International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Predicting Student Performance using Machine Learning Approach

Manisha Bhimrao Mane and Dhanashri Kulkarni

Department of Computer Engineering. D.Y Patil College of Engineering, Ambi, Pune

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Predicting Students performance heretofore can be extremely helpful for educational institutions to improve their instructing quality. This paper proposes to predict students performance by thinking about their scholarly subtleties. Educational associations are extraordinary and assume utmost significant job for the improvement of any country.As Education changes the lives of people, families, networks, social orders, nations and at last the world! This is the reason we live agreeable lives today. Presently a day's training isn't restricted to just the homeroom instructing however it goes past that like Online Education System, Web-based Education System, Seminars, Workshops, MOOC course. turns into It's all the more testing to Predict understudy's performance in view of the colossal main part of information put away in the conditions of Educational databases, Learning Management databases. Students' performance can be assessed with the assistance of different accessible techniques.It is advancing territory of concentrate that accentuation on different strategies like characterization, prediction, include determination. It is utilized on learning records or information identified with training to predict the students' performance and learning conduct by extricating the hidden knowledge.

Keywords: *College Education, Machine Learning, Result Prediction,supervised learning .*

Introduction

Today every educational institution handles and deals with large amount of student data which can be beneficial for a number of reasons. One of the important application of such data is predicting student performance. Such a prediction can be useful not only for the students but also for teachers/mentors.

Mentors can provide special assistance to the students who are on the verge of failing. In order to determine which category a student lies, such data can be quite helpful. This application can be used by any prominent school or colleges. It can be used to predict the pointer ranges or percentage range for future semester exams. These ranges can be predicted using a number of data mining algorithms such as classification algorithms, rule-based algorithms, ensemble methods, and neural networks. The main aim of this project is the selection of features that show a strong relationship with a target attribute that is to be predicted from a high dimensional data-set.

We have evaluated and compared the number of algorithms such as decision tree, random forest, support vector machine, naive Bayes and neural networks by applying them on the data-set The rest of the paper provides an explanation on nature of neural networks along with the results of our evaluation. Machine learning is used for analyzing data based on past experience and predicting future performance .Reinforcement machine learning algorithms is a branch of artificial intelligence .It automatically determines the behaviour of environment and maximizes its performance .

A. Motivation

We will be focusing on the improvement of Prediction classification techniques which are used to analyze the skill expertise based on their academic performance by the scope of knowledge. Measuring student performance using classification technique such as decision tree. The task can be processed based on the several attributes to predict the performance of the student activity respectively.

B. Objectives

1. Our system can be used for various different waysin various institutions universities ,such as we can identify the bright students for scholarship and further can identify the weak students earlier who can be guided at right time for exact guidance.
2. To automate the process of student analysis and result making traditional ways.

Review of Literature

In the [1] work, author has used DT and BN classes of MLAs for predicting the undergraduate and post

graduate results of two universities in Thailand. The total number of student records 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 classes of predictions DT yields better results than BN by 3 to 12 percent. Re sampling was used to improve the prediction accuracy.
In the [2] work, Kotsiantis et al [4] described a model to predict student results for a distance learning course in Hellenic Open University. Predictions were done on the basis of marks obtained in written assignments.

The algorithms used for this prediction are C4.5, Na¨ıve Bayesian Network (NBN), Back Propagation (BP), 3-Nearest Neighborhood (3-NN) and Sequential Minimal Optimization (SMO). A set of 510 students of the university was chosen for experimental purpose. It was found that the NBN algorithm generates the best results .

In the [3] work, Rama-swami et al [5] developed a predictive data mining model for student performance to identify the factors causing poor performance in higher secondary examination in TamilNadu. A data set for 772 students collected from regular students and school offices were used for this prediction.Algorithm used for this prediction is Chi-Square Automatic Interaction Detection (CHAID) DT. This tree was used to generate a set of decision rules used for predicting student grades.

In the [4] work, Menaei-Bidgoli et al[6] applied data mining algorithms on "logged data" in a educational web based learning system. The system was tested with a data set of 227 students enrolled in a physics course in Michigan State University.Classification was initially performed using Quadratic BN, 1-NN, Prazen Window, Multi layer Perception (MLP) and C5.0 DT. It was seen that combining these classifiers increases prediction accuracy. Genetic Algorithms (GA) were further used to improve prediction accuracy by 10 percent

Kovacic [7] explores the "socio-demographic" and "study environment" factors that results in student dropout in a polytechnic college in New Zealand. He uses student enrollment data like age, gender, ethnicity for this purpose.The total number of student records used for purpose was 450.Algorithms used for this prediction are CHAID and Classification and Regression Trees (CART). It was found that CART obtained a higher degree of accuracy (60.5). Based on the results of Confusion Matrix and ROC curve he concluded that decision trees based on enrollment data alone are not sufficient to classify students accurately.

Karamohzis and Vrettos [8] have used ANN for predicting student graduation outcomes at Waubonsee College. The prediction model was constructed from a profile of 1407 students of which 1100 were used for training and 307 were used for testing purpose. The average predictive efficiency for training and test sets were 77 percent and 68 percent respectively.

Proposed Methodology

This approach is based on machine learning so we work on the parameters specified by teacher for individual students.

Admin will be the central authority for User activation purpose then the .

Once teacher logged in the he/she can update the each and every students records and can generate the results as per parameters.

Machine learning and Deep learning classifiers are given to the data set collected from educational environments. Data is preprocessed and check for missing values. Classifiers are applied on the data set to build the models. Models are tested with test data to predict the students' performance and the best models yielding high accuracy are considered.

1. Firstly, MLP-a Deep Neural network classifier andclassifiers of data mining namely Bayes Net, Classifiers. Training will be performed on Educational data-set and a Model was obtained.

2. The Obtained Model from the first phase is suppliedwith test data set and the results are obtained.cross validation is done here in the second phase which involves both training and testing.

3. The obtained results are assessed on parameterslike Age, M Education, F education , the best model yielding high results is selected in the third phase.

Advantages of Proposed System:

1. The manual prediction methods are overcome by ourproject.

2. Students will be able to focus on the topics and subjects.

3. This proposed system effectively able to record thescrolling percent of web page.

4. This proposed system accurately calculate the students performance based on the data.

5. The model proposed in this study of prediction problemis The students data set is loaded in weka and converted to arff file. The pre-processing step is required as it applies a filter to change the ordering of data elements in the data set. Randomize and Remove Percentage Filters are chosen to perform this task.

6. Classification is one of the most frequently studiedproblems by data mining and machine learning (ML) researchers. It consists of predicting the value of a (categorical) attribute (the class) based on the values of other attributes (the predicting attributes). There are different classification methods.Bayesian classification is an algorithm that is based on Bayes rule of conditional probability. Bayes rule is a technique to estimate the likelihood of a property given the set of data as evidence or input.

A. Architecture

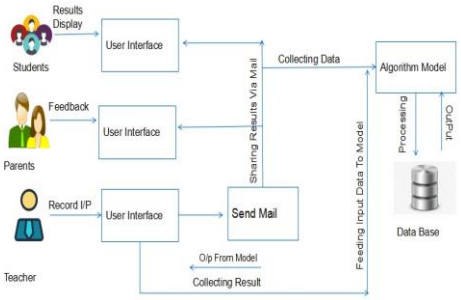

Fig. 1. Proposed System Architecture

Explanation:

Machine Learning
In this step we are going to use the machine learning methodology for predicting the students performance by using the external parameters that are going to affect the performance.
Project flow Steps:
1) The users will be registering first. 2) Then the admin will be activating all users like teacher ,student ,parents.

3) Teacher can log in and add students and can update individual students records.
4) The teacher can also generate the results and share the results with students and parents.

B. Algorithms
1. Naive Bayes Algorithm:
Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier. The Naive Bayes algorithm is called "naive" because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features. Here we classify the heart disease based on heart check up attributes. Naive Bayes or Bayes' Rule is the basis for many machine learning and data mining methods. The rule (algorithm) is used to create models with predictive capabilities. It provides new ways of exploring and understanding data.
Why to prefer naive Bayes implementation:

- When the data is high.
- When the attributes are independent of each other.
- When we expect more efficient output, as compared to other methods output.

Based on all these information and steps we classify to predict the heart disease depending on heart check up attributes.
Steps:
1. Given training dataset D which consists of documents belonging to different class say Class A and Class B 2. Calculate the prior probability of class A=number of objects of class A/total number of objects Calculate the prior probability of class B=number of objects of class B/total number of objects

3. Find NI, the total no of frequency of each class

Na=the total no of frequency of class A Nb=the total no of frequency of class B

4. Find conditional probability of keyword occurrence givena class:

P (value 1/Class A) =count/ni (A)
P (value 1/Class B) =count/ni (B)
P (value 2/Class A) =count/ni (A)
P (value 2/Class B) =count/ni (B)
.....
.....
.....
P (value n/Class B) =count/ni (B)

5. Avoid zero frequency problems by applying uniform

distribution

6. Classify Document C based on the probability p(C/W)

a. Find P (A/W) =P (A)*P (value 1/Class A)* P (value

2/Class A)..... P(value n /Class A)

b. Find P (B/W) =P (B)*P (value 1/Class B)* P (value 2/Class

B)..... P(value n /Class B)

7. Assign document to class that has higher probability.

C. Mathematical Model
1. Mathematical equation:
The algorithm implemented in this project is describe as: Algorithm
P(class/features) = P(class) * P(features/class)*P* _____(*features*)

- P(class/features) : Posterior Probability
- P(class) : Class Prior Probability
- P(features/class) : Likelihood
- P(features) : Predictor Prior Probability

A. Normal distribution
Normal distribution The probability density of the normal distribution is:

$$f(x|\mu,\ \sigma^2) = \frac{1}{\sqrt{2\,\pi\sigma^2}}\ ^{-\frac{(x-\mu)^2}{2\,\sigma^2}} \quad (1)$$

Where

- 'μ' is the mean or expectation of the distribution,
- 'σ' is the standard deviation, and
- 'σ^2' is the variance.

Result and Discussion

We will be recording the details of all external parameters specified by teacher to the particular students . On the basis of input given from teacher side the performance prediction of the student will be done and the generated output will be sent to parents and students for further purpose of operation.
It is experimented that MLP technique, Naive Bayes, performed well in predicting student's performance than existing systems SVM classifier. Techniques that gave optimal results are MLP, Naive Bayes with maximum accuracies of 90.00 Percent.
Based on the results, MLP technique is more efficient compared to other technique in prediction of students' performance. Rules can be mined and accuracy needs to be improved in SVM, K-NN as part of the future work.

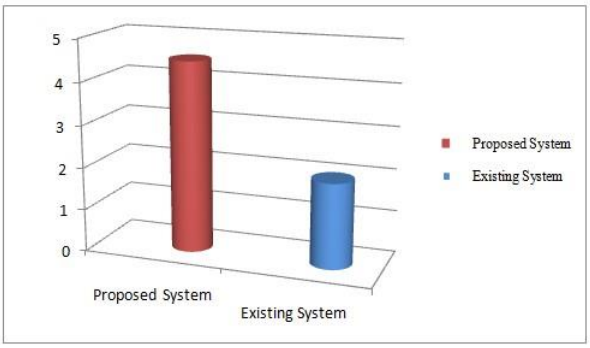

Fig. 2. Algorithms Comparison

Conclusion

Machine learning procedures can be valuable in the field of students performance prediction thinking about that they distinguishes from the starting point of academic year.

The point of this paper is to apply machine learning calculations for forecast of students performance. An early examination of students having horrible showing enables the administration to make auspicious move to improve their performance through foreseeing their academic subtleties. Precisely predicting students performance dependent on their continuous academic records is anticipated. Additionally we conclude that proposed framework is helping us to improve the candidates performance. In this paper machine learning can end up being amazing asset and all algorithms we utilized increases with increase in data-set size size.

Acknowledgment

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the authorities of University of Pune and concern members of cPGCON 2019 conference, organized by, for their constant guidelines and support. We are also thankful to the reviewer for their valuable suggestions. We also thank the college authorities for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

References

[1] Romero, C. and Ventura, S.2007, Educational data mining: A survey from 1995 to 2005, Expert Systems with Applications, 135–146.

[2] Castro, F., Vellido, T., Angela Nebot, and Mugica F.,Applying Data` Mining Techniques to e-Learning Problems.

[3] Nguyen, N. N., Janeck, P, Haddawy, P., 2007, A Comparative Analysis of Techniques for Predicting Academic Performance, 37th ASEE/IEEE Frontiers in Education Conference.

[4] Kotsiantis S. Piarrekeas C. Pintelas, P.,2007.Predicting Students performance in Distance Learning using Machine Learning Techniques, Applied ArtificialIntelligence, 18:411-426.

[5] Ramaswami, M., Bhaskaran, R.,2010, A CHAID Based Performance Prediction Model in Educational Data Mining, International Journal of Computer Science.

[6] Minei-Bidgoli, B., Kashy, D., Kortemeyer G., Punch W F, 2003. Predicting Student Performance: An Application of Data Mining Methods with the Educational Web-Based System LON-CAPA, 33rd ASEE/IEEE Frontiers in Education Conference.

[7] Kovacic, Z. J.,2007, Early Prediction of Student Success: Mining Students Enrolment Data, Informing Science IT Education Conference .

[8] Livieris, E., Drakopoulou, E., Pintelas, P., Predicting students' performance using artificial neural networks.

[9] Stefanowski, J., An Experimental Study of Methods Combining Multiple Classifiers - Diversified both by Feature Selection and Bootstrap Sampling.

[10] Zhao, Y., and Zhang, Y.,Comparison of decision tree methods for finding active objects.