Students and Staffs Performance Prediction Using Machine Learning

Abishake J[1], Roshan Kumar J[2], Abishek Js[3], R. Shobana[4]
[1,2,3]*Department of computer science and engineering, Aarupadai veedu institute of technology, Chennai*
[4]*Guided, Assistant Professor, Dept of CSE, Aarupadai veedu institute of technology, Chennai*

Abstract **- In recent years, educational data mining (EDM) has become a new research field due to the development of various statistical methods to study data in the educational context. One such application is the early prediction of student performance. Weak disciples" in order to arrange some form of correction for them. Since the number of attributes is large enough, a feature selection algorithm is applied to the data set to reduce the number of features. Then five types of machine learning algorithms (MLA) were applied to the data set, and it was found that the decision tree algorithm class gave the best results. Students and staffs performance prediction is very important to know a student progress rate. during this Research, we are attempting to seek out out student' current status, previous status and predict his/her future results. once the outcome, lecturers can provide him/her correct recommendation to avoid the poor result and can also groom the student. a lecturer can't monitor each and every single student at once. If a system can facilitate a lecturer regarding the scholars like that student needs which sort of help. The aim helps the student to avoid his/her predicted poor result and to improve their extra info activities based on the student's performance level using Artificial Intelligence. This research would be useful for the students and staffs.**

Index Terms **- Student Performance, Algorithm, Machine Learning, Education, Prediction, Data Mining.**

1.INTRODUCTION

In any shape of higher training it is indispensable to predict a student‟s academic overall performance [1,2]. There are two reasons for this: it is integral to identify which set of college students would do well in semester give up examination so that they can be awarded scholarships and more importantly to identify the students who may fail in semester quit examinations so that some structure of remediation can also be provided to them. The academic overall performance of students relies upon on a number of factors; some of these are previous tutorial records, financial status, family background, overall performance in mid semester examinations etc.

Based on these factors, classification models using MLAs can be developed which can predict student results. Various classes of MLAs [3] have been used by the researchers for this purpose. These consist of Decision Trees (DT), Bayesian Networks (BN), Artificial Neural Networks (ANN), Support Vector Machines (SVM) to title a few. As hinted above various studies have been accomplished through researchers to predict the academic performance of college students in various examinations. These encompass higher secondary, graduation, publish graduation, engineering as well as clinical courses. Predictions have been achieved for ordinary university courses as nicely as distance getting to know courses. For Web based totally learning publications [5] this kind of study attains a greater diploma of magnitude due to the lack of face-to-face interplay between the student and the instructor. In such cases MLAs are applied on pupil facts generated from “system logs” for predicting scholar results. Varied levels of success have been mentioned by using the researches in terms of prediction accuracy in all these fields while establishing these models. Brief report on some of these works is next discussed. Thai Nghe et al[3] has used DT and BN lessons of MLAs for predicting the undergraduate and put up graduate results of two universities in Thailand. The whole wide variety of pupil data used for this prediction is 20492 and 932 respectively. Algorithms used for this prediction are C4.5, MSP and Naive Bayes. They concluded that for all lessons of predictions DTyields better effects than BN by means of 3 to 12%. Resampling was used to improve the prediction accuracy. Kotsiantis et al [4]

described a model to predict scholar effects for a distance gaining knowledge of direction in Hellenic Open University. Predictions had been performed on the basis of marks got in written assignments. The algorithms used for this prediction are C4.5, Naïve Bayesian Network (NBN), Back Propagation (BP), 3-Nearest Neighborhood (3-NN) and Sequential Minimal Optimization (SMO). A set of 510 students of the university used to be chosen for experimental purpose. It was once discovered that the NBN algorithm generates the excellent effects (accuracy72.48%). Ramaswami et al [5] developed a predictive fact mining model for student overall performance to end up privy to the factors causing negative overall performance in more secondary exam in India. A records set for 772 college students collected from ordinary college students and university workplaces have been used for this prediction. This tree become used to generate a hard and fast of choice policies used for predicting student grades. The preferred prediction accuracy finished become 44.69%.Menaei-Bidgoli et al[6] carried out records mining algorithms on"logged records" in a instructional net primarily based totally definitely learning gadget. The gadget become examined with a records set of 227 college students enrolled in a physics route in Michigan State University. Classification become as soon as at the start completed using Quadratic BN, 1-NN, Prazen Window, Multilayer Perceptron (MLP) and C5.zero DT. It become as soon as taken into consideration that combining those classifiers will increase prediction accuracy. Genetic Algorithms (GA) have been in addition used to enhance prediction accuracy with the aid of using 10%.Kovacic [7] explores the "socio-demographic" and "study environment" elements that results in student dropout in a polytechnic college in New Zealand. He makes use of pupil enrollment records like age, gender, ethnicity for this motive. The whole variety of student documents used for motive become 450. Algorithms used for this prediction are CHAID and Classification and Regression Trees (CART). It was once determined that CART were given a better degree of accuracy (60.5%). Based at the results of Confusion Matrix and ROC curve he concluded that choice timber primarily based totally on enrollment records on my own aren't sufficient to categorise college students accurately. However, making use of those algorithms on pupil records set isn't green in maximum conditions because it contains alarge amount of factors and records records [10]. In this context the intention of this paintings is as follows:

1. To find out the attributes and exam pattern of a hard and fast of college students majoring in Computer Science in a few undergraduate faculties in India.
2. To carry out elements decision in this records set to decide the last variety of features; those factors could then be used for classification.
3. To have a look at MLAs in this decreased feature set to understand the scholars who can also additionally moreover feature exceptionally well and greater importantly folks who also can carry out poorly.
4. To feature a parametric assessment with the modern-day literature noted above. In slight of this the corporation of the paper is as follows. The subsequent vicinity gives a short evaluation of the MLAs used on this paper. The following component describes pupil characteristic and records set used on this paintings at the side of exam pattern. The outcomes and the inferences are subsequent discussed.

Finally acomparative discover approximately of this paintings is completed with similar works in applicable literature.

2. MACHINE LEARNING MACHINE

Learning is a sub-vicinity of synthetic intelligence, wherein the time period refers back to the capacity of IT structures to independently discover answers to troubles with the aid of using spotting styles in databases. In different words: Machine Learning permits IT structures to apprehend styles on the premise of present algorithms and records units and to broaden good enough answer concepts. Therefore, in Machine Learning, synthetic understanding is generated on the premise of experience.

3. TEXT MINING ALGORITHM

Text mining (additionally called textual content analytics) is an synthetic intelligence (AI) era that makes use of herbal language processing (NLP) to convert the free (unstructured) textual content in files and databases into normalized, established records appropriate for evaluation or to power gadget learning

(ML) algorithms. Text evaluation offers qualitative outcomes and textual content analytics offers quantitative outcomes. If a gadget plays textual content evaluation, it identifies vital records in the textual content itself, however if it plays textual content analytics it well-known shows styles throughout heaps of texts, ensuing in graphs, reports, tables etc.

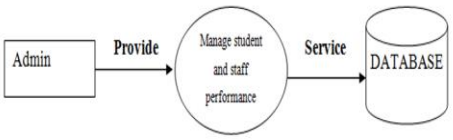
Attributes and Domain

Attribute Number	Attribute Name	Description	Domain
1	Gender	Student's sex	m-male; f-female
2	Caste	Student's caste	1-UR; 2-SC; 3-ST; 4-OBC
3	Religion	Student's religion	1-Hindu; 2-Muslim; 3-Christian; 4-Others
4	Fsize	Student's family size	Numeric eg.(2,4,6)
5	Board	Student's board at higher secondary level	1-State boards; 2-CBSE; 3-ISC;4-other boards
6	Sorigin	Student's state of origin	1-TN;0-Other States
7	Income	Student's family income	1- <=11K;2- >11K but <=21K;3- >21K but <=31K;4- >31K but <=41K;5- >51K
8	Boardmarks	Student's aggregate % of marks at higher secondary level	1- >=89%; 2-89% to 79%;3-79% to 70%;4-69% to 60%;5-59% to 50%; 6-<50%
9	Hday	Number of hours studied by the student per day	1-9 hours
10	Atten	Students % attendance in class for the semester	1- >=90%; 2-89% to 80%;3-79% to 70%;4-69% to 60%;5-59% to 50%; 6-<50%
11	Medium	Medium of study at higher secondary level	1-English; 2-Tamil; 3-Hindi;4-other
12	Grade (Dependent Variable)	Grade secured by the student in final semester	O-90% to 100%;E-80% to 90%;A-70%-80%;B60% to 70%;C-50% to 60%;D-40%- 50%;F<=40%

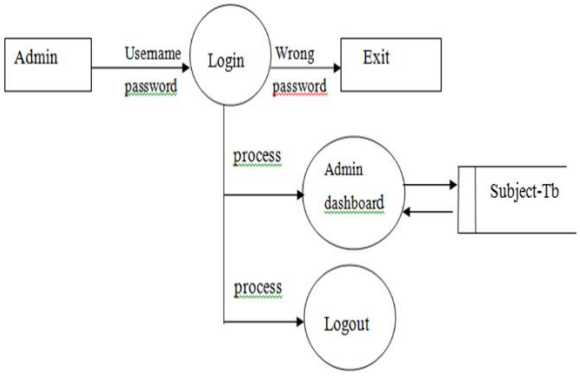
Needless to say, students have to pay a lot of money to the tutor. It has been observed that students who do well in exams tend to do well in final exams. Another indicator in this sense is the student's test scores in the middle of the semester. The number of hours of study also determines the quality of his academic performance, there are These attributes are used to predict student performance as a seventh-grade question for the final exam: Specifically, O means that the student is very good and can receive a scholarship, and F means that the student is very poor. Therefore, corrective care may be required. Other qualifications are fully defined. The attributes, predictors, and corresponding domains of the students are listed in Table 1.The 2006-10 data set was initially collected and used for training purposes. The total number of students surveyed in this period is 309. After deleting the data records with missing values, the total number of complete data records for TH11, PR11, and PR12 files are 287, 292, and 297 respectively. The data records are as followsDS1. The data from 2011 to 2012 are used for testing.

4. DATA FLOW DIAGRAM

Level-0: The Figure shows the admin managing student and staff performance

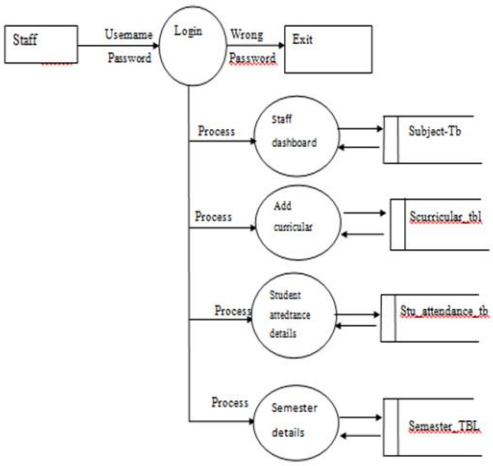


Level-1: The Figure shows admin login and dashboard.

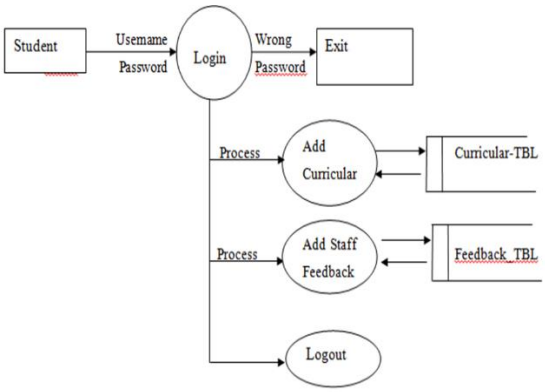


Level-2:The Figure shows Staff login and the process

of adding curricular and student's attendance details.



Level-3: The figure shows Student's login and adding feedback and curricular.



5. RESULTS AND DISCUSSION

The attribute set shown in Table 1 includes 14 independent attributes and one dependent attribute for prediction. Classification of data with such a large attribute set will lead to many shortcomings [10]: high computational complexity, poor interpretability of the model, and high degree of data overfitting, which reduces the possibility of generalization. Acharya and Sinha [18] conducted a detailed study on the influence of various filter and wrapper-based feature selection methods on DS1.Three filter-based methods are used: correlation-based feature selection (CBFS), chi-square feature evaluation (CBFS), and information gathering attribute evaluation (IGATE). In the bowl, they use learning algorithms to select attributes. Naive Bayes and 1 block. They found that the CBFS algorithm has 8 functionsBetter ranking results. To ensure consistency, the CBFS algorithm with the 8 most classified features was selected to apply the classifier in each of the three articles (TH11, PR11, and PR12).

Fig 5.1 STUDENT REGISTRATION

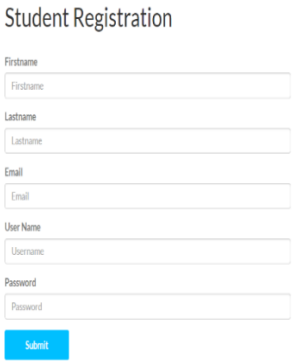


Fig 5.2 STUDENT LOGIN

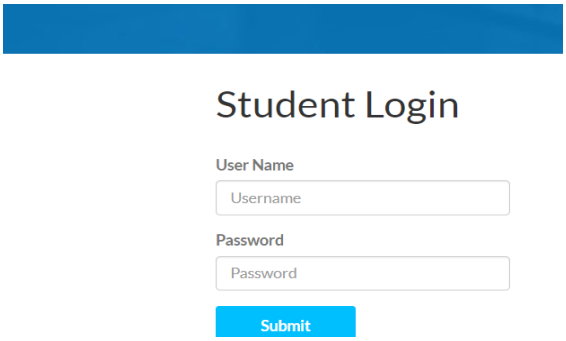


Fig 5.3 STAFF LOGIN

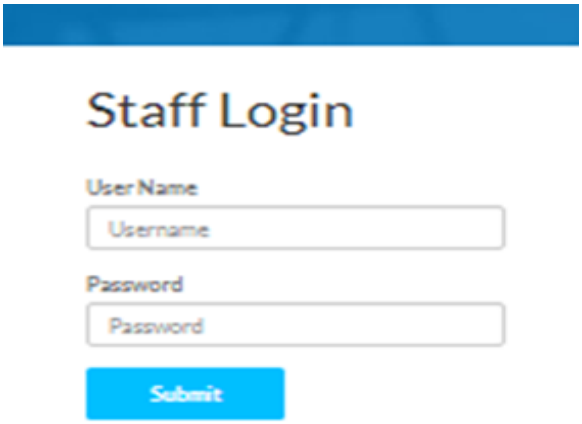


Fig 5.4 ADMIN LOGIN

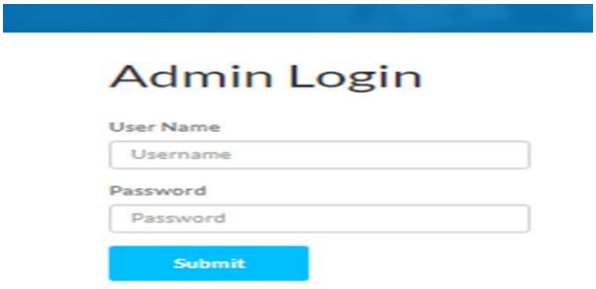


Fig 5.5 VIEW STUDENT PREDICTION

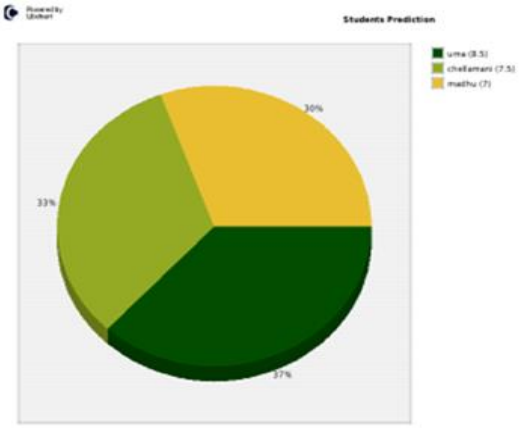


Fig 5.6 VIEW STAFF PREDICTION

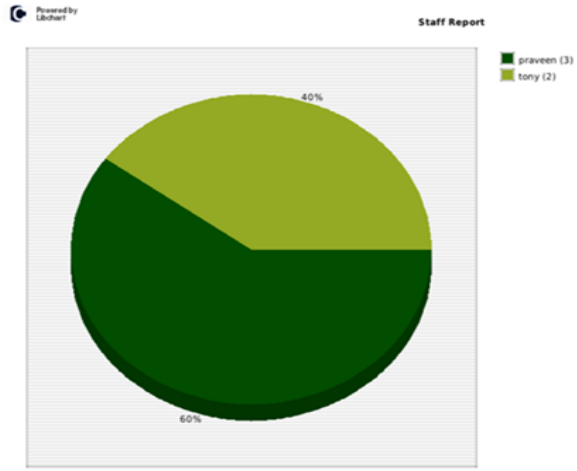


Fig 5.7 SEMESTER DETAILS

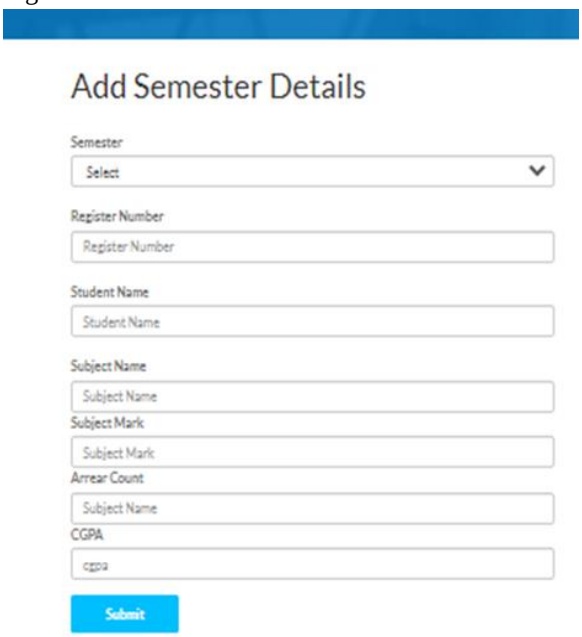
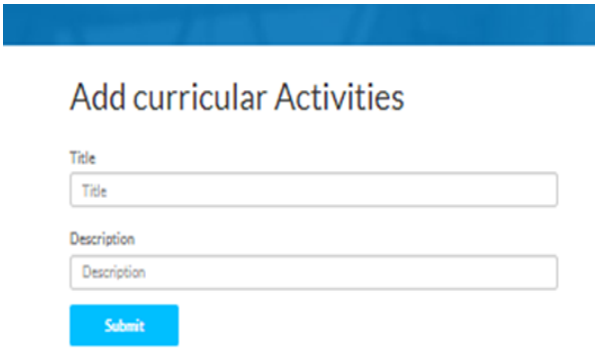


Fig 5.8 CURRICULUM DETAILS



6. COMPARATIVE STUDY WITH CONTEMPORARY LITTERATURE

As mentioned in Section 1, numerous researchers [3,4,5,6,7,8,9] have already used MLAs to are expecting scholar results. Some of those are said in phase 1. A first-rate distinction between the ones works and modern paintings is that right here characteristic choice is explored in detail [18] earlier than appearing classification. Various fashions evolved can be farther in comparison at the foundation of the subsequent parameters:

1. Machine Learning Algorithm used: Several training of MLAs had been utilized by special researchers like Decision Trees, Bayesian Networks, Neural Networks, Support Vector Machines etc.
2. Number of training predicted: This offers the variety of classes into which statistics is to be categorized. For example, 6-magnificence suggests that scholars are categorized into 6 classes.
3. Experiments: The accuracy of the version constructed may be predicted via way of means of making use of it on scholar dataset.
4. Accuracy of prediction: The usability of the version relies upon at the prediction accuracy.
5. Experiments: The accuracy of the version constructed may be predicted via way of means of making use of it on scholar dataset.
6. Accuracy of prediction: The usability of the version relies upon at the prediction accuracy.
7. Sample size: The better the variety of scholar records, the more is the prediction accuracy.
8. Simulating device used: This is the software program package deal used for appearing prediction.

7. CONCLUSION

The purpose of this article is to apply MLA to predict student performance. To this end, suitable representatives of the five MLA classes were trained and then evaluated. They have been found to give useful results.

Excellent learners, and more importantly, they may perform poorly in the last quarter. These attributes and data sets come from a survey of a group of computer science students from selected universities. It should be noted that the research method developed is universal in nature: it can be applied continuously or remotely.

Educational courses, including courses that use online learning. Second, a few students who are registering for courses are shown in the middle. Semester exams, but for some reason, they cannot take the exams in the last quarter.

These students are not included in the prediction because they contain missing attributes. Finally, the prediction efficiency can be improved by combining multiple classifiers (CMC). It is also suitable for this purpose.

REFERENCE

[1] 10th International conference on computing, communication and Networking Technologies (ICCCNT) DOI:10.1109/ICCCNT45670.2019. 895 5629 Publisher IEEE

[2] International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue IV, April 2018- Available at www.ijraset.com

[3] International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) ISSN:2277 128X Volume 7, Issue 2, February 2017 Available at www.ijarcsse.com

[4] Cortez,Paulo, Alice Maria and Gonclaves Silva,(2008),Using data mining to predict secondary school performance .