predicting students performance has motivated us to carry out further research to be applied in our environment. It will help the educational system to monitor the students performance in a systematic way

REFERENCES

[1]. Pena A. Educational data mining: A survey and a data mining based analysis of recent works. Expert systems with applications, vol.41 (4), 1432–1462, 2014.

[2]. Livieris, et al. Predicting students performance using artificial neural networks, 8th PanHellenic conference with International participation Information and communication technologies, pp.321-328, 2012.

[3]. Sultana et al. Prediction of Sentiment Analysis on Educational Data based on Deep Learning Approach.21st Saudi Computer Society National Computer Conference (SCSNCC). IEEE, 2018.

[4]. Amrieh et al. Mining Educational Data to Predict Student's academic Performance using Ensemble Methods, International Journal of Database Theory and Application, Volume 9, issue 8, pp.119-136, 2016.

[5]. Yukselturku et al. Predicting dropout student: an application of data mining methods in an online education program. European Journal of Open, Distance and eLearning, vol.17(1),pp.118- 133, 2014.

