

Student Performance Analysis using Machine Learning



P.Kalpana, E.Arunmaran, S.Hanif, T.Deebak

Abstract: Predicting student data to improve instructor and learner more efficiently in teaching and learning. It also strengthens contact between administrators, teachers and learners and helps monitor the behavior of students at multiple levels, such as class assignments, workshops, internal assessments and final exams. This program was built across three fields. We are Learning, Psychology and Computer Science. Educational institutes are increasingly using educational systems in recent years to assess their performance in order to construct plans for further growth and future actions. Such activities concentrate on discovering and improving approaches that can improve student academic performance, indirectly helping institutes attract more new students and maintain older students, the algorithms used in these systems are known as Educational Data Mining or EDMs. The prediction of student performance is an important aspect of EDM, which is the main area of this research work. Predicting student performance is a process that focuses on inferring information to learn from the student performance data. It can provide accurate collection of data on learning activities, such as time-on-task and evaluation scores, allowing for useful progress estimates for both the student and the trainer. In order to improve their performance, early warning of "at risk students" can be obtained which can help trainers to increase their focus on them. This provides a better way of predicting student performance to improve their skills at an earlier stage. Thus, student performance prediction helps to easily adapt, personalize and interfere.

Keywords : Student data, Performance , Prediction, future actions.

I. INTRODUCTION

Under studies get dropped out of school because of a few reasons one of them is the understudies ' horrible showing in scholastics that outcomes in terrible evaluations. One approach to forestall such a circumstance is by assessing the understudies ' last grades from past outcomes right off the bat in their examinations. On the off chance that that should be possible, the instructors just as the understudies would think that its supportive to help the scholarly exhibition.

Revised Manuscript Received on April 30, 2020.

* Correspondence Author

Ms P.Kalpana*, CSE, Sri Krishna College Of Technology, Coimbatore, India., Email:p.kalpana@skct.edu.in

E.Arunmaran, CSE, Sri Krishna College Of Technology, Coimbatore, India.Email: arunmaran.e@gmail.com

S.Hanif, CSE, Sri Krishna College of Technology, Coimbatore, India. Email: 16tucs041@skct.edu.in

T.Deebak, CSE, Sri Krishna College of Technology, Coimbatore, India. Email: deebakjay1999@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license ([http://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/))

Through assessing the estimates educators will get a general perspective on a particular understudy's prosperity and in what zones the individual needs to work more to accomplish passing marks and understudies will get a general thought of their investigations.

The point of the SA[System Analysis] is to produce the concise assignment of examination and furthermore to give total data about the definition, activities and different imperatives, for example, execution appraisal and framework improvement. SA's point[10] is to accurately decide the specialized subtleties for the key thought. Here is a need to assess the nature of the devices used to test structure and threatening issues. Labor, time and gadget details are the apparatuses to be broke down. Three-part groups take an interest in the whole life pattern of the undertaking aside from the testing stage. The exploration is completed by the certified analyzers before the genius is actualized.

II. FEASIBILITY STUDY

Possibility Study is the most elevated variant of the whole framework to help answer a few inquiries, for example, What's your concern? Indeed, even that is an issue worth understanding? The examination is utilized to decide the achievability of the proposed framework, contemplating the specialized, authoritative, and monetary variables. The administration will get an away from of the proposed framework by attempted a practicality study. For the undertaking, the accompanying sorts are considered to guarantee the plan is correct and don't represent any huge deterrents.

Feasibility study[13] encompasses the following things:

- Technical Feasibility
- Economic Feasibility
- Operational Feasibility

authors of the accepted manuscripts will be given a copyright form and the form should accompany your final submission. It is noted that:

- Each author profile along with photo (min 100 word) has been included in the final paper.
- Final paper is prepared as per journal the template.
- Contents of the paper are fine and satisfactory. Author (s) can make rectification in the final paper but after the final submission to the journal, rectification is not possible.

III. RELATED WORKS

The understudy execution expectation depends on different elements of understudy factors,

for example, a gathering, individual, ecological, and mental. Loads of research has been done in the course of the most recent couple of years to foresee the achievement of the understudies. Here a portion of the examination papers are considered and assessed for factors that influence the forecast of understudy scholastics in various understudies. Around 35-40 research papers, contemplates, book parts are considered for review.[1] Academic has demonstrated that institutional outer information sources (EDS) and interior information sources (IDS) have created the best result than any model dependent on existing institutional understudy databases.

This exploration utilized the understudy stamps in HSD[4] and mid-term, research center grade, workshop results, task, participation, schoolwork, understudy cooperation to predict[3] in another example, Internal Assessment Test grade for task accommodation and rating, right reaction, fearlessness, enthusiasm for the specific course and degree aspiration for anticipating understudy scholarly performance[2].

They likewise utilized the different kinds of prescient fields, for example, Sex, Family Information, Distance, High School Percentage, GPA, Entrance test marks, Scholarships, Time, Resources, the Internet, Degree Importance, Earnings[6] in their examination they utilized a few scholastic groups of CS&IT office at NED University, Pakistan. For their study[7] they utilized HSC marks, MPC marks, Math stamps in HSC, even checks in different subjects learned in the normal course of a programming language, CSA, Logic plan, OOP, DBMS, ALP, FAM, SAD, Data structure and so on.

In any case, in the wake of preprocessing the information, they found that characteristics, for example, Sex, Specialty, City, HSC Marks, type of SSC school, College CGPA are most significant[8] utilized for breaking down Bachelor of Computer Science first-semester understudy records. They utilized a few characteristics like Sex, Race, Location, University Entry Mode, Family Income for information collection[9] gathered Nigerian College of Education information from the understudies. In their exploration, they saw software engineering course Data Structure as one of the most significant themes and accordingly gather information for this subject individually.

Understudy traits, for example, parent training, parent calling, classification, SSC board, sort of confirmation, SSC medium, SSC class, first semester result, second semester, third semester, fourth semester, fifth semester and 6th semester result as most significant qualities accumulated information of around 350, BE (CSE) from KLN College of Information Technology understudies for expository purposes. They at first picked 33 qualities for survey however in the long run characteristics with the higher positioning are considered with the end goal of order. The property picked is CGPA, unfulfilled obligations, participation, SSC scores, cut-off building, vehicle of-training and the kind of information utilized by the board.

They gave information identified with understudy capabilities and qualities just as their segment data, utilizing information from chose understudies from Babcock University, Nigeria. In light of the data gathered, they considered a portion of the components, for example, age,

sexual orientation, conjugal status of guardians, parent capability, occupations of guardians, SSC score, HSC score, CGPA first year they attempted to make sense of the best approach to anticipate the last grade of designing understudies at the University of Computer Science, Greece.

IV. QUANTIZATION

During Quantization the key factors for student study can be identified. That includes factors such as absences, study time, travel time, family size, health etc. The derived characteristics are used to find accurate data predict performance of students. In this method, the Linear Regression Algorithm is used to find exact results. These features include progress user interface, performance estimate for students, visualization, and report development. A good user interface provides an interface that is easy to use, and not frustrating. In the meantime the students ' success forecast is included in the system to ensure the targets.

Existing systems have some features that are used during the new system's design and implementation process. A good user interface provides an easy-to-use interface which is not complicated. Meanwhile, the student performance forecast is integrated into the system to ensure the goals are met. Furthermore, report generation in Portable Document Format (PDF)[11] and display of diagrams such as PDF charts promote student study. From these features all requirements given by the user are met.

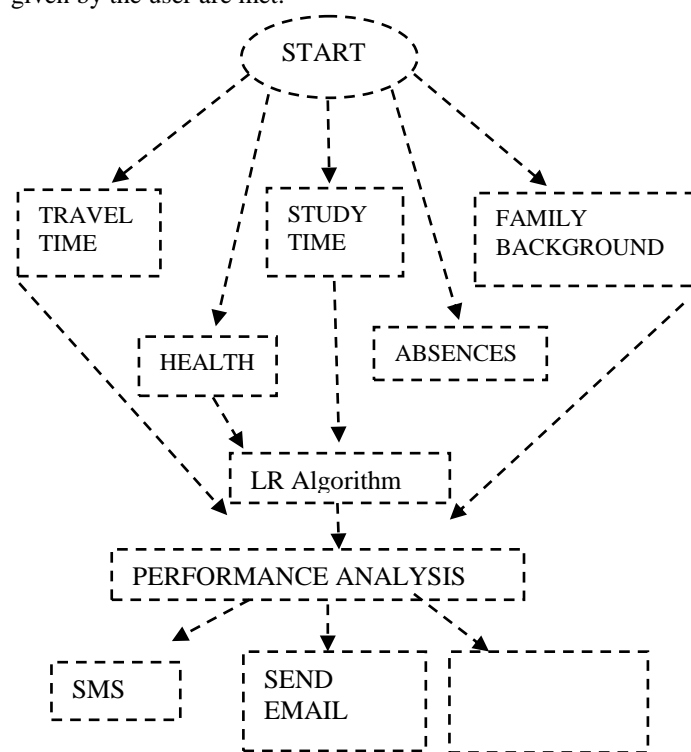


Fig. 1. Flow diagram of Student Performance Analysis

ALGORITHM & IMPLEMENTATION

Before you begin to format your paper, first write and save the content as a separate text file.

Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. Random Forest:

Irregular timberlands or arbitrary woods of choice is an intelligent technique for figuring out how to order different undertakings, which works by developing a choice tree at the hour of preparing and giving the class which is class mode (arrangement) or mean expectation (relapse) [12]of the trees.

B. Decision Tree:

Decision trees are graphic structures in which each future decision produces a new node resulting in a tree-like graph. Figure 2 demonstrates the definition of a decision-tree. This tree is used to assess if a student passed the test by looking at the values of GPA and age. In the margins, "Y" and "N" state whether the "GPA > 3" and "Age > 15" conditions are met. For machine learning, Decision trees partition the data set into appropriate values until a tree structure has formed. The process is called recursive-effect partitioning. If a completely homogeneous part can not be produced, more common value is chosen. Build a student data showing decision tree.

Determining the algorithm for the decision tree regularly does so. Classify whether or not an email is a spam, classify whether a tumor is malignant, classify whether or not a website is fake, etc. These are typical examples of how machine learning algorithms can make our lives a lot easier. The logistic regression algorithm is a very simple classification algorithm which is rudimentary and useful. Let's look at logistic regression more closely now.

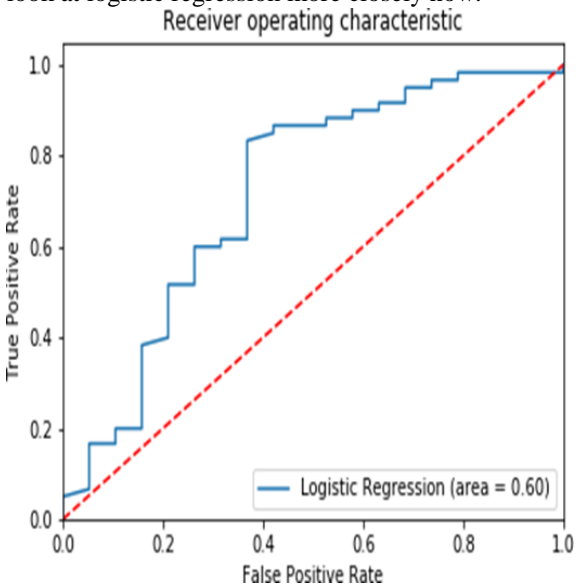


Fig. 2. Logistic Regression

V. RESULT

In Logistic regression algorithm single predictors are used to predict the output. The theoretically estimated value can be

from infinity everywhere. The result must be a gender vector, i.e. 0-"N," "1-"Y. The output of the equation should be between 0 and 1. In order to squash the predicted value between 0 and 1, the sigmoid function which takes the linear equation output(z) and gives the g(x) function which returns a squashed value is within the range of 0 to 1. To understand how sigmoid functions squash values within the set, let's see the plotting graph.

Logistic regression uses a representation equation which is similar to linear regression. Input values (x) are combined linearly using values to estimate a value (y) for output. The only distinction from logistic regression is that the values of binary variables are written.

Logistic regression ex:

$$y = \frac{e^{(a_0 + a_1 * x)}}{1 + e^{(a_0 + a_1 * x)}}$$

A0 is the intercept term where the source is y, and a1 is the single input value coefficient(x). The equation's data value must be gathered from the dataset.

The most simple form of data analysis using a single set of values is the representation of the model that would be stored in memory as the coef (coefficient) in the equation of Univariate analytics. Results are limited to one variable. This is not about causes or relationships and its main aim is to summarize the data.

Multivariate analysis uses different methods to evaluate data sets containing more than one variable and the techniques are of particular value when dealing with correlated variables.

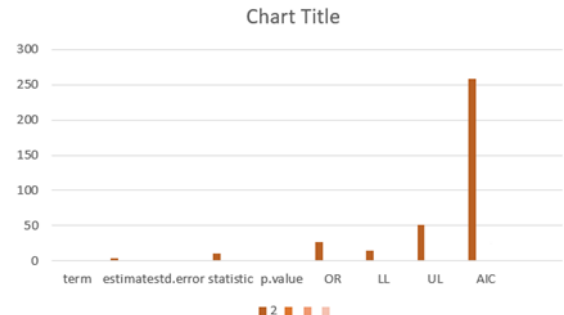


Fig. 3. G1 Score

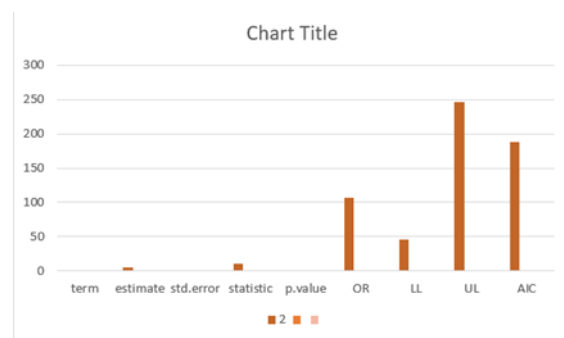


Fig. 4. G2 Score



Fig. 5. Multivariate Analysis

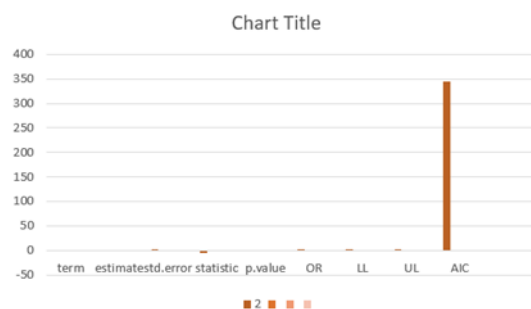


Fig. 6. Fail Odds Ratio

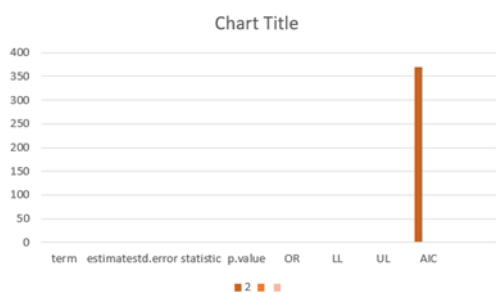


Fig. 7. Female Education Ratio

VI. CONCLUSION

Training is the fundamental pillar of our culture. Business Intelligence (BI)/Data Mining (DM) techniques enabling the extraction of high-level knowledge from data which offers opportunities for the educational domain. Several studies have employed BI / DM approaches to strengthen and improve the management of school resources. Here the estimation of secondary-school grades of two core classes (Mathematics) was presented by using school-like data for comparison past school grades.. Distinct input options (e.g. with or without past grades) were also discussed. The results obtained indicate that, provided the first and/or s, a high predictive accuracy can be achieved. It suggests the student achievement assumption is heavily influenced by previous performances. An overview of information generated by the best predictive models also reveals that there are other important features in some situations, such as: number of absences, reason for selecting the particular school, extra school support, lifestyle for students, and social variables for parents.

FUTURE ENHANCEMENT

Digitalizing the data by providing management dashboard to the professors and students to track the data more efficiently in order to achieve more accuracy. This helps the students to know their level of performance easily. And it will be more comfortable for the professors to maintain the students data in a centralized storage.

REFERENCE

1. Farhana Sarker, Thanassis Tiropanis and Hugh C Davis, Students' Performance Prediction by Using Institutional data sources like internal and external.
2. D. M. D. Angelina, Association rule generation for student performance analysis with help of apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) 1 (1) (2013) p12 – 16.
3. Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby, Data Mining: A Classification Method for prediction of Student's Performance, World Journal of Computer Application and Technology 2(2): 43-47, 2014.
4. Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, The Students Academic Performance prediction Using Data Mining Techniques, Applied Mathematical Sciences, Vol. 9, 2015, no. 129, 6415 - 6426HIKARI Ltd.
5. Mashael A. Al-Barrak and Mona S. Al-Razgan, Analysis of students performance using classification: a case study, Theoretical journal Applied Information Technology 20th May 2015. Vol.75. No.2.
6. Ms. P. Kalpana, R. Kirubakaran,Dr.P.Tamije selvy, "Hybrid SVD Model for Document Representation", International Journal of Engineering and Advanced Technology., ISSN:2249 – 8958, Volume-8 Issue-2, sep 2019.
7. Siamala Devi S, Shanmugam A., "An Integrated Harmony Search Method for Text Clustering using a Constraint based Approach", Indian Journal of Science and Technology, Vol 8(29), 73986, 2015.
8. Ms. P. Kalpana, Dr.P.Tamije selvy "A Probe on Document clustering methodologies and it Performance Metrics" in International Journal of Recent Technology and engineering(IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019.
9. Siamala Devi S, Dr. A. Shanmugam, "Hybridization of K-Means and Harmony Search Based On Optimized Kernel Matrix and Unsupervised Constraints", Asian Journal of Information Technology, Vol.15, No.22, pp.4512-4521, 2016.
10. Selvy P.T., Palanisamy V., Purusothaman T.," Performance analysis of clustering algorithms in brain tumor detection of MR images", European Journal of Scientific Research,2011.
11. Sreeja N.K., Sankar A." Pattern matching based classification using Ant Colony Optimization based feature selection", Applied Soft Computing Journal,2015.
12. Bhuvaneswari K., Rauf H.A." Edgelet based human detection and tracking by combined segmentation and soft decision, 2009 International Conference on Control Automation, Communication and Energy Conservation, INCACEC 2009.
13. Sathya Bama S., Irfan Ahmed M.S., Saravanan A.,"A mathematical approach for improving the performance of the search engine through Web content mining", Journal of Theoretical and Applied Information Technology,2014.

AUTHORS PROFILE



Ms.P.Kalpana received her B.E(CSE)degree under Anna university, India in the year 2011 and M.E(CSE) degree from Sri Krishna College of Engineering and Technology, India in the year 2013.she is pursuing her Ph.D under Anna university, India she actively publishing lot of research papers in her thirist. Her areas of interests are, data mining,

machine learning & image processing.





E.Arunmaran pursuing BE CSE. at Sri Krishna College of Technology. His field of interests include Cloud Computing, Artificial Intelligence.



S.Hanif pursuing BE CSE. at Sri Krishna College of Technology. His field of interests include Cloud Computing, Artificial Intelligence, Networking.



T.Deebak pursuing BE CSE. at Sri Krishna College of Technology. His field of interests include Cloud Computing, Artificial Intelligence.