Student Performance Analysis Behavior Using Unsupervised Approach

M. Srikanth[1],Guntaka Sai pramidha[2],Gurram Svethalina[2], Kuppa Naga Soumya[2]

[1]Professor, Department of ECE, Malla Reddy Engineering College for Women, Hyderabad, Telangana, India.
[2]UG Scholar, Department ofECE, Malla Reddy Engineering College for Women, Hyderabad, Telangana,India.

Abstract:

Performance analysis of outcome based on learning is a system which will strive for excellence at different levels and diverse dimensions in the field of student's interests. This system developed to analyze and predict the student's performance only. The proposed framework analyzes the students' demographic data, study related and psychological characteristics to extract all possible knowledge from students, teachers and parents. Seeking the highest possible accuracy in academic performance prediction using a set of powerful data mining techniques. The framework succeeds to highlight the student's weak points . The realistic case study that has been conducted on 200 students proves the outstanding performance of the proposed framework in comparison with the existing ones.

Existing System:

The previous predictive models only focused on using the student's demographic data like gender, age, family status, family income and qualifications. In addition to the study related attributes including the homework and study hours as well as the previous achievements and grades. These previous work were only limited to provide the prediction of the academic success or failure, without illustrating the reasons of this prediction. Most of the previous researches have focused to gather more than 40 attributes in their data set to predict the student's academic performance. These attributes were from the same type of data category whether demographic, study related attributes or both, that lead to lack of diversity of predicting rules.

Disadvantage:

- As a result, these generated rules did not fully extract the knowledge for the reasons behind the student's dropout.
- Apart from the previously mentioned work, there were previous statistical analysis models from the perspective of educational psychology that conducted a couple of studies to examine the correlation between the mental health and the academic performance.
- The type of the recommendations was too brief, they missed illustrating the methodologies to apply them.

Proposed System:

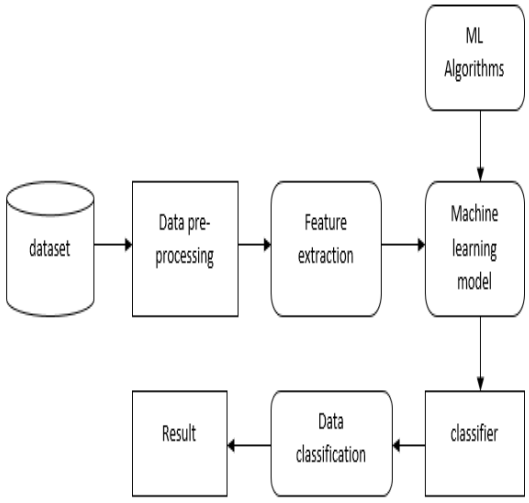
The proposed framework firstly focuses on merging the demographic and study related attributes with the educational psychology fields, by adding the student's psychological characteristics to the previously used data

set (i.e., the students' demographic data and study related ones). After surveying the previously used factors for predicting the student's academic performance, we picked the most relevant attributes based on their rationale and correlation with the academic performance.

Advantage:

- The proposal aims to analyze student's demographic data, study related details and psychological characteristics in terms of final state to figure whether the student is on the right track or struggling or even failing. In addition to extensive comparison of our proposed model with the other previous related models.

System Architecture:



System Requirements:

Software:

- Python
- Anaconda navigator

Hardware :

- Windows 7,8,10(64 bit)
- RAM 3GB
- Raspberry pi 3 B+

Raspberry Pi

Raspberry Pi is a small single-board Computer developed in UK by Raspberry Pi foundation to promote the teaching of computer science in schools and in developing countries.

Original model become far more popular than anticipated sealing outside of its target market, for uses such as robots.

Processor

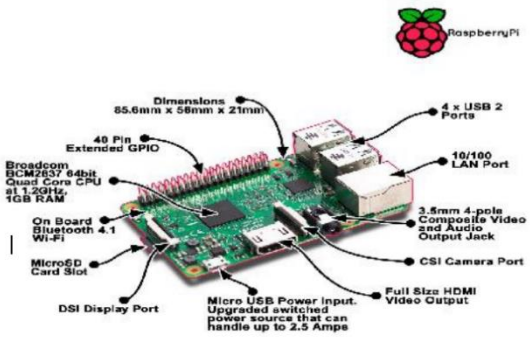
The processor at the heart of the Raspberry Pi is a Broadcom BCM28XX.

This is the Broadcom System on Chip (SOC) chip use in the Raspberry Pi. The processor from first to third generations include:

- Raspberry Pi 1: Broadcom BCM2835 SOC with 700MHz CPU speed, L2 cache of 128kb with ARM compatibility AR1176JZF-S (ARMv6) 32-bit RISC ARM.
- Raspberry Pi 2: Broadcom BCM 2836 SOC with 900MHz CPU speed, L2 cache of 256kb with 32-bit quad-core ARM cortex-A7 (ARMv7).

- Raspberry Pi 3: Broadcom BCM2837 SOC with 1.2GHz 64-bit quad-core –A53 with 512 kb shared L2 cache (64-bit instruction set ARMv8).

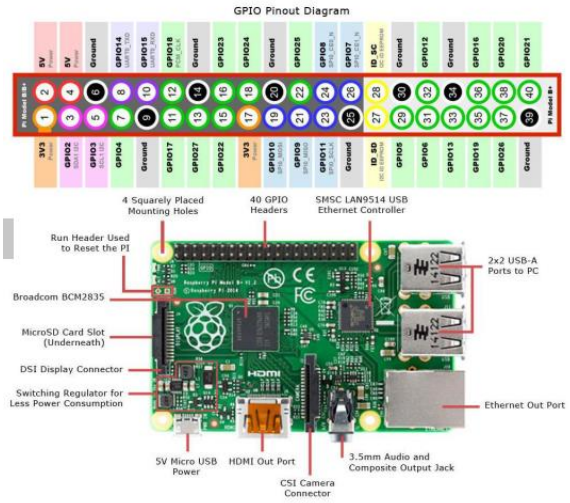
Raspberry Pi 3



In this project we are using Latest version of Raspberry i.e. Raspberry Pi 3The processor at the heart of the Raspberry Pi 3 is a Broadcom BCM2837, and the later models of the Raspberry Pi 2. The underlying architecture of the BCM2837 is identical to the BCM2836. The Only significant difference is the replacement of the ARMv7 quad core cluster with quad-core ARM Cortex A53 (ARMv8) cluster.

The ARM cores run at 1.2GHz, making the device about 50% faster than the Raspberry Pi 2 with a on board memory of 1GB RAM. The videocoreIV runs at 400MHz.

One powerful feature of the Raspberry Pi is the row of GPIO (general purpose input/output) pins along the top edge of the board. These pins are a physical interface between the Pi and the outside world. At the simplest level, you can think of them as switches that you can turn on and off (output).Of the 40 pins, 26 are GPIO pins and the others are power or ground pins (plus two ID EEPROM pins).



Modules

- **DATA COLLECTION**
- **DATA PRE-PROCESSING**
- **FEATURE EXTRATION**
- **EVALUATION MODEL**

DATA COLLECTION

Data used in this paper is a set of student details in the school records. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called *labelled data*.

DATA PRE-PROCESSING

Organize your selected data by formatting, cleaning and sampling from it.

Three common data pre-processing steps are:

1. **Formatting**
2. **Cleaning**
3. **Sampling**

Formatting: The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

Cleaning: Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonym zed or removed from the data entirely.

Sampling: There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

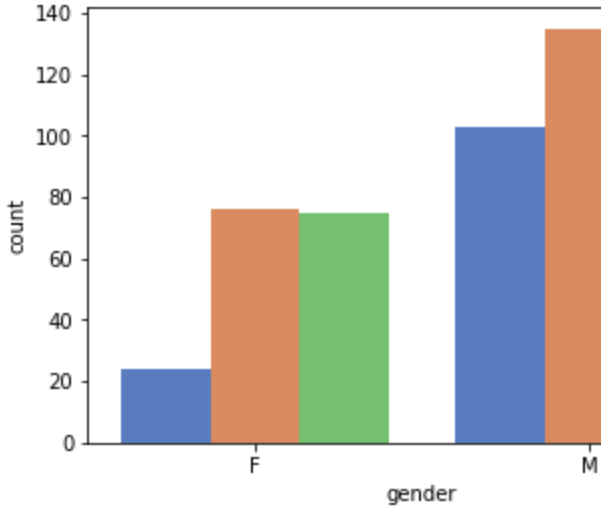
FEATURE EXTRATION

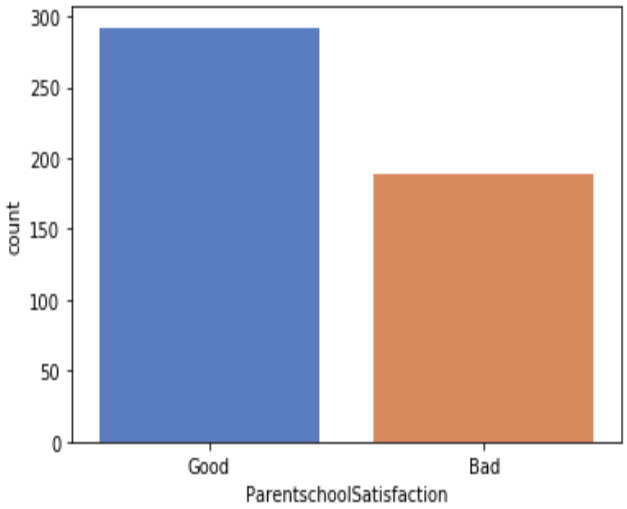
Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Finally, our models are trained using Classifier algorithm. We use classify module on Natural Language Toolkit library on Python.We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.

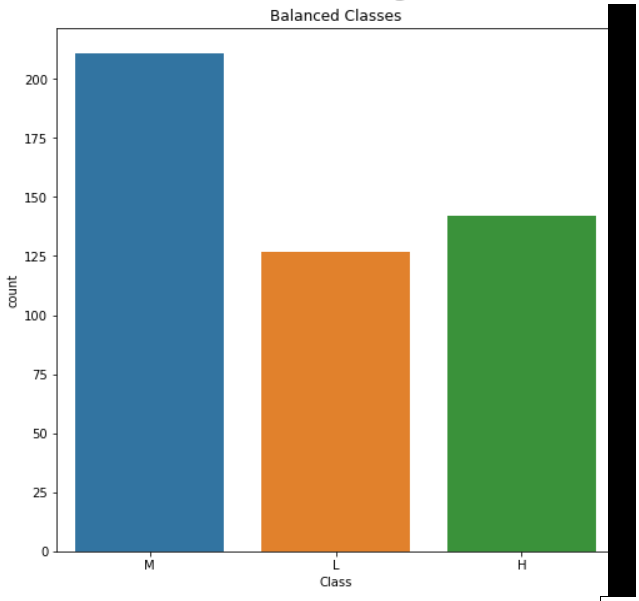
EVALUATION MODEL

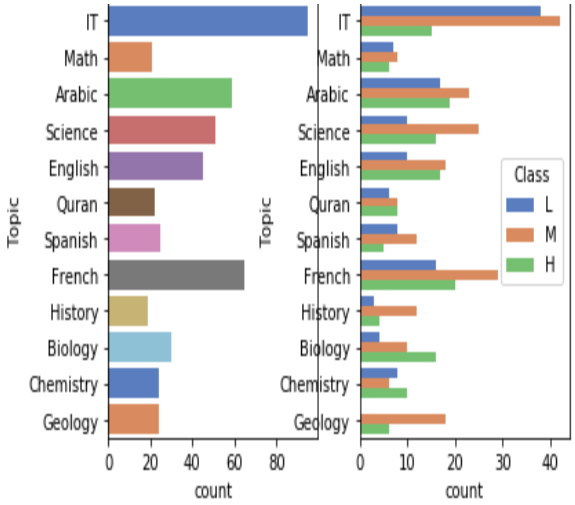
Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over fitted models. There are two methods of evaluating models in data science, Hold-Out and Cross-Validation to avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance. Performance of each classification model is estimated base on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs. Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

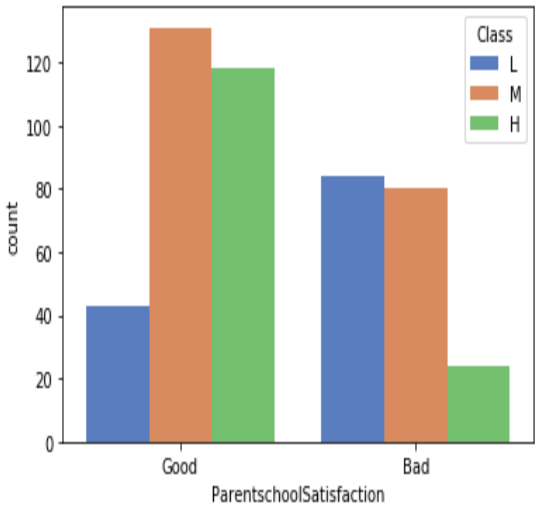
Screen Shots:

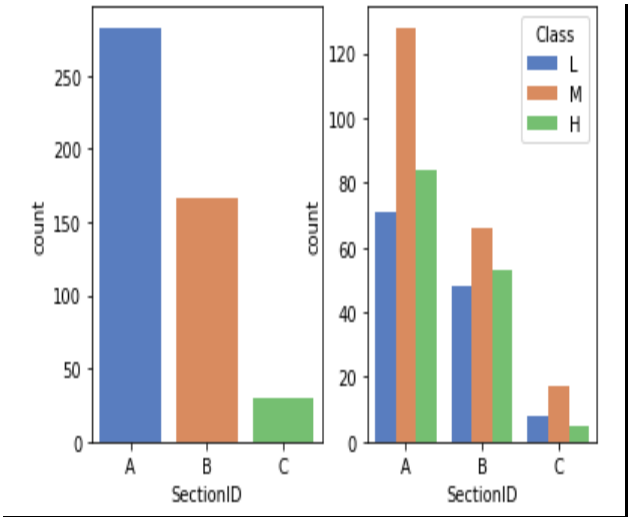

count
gender
F
M

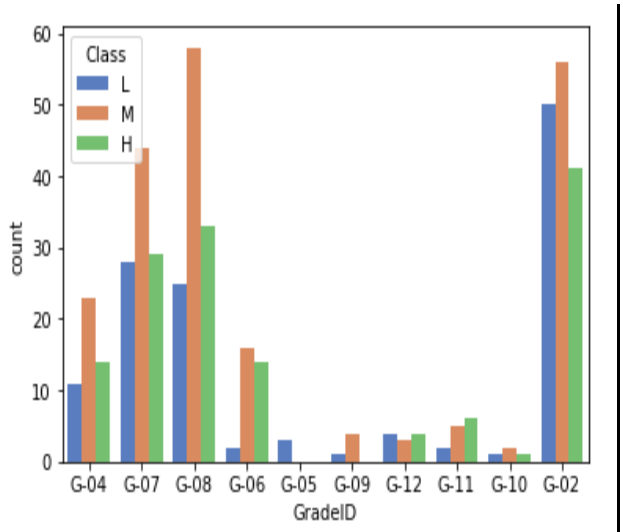

count
ParentschoolSatisfaction
Good
Bad

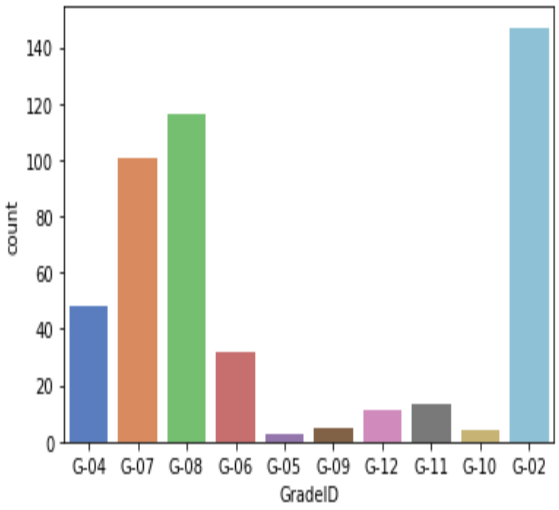

Balanced Classes
count
Class
M
L
H

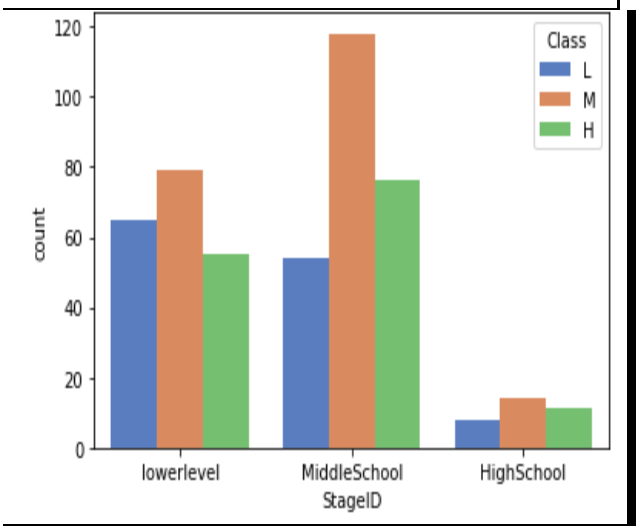

Topic
count
Class
L
M
H
IT
Math
Arabic
Science
English
Quran
Spanish
French
History
Biology
Chemistry
Geology


count
ParentschoolSatisfaction
Class
L
M
H
Good
Bad


count
SectionID
Class
L
M
H
A
B
C







Conclusion

Finally, performance analysis for students are a major problem. It is important that they are countered. The work reported in this thesis indicates the machine learning techniques with supervised learning algorithms to understand the performance of algorithm with respect to student records where we analyses the performance of student and categorized it into three classes as high , average, low with the accuracy of 79% .

Future Work

In the future we provide some technical solution by improve the efficiency of student performance .The user interaction model could be derived for giving the record of student dynamically and it could give staff an alert message about those students who are having low performance . We could build the prediction using Neural Network and can expect improvised results. We can add non- academic attributes along with academics attributes.

References :

[1] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From data mining to knowledge discovery databases.,” AI Magazine, pp. 37-53, 1996.

[2] “ÍndiceBrasscom de Convergência Digital,” 2015. [Online]. Available: www.brasscom.org.br. [Accessed: 14-May-2016].

[3] H. Witten, E. Frank and M. A. Hall, Data mining: Practical Machine Learning Tools and Techniques. Burlington: Morgan Kaufmann, 2011.

[4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," in ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, Jun. 2009.

[5] B. K. Baradwaj and S. Pal, "Mining Educational Data to Analyze Students' Performance," in International Journal of Advanced Computer Science and Applications – IJACSA, vol. 2, no. 6, pp. 63-69, 2011.

[6] J. R. Quinlan, C4.5, San Francisco: Morgan Kaufmann Publishers Inc., 1993. [7] E. L. Ferreira, H. Rausch, S. Campos, A. Faria-Campos, E. Pietra and L. da Silva Santos, "Medical data mining: A case study of a Paracoccidioidomycosis patient's database," e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on, Natal, 2014, pp. 275-280.

[8] F. H. D. Araujo, A. M. Santana and P. A. S. Neto, "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process," in IEEE Latin America Transactions, vol. 13, no. 1, pp. 299306, Jan. 2015.