

EVALUATION OF EDUCATIONAL DATA USING MACHINE LEARNING

Surabika Hota - Faculty at Gandhi Institute For Technology, CSE Dept.
(Affiliated to Biju Patnaik University of Technology)

Bhagaban Swain - Student at Gandhi Institute For Technology, CSE Dept.
(Affiliated to Biju Patnaik University of Technology)

Sambit Kumar Nayak - Student at Gandhi Institute For Technology, CSE Dept. (Affiliated to Biju Patnaik University of Technology)

Saswat Kumar Behera - Student at Gandhi Institute For Technology, CSE Dept. (Affiliated to Biju Patnaik University of Technology)

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.

Keywords: — *Student performance, Machine Learning, RF.*

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 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 e 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”.

Objective

- The objective of this project to build a model which predict the average score Of a student by learning the trained data.
- The logic behind the code is to train machine by giving large number of data Of a student using Random-Forest Algorithm.
- We are using supervised machine-learning, because in supervised learning training data provided to the machine work as the supervisor that teaches the Machine to predict the output correctly.

II.EXPLORATORY DATA ANALYSIS (EDA)

To study and observe the behavior 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:

III. LITERATURE SURVEY

Machine learning (ML) is the study of computer algorithm that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email, filtering, speech recognition, and computer vision , where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics.

IV. Methodology:

Tis section describes the details of the dataset, pre-processing techniques, and machine learning algorithms used in this project.

```
[ ] df.head()
```

```
[ ] from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=1/4, random_state=42)
```

▾ bulid Model

```
[ ] from sklearn.ensemble import RandomForestRegressor
model=RandomForestRegressor()
```

▾ Fit the training data set to the model

```
[ ] model.fit(x_train, y_train)
```

IV.1. Dataset

Educational institutions regularly store all data that are available about students in electronic medium. Data are stored in databases for processing. These data can be of many types and volumes, from students' demographics to their academic achievements. In this study, the data were taken from the Student Information System (SIS), where all student records are stored at an Odisha state Government Portal.

The dataset contains Gender, Maths score, writing score, reading score, Lunch, ethnicity, Level of Education.

IV.2. Data Identification & Collection:

At this phase, it is determined from which source the data will be stored, which features of the data will be used, and whether the collected data is suitable for the purpose. Feature selection involves decreasing the number of variables used to predict a particular outcome. The goal; to facilitate the interpretability of the model, reduce complexity, increase the computational efficiency of algorithms, and avoid overfitting.

IV.3. Establishing DM model & Implementation of Algorithms

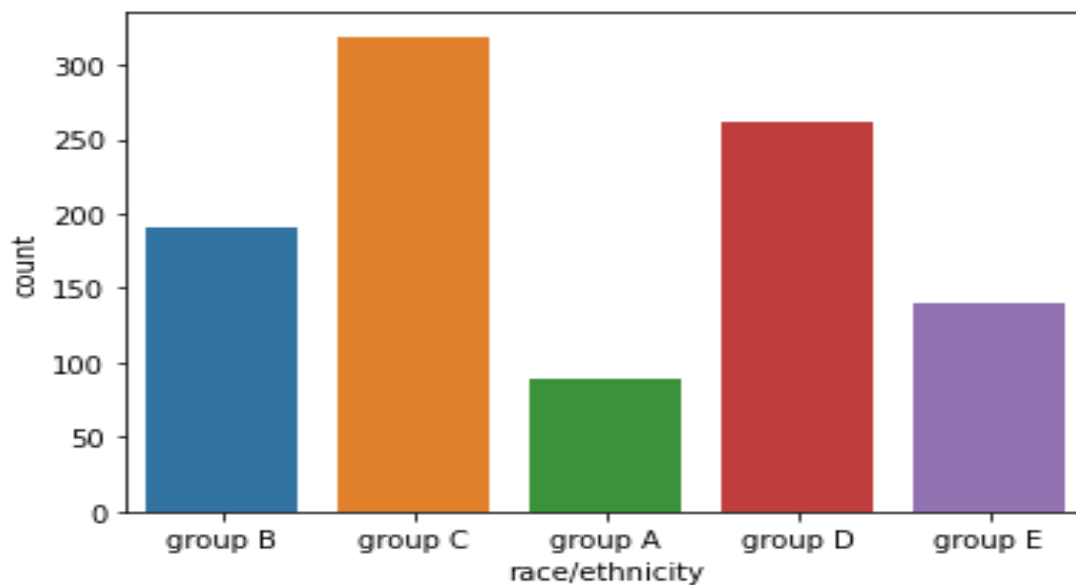
RF, LR, SVM and NB were employed to predict students' academic performance. The prediction accuracy was evaluated using tenfold cross validation. The DM process serves two main purposes. The first purpose is to make predictions by analyzing the data in the database (predictive model). The second one is to describe behaviors (descriptive model). In predictive models, a model is created by using data with known results. Then, using this model, the result values are predicted for data sets whose results are unknown. In descriptive models, the patterns in the existing data are defined to make decisions.

IV.4. Spilt the data and set in to Train and test dataset and Build Model

IV.4. Evaluation Of the Model Performance

performance of model was evaluated with confusion matrix, classification accu-racy (CA), precision, recall, f-score (F1), and area under roc curve(AUC) metrics.

V. GRAPH:



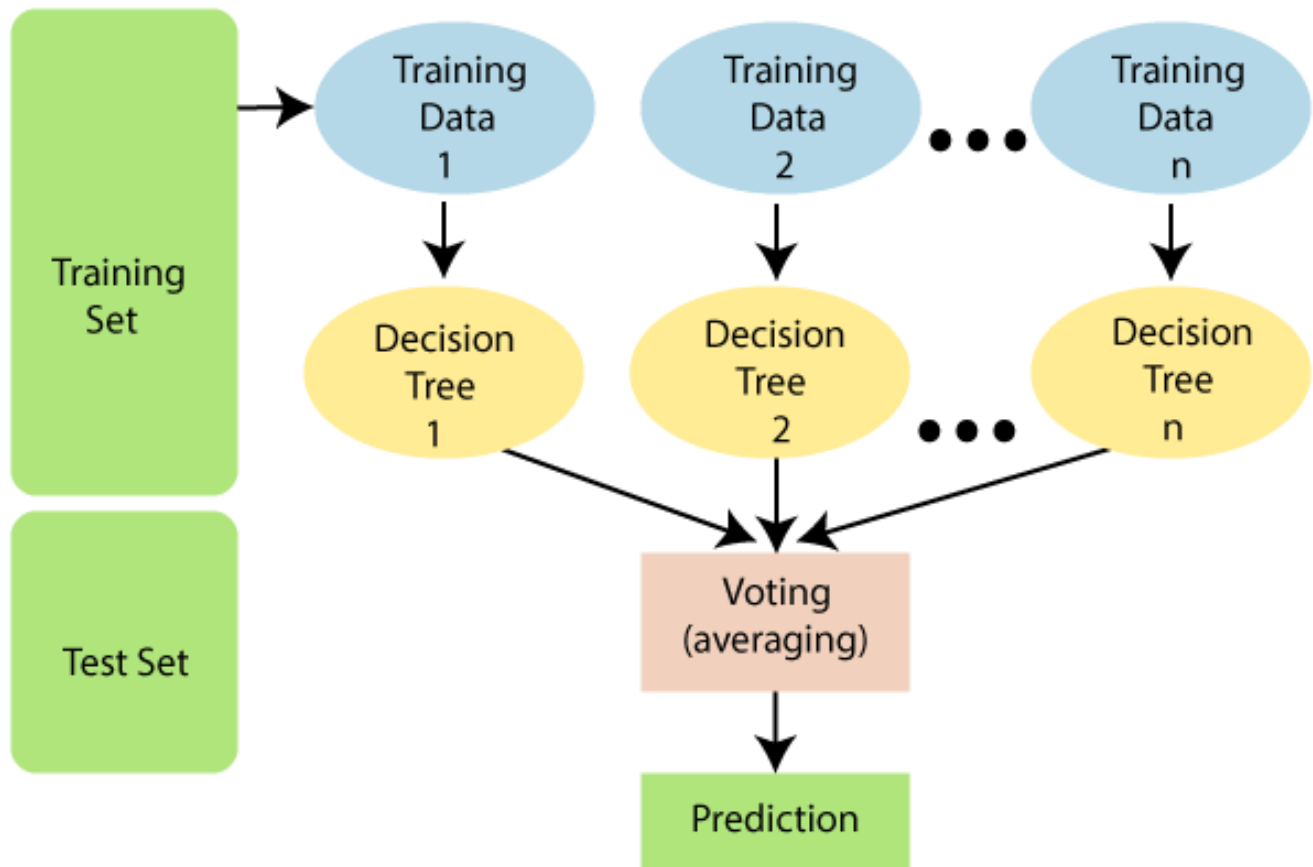
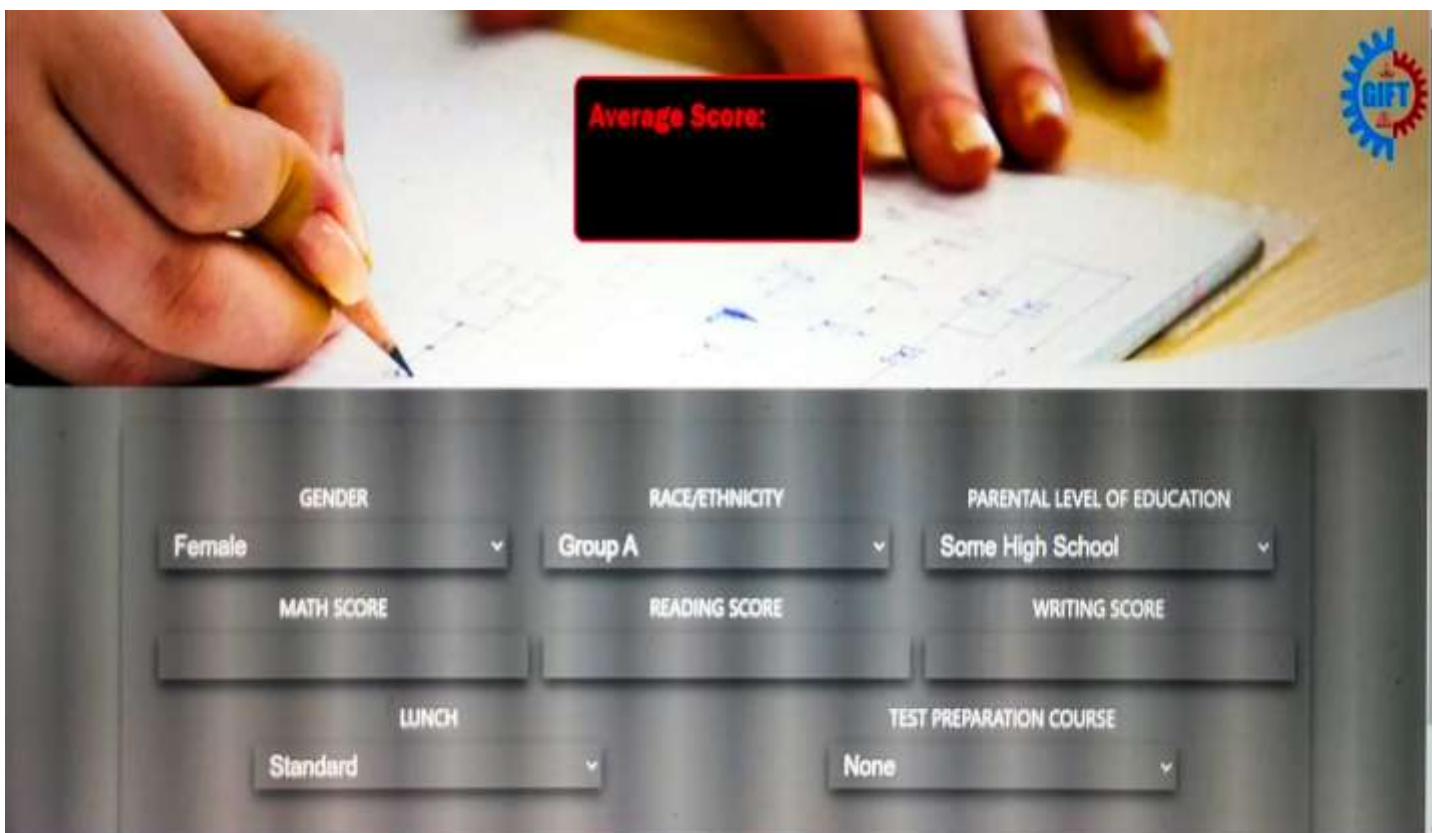
VI. ARCHITECTURE DIAGRAM

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	average_score
0	female	group B	bachelor's degree	standard	none	72	72	74	72.666667
1	female	group C	some college	standard	completed	69	90	88	82.333333
2	female	group B	master's degree	standard	none	90	95	93	92.666667
3	male	group A	associate's degree	free/reduced	none	47	57	44	49.333333
4	male	group C	some college	standard	none	76	78	75	76.333333
...
995	female	group E	master's degree	standard	completed	88	99	95	94.000000
996	male	group C	high school	free/reduced	none	62	55	55	57.333333
997	female	group C	high school	free/reduced	completed	59	71	65	65.000000
998	female	group D	some college	standard	completed	68	78	77	74.333333
999	female	group D	some college	free/reduced	none	77	86	86	83.000000

1000 rows x 9 columns

```
[ ] from sklearn.metrics import r2_score
    print(r2_score(predictions, y_test))
```

0.9938159834326161

**OUTPUT DIAGRAM:**

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 student's 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.