International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 3, Issue 3, pp 6-10, March-2019

Result Prediction System using Data Science and Behaviour Analysis

[1]Meet Mehta, [2]Hiresh Joshi, [3]Anand Singh, [4]Shiwani Gupta

[1,2,3]Student, Department of Computer Engineering, Thakur College of Engineering & Technology, Maharashtra, India
[4]Assistant Professor & Deputy HOD, Department of Computer Engineering, Thakur College of Engineering & Technology, Maharashtra, India

***Abstract* - Students life today have become Monotonous. In the whilst, there's a need of an hour to develop a predictive model that helps student to know their pointers in advance. The main purpose is to implement a realistic model that helps students to predict their grades. The model predicts the forthcoming result of the respective student by a blend of both Past Data analysis and Present Student Behavior such as daily/weekly average time spent on PC, wake up time, time devoted on Social media, time devoted on studies, travel time. Although there are many predictive algorithms which helps to calculate the grades, but the main thing lies in efficiency. This system bucolically uses the concept of data mining, data modeling, linear regression. We have built a predictive model keeping into mind the behavioral aspects, by building some formulas on the basis of the weights assigned. The predictive based model does not completely depend on accuracy, but it can sanguinely predict the approximate range, although in many other systems, different approaches have been used, mentioned in the latter part.**

Keywords: Monotonous, bucolically, linear regression, predictive, Data Mining, Predictive.

I. INTRODUCTION

A SCRUM based Model developed for the Students, focuses on predicting the results in advance but taking into consideration many implicit as well as explicit factors. The Result Prediction system is a holistic approach towards determining the future result. To implement a realistic model that helps students to predict their grades, scores or which class they fall into prior to their result. The model predicts the forthcoming result of the respective student by a blend of both Past Data analysis and Present Student Behavior such as Daily/Weekly average spent on pc, wake up time and sleeping time, weekly average time spent on social media/watching to/series/movies. The application will be able to predict the result/performance of student using various classification algorithms and. The main purpose of the project is to make the result system automated without any manual calculations which reduces the chance of mistakes. The project also determines which class will the student fall into and also gives suggestion to the student to improve.

II. LITERATURE SURVEY

Predicting Student Performance using Advanced Learning Analytics: The research problem of students' performance prediction can be analyzed through diverse angles. In the current literature, a number of complimentary approaches provide a baseline for such an analysis. A prediction model (CHAID) is developed to predict the performance of higher secondary school students, which is critical before getting admission into universities. [1]. The gaps identified in the above survey are as follow The model only predicts whether the student will "pass" or "drop out". It makes use of student academic performance, family income & assets, family expenditure and student personal information. It doesn't include Behavior Analysis, midterm scores (Term Test scores) & Attendance. In this model the data set collected consists of "Scholarship Holding student's" from different universities. This model will be only beneficial for recruiting agencies. As hinted several studies have been done by researchers to predict the academic performance of students in various examinations.

Bhardwaj and Pal conducted a research on a group of 50 students enrolled in a specific course program across a period of 4 years (2007-2010), with multiple performance indicators, including "Previous Semester Marks", "Class Test Grades", "Seminar Performance", "Assignments", "General Proficiency", "Attendance", "Lab Work", and "End Semester Marks". They used ID3 decision tree algorithm to finally construct a decision tree, and if-then rules which will eventually help the instructors as well as the students to better understand and predict students' performance at the end of the semester.

Furthermore, they defined their objective of this study as: "This study will also work to identify those students which needed special attention to reduce fail ration and taking

© 2019 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 3, Issue 3, pp 6-10, March-2019

appropriate action for the next semester examination". Bhardwaj and Pal selected ID3 decision tree as their data mining technique to analyze the students' performance in the selected course program; because it is a "simple" decision tree learning algorithm. [2]. Early prediction of student results for identifying the "weak" students so that some form of remediation may be organized for them. In this paper a set of attributes are first defined for a group of students majoring in Computer Science in some undergraduate colleges in Kolkata. Early Prediction of Students Performance using Machine Learning Techniques - As hinted several studies have been done by researchers to predict the academic performance of students in various examinations.

These include higher secondary, graduation, post-graduation, engineering as well as medical courses. Predictions have been done for traditional university courses as well as distance learning courses. This literature is a similar example of our project predicting in which class the particular student will fall. [3] The gaps identified in the above survey are as follow the model only predicts which class the student falls ranging from "O – F". Well this model has some features for behavior analysis such as "No. of hours studied" but it lacks; it does not include "co-curricular" or "Extra-Curricular" Activities of the student. It doesn't consider students whose midterm scores are missing in the record & does not provide a report with suggestions to improve grades.

III. DESIGN & DEVELOPMENT USER - INTERFACE

a) Take input from students (Semester pointers)

Figure 1: Input to system (Pointer)

b) Attendance, Activities and Behavioral Inputs

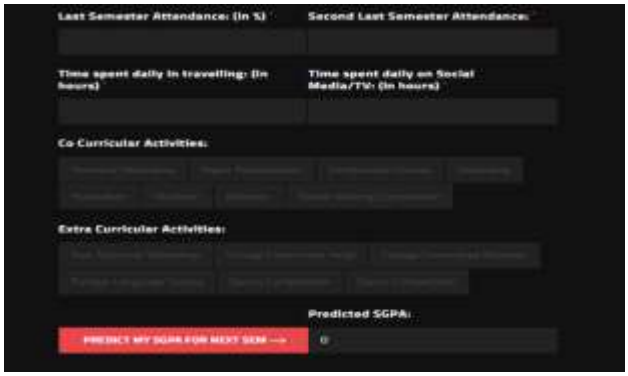

Figure 2: Inputs to system (Attendance and Behavioral)

IV. EXPERIMENTATION

The CHAID based predictive model showed accuracy in very limited number of classes. Although we are going to make an application based model, showing students their grades and their analysis report and also predict their forthcoming results. The proposed system is concerned with the students and their academic records. At initial phase the scope of the system will be at student academic level which will be helpful for the students and later it may be expanded for the faculties also for the analysis of teaching record. Technology used is already mentioned above. We'll be using MS Excel for Data Collection. Data Cleaning will be done using MS Excel & Python libraries such as Pandas and Numpy. Creating and Training Data sets and integrating them will be done on Google ML Cloud Server. The front end will be deployed directly over the web using PHP.

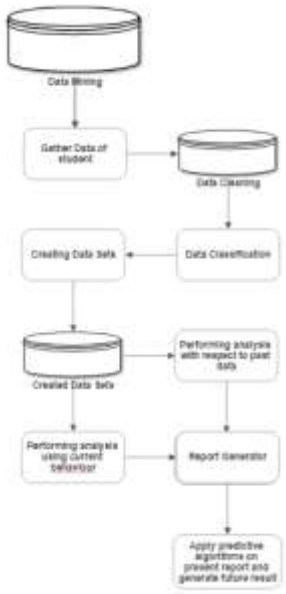

Figure 3: Block diagram of result prediction system

© 2019 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 3, Issue 3, pp 6-10, March-2019

a) Evaluation Measure

We focused on student graduation prediction. Since the true binary target label (1: graduate, 0: not graduate) is imbalanced (i.e., number of 0s outweighs number of 1s), accuracy is not an appropriate metric. Instead, we used the Receiver Operating Characteristic (ROC) for evaluating the quality of the predictions. An ROC curve was created by plotting the true positive rate (TPR) against the false positive rate (FPR). In this task, the TPR is the percentage of students who graduate, which the Grit Net labels positive, and the FPR is the percent of students who do not graduate, which the Grit Net incorrectly labels positive. The accuracy of each system's prediction was measured by the area under the ROC curve (AUC) which scores between 0 and 100% (the higher, the better) with random guess yielding 50% all the time. We used 5-fold student level cross validation.

Fundamental of Linear Regression

The backbone of this project us Linear regression algorithm. It is the most basic & common type of predictive analysis used. The general thought of regression is to look at two things: (1) does a lot of indicator factors work admirably in foreseeing a result (subordinate) variable? (2) Which factors specifically are huge indicators of the result variable, and how do they– demonstrated by the extent and indication of the beta estimates– sway the result variable? These regression estimates are used to find the relationship between one dependent variable and one or more independent variables. The most straightforward form regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

The general syntax of Linear Regression is given below:

```
from sklearn.linear_model import LinearRegression
from sklearn.datasets import make_regression
# generate regression dataset
X, y = make_regression(n_samples=100, n_features=2,
noise=0.1)
# fit final model
model = LinearRegression()
model.fit(X, y)
# define one new data instance
Xnew = [[-1.07296862, -0.52817175]]
# make a prediction
ynew = model.predict(Xnew)
# show the inputs and predicted outputs
print("X=%s, Predicted=%s" % (Xnew[0], ynew[0]))
```

V. RESULTS AND DISCUSSION

Initially, the first and foremost thing of any system is to calculate its feasibility. Following are some of the feasibility criteria as discussed:

Technical Feasibility

- Method of production: We will be using SSIS, Tableau, MySQL Server, VS 6.0, WEKA to develop the system
- We are using a framework of Agile Methodology (SCRUM model)
- Technology used for data storage: Excel , .csv file system
- Resources Required: Manpower, Programmers, data analyst , debuggers
- Software required: Tableau , SSIS , python interpreter
- Editor: Sublime Text

Economical Feasibility

The Cost Benefit Analysis summarizes the revenues and costs involved with the proposed project. As the proposed system will be used for the benefits of the students, no additional cost will be paid by them. No hardware system is included in our project, so the hardware cost gets minimized.

Modules

The system is homogeneous in nature as it performs only one function i.e. predicting the current SGPA of a student. Therefore, the system is broken into fewer modules:

a) Pointer based Analysis

The first and most important prerequisite of the model is availability of past students' data (pointer). In the data set we have collected the data of students showing their previous semester's marks and pointers. Trend analysis is used to plot the graphs between the pointers of two semesters and a gradient value is calculated after each analysis.

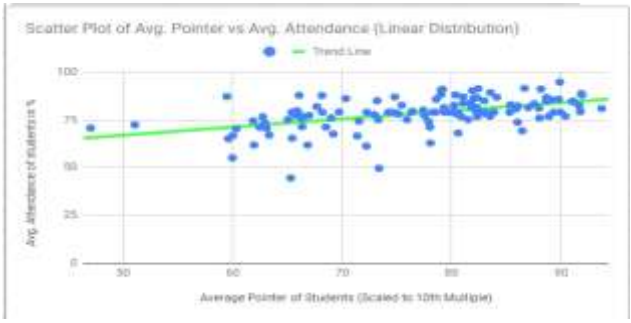

Figure 4: Average pointer of students (Scaled to10th Multiple)

© 2019 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 3, Issue 3, pp 6-10, March-2019

In this manner the gradient value of all six semesters is calculated and the graph is plotted. A Trend line shows the average of Pointers, if the predicted pointer is high then average then it is plotted above the trend line and if it is low then it is plotted below.

b) Attendance and Behavior based Analysis

Although attendance is mandatory for students in college but it does not particularly impact on the pointer because it is nonlinear. Therefore we have tried to make a little impact on the pointer by assigning a linear weight to the attendance. The attendances of previous two semesters are taking into consideration and then a mean of attendance matters. We have scaled the attendance on the range of 1-5 and the student having the highest range impacts a bit greater on pointer then the student who has the lowest range. Same is applied for Behavior, i.e. Time spent on Travelling, time spent on social media etc.

c) Graphical analysis of the pointer

Previously we have seen that trend analysis helps to predict the current semester's pointer. Therefore, a student must know how its progress graph looks like and what the chance of improvement is. The student can see his/her progress graph and then act accordingly for the next semester. Also tricks and tips links are provided at the end for motivation.

d) Progress Graph of the student (with tricks and tips for improvement)

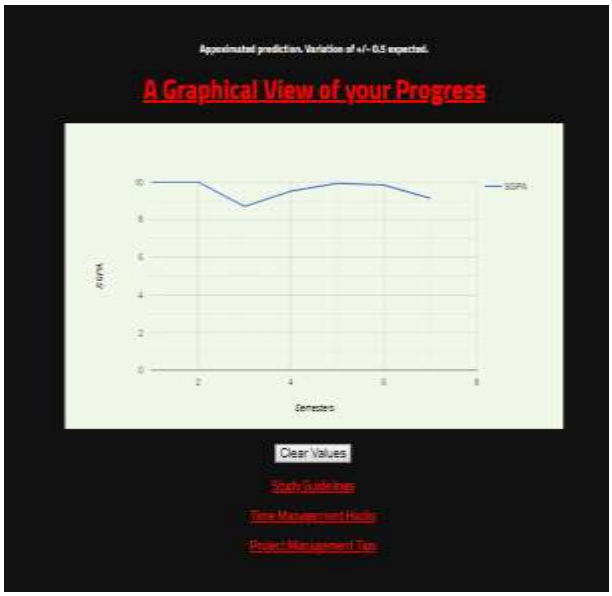

Figure 5: Progress Graph

VI. CONCLUSION

In the beginning while collecting the raw data, we were pretty sure about the conceptual view of the system in our mind. The first and foremost thing was collection of data, cleaning it and transforming it. The data was not magnified in number, earlier we had just 70-80 tuples of data, and then after transformation the tuples size increased to 120. But this number was not sufficient to train for a particular SVM or Decision tree model. Data Science is an ephemeral field in which the main heart is the data. Without sufficient data, the outcome cannot be generated. Therefore, we had to shift our model to linear regression. Also the number of tuples to be trained was limited because the data was not peculiar. The use case of the system also varied after the development of the system. Training the data till 6th semester was mandatory because if number of data tuples were less than it would result in vague result. However, the system correctly predicts the result approximate to range of +, - 0.5 pointer.

Predicting students' performance is mostly useful to help the educators and learners improving their learning and teaching process. This project has reviewed previous studies on predicting students' performance with various analytical methods. Most of the researchers have used cumulative grade point average (CGPA) 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] Ali Daud, Naif Radi Aljouhani, Farhaat Abaas and Rabeeh Ayaz Abbasi', "Predicting Student Performance using Advanced Learning Analytics" *WWW'17 Companion ACM Digital Library* ISBN: 978-1-4503-4914-7.

[2] Brijesh Kumar Bhardwaj, Saurabh Pal', "Data Mining: A prediction for performance improvement using classification" (*IJCSIS) International Journal of Computer Science and Information Security*, Vol. 9, No. 4, April 2011.

[3] Anal Acharya and Devdatta Sinha', "Early Prediction of Students Performance using Machine Learning Techniques" *International Journal of Computer Applications (0975-8887)* Volume 107- No.1.

© 2019 IRJIET All Rights Reserved

ISSN (online): 2581-3048

Volume 3, Issue 3, pp 6-10, March-2019

[4] M Ramaswami and R Bhaskaran', A CHAID Based Performance Prediction Model in Educational Data Mining, *IJCSI International Journal of Computer Science Issues*, Vol. 7, Issue 1, No. 1, January.

[5] https://luis-goncalves.com/what-is-scrum-methodology/

[6] https://medium.com/deep-math-machine-learningai/chapter-4-decision-trees-algorithms-b93975f7a1f1(14/10/18)

[7] http://www.statisticssolutions.com/non-parametricanalysis-chaid/ (12/10/18)

Citation of this article:

Meet Mehta, Hiresh Joshi, Anand Singh, Shiwani Gupta, "Result Prediction System using Data Science and Behaviour Analysis", Published in *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, Volume 3, Issue 3, pp 6-10, March 2019.

****

© 2019 IRJIET All Rights Reserved