International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952106

A Machine Learning Approach for Predicting Student Performance

C. Selvi*[1], R. Shalini[2], V. Navaneethan[3], L. Santhiya[4]

[1]Associate professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Anna University, Tamilnadu, India

[2,3,4]Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Anna University, Tamilnadu, India

ABSTRACT

An University's reputation and its standard are weighted by its students performance and their part in the future economic prosperity of the nation, hence a novel method of predicting the student's upcoming academic performance is really essential to provide a pre-requisite information upon their performances. A machine learning model can be developed to predict the student's upcoming scores or their entire performance depending upon their previous academic performances.

Keywords : Machine learning- supervised learning - prediction models-EPP Algorithm -naïve bayes algorithm - Decision tress - Regression models-Statistics

I. INTRODUCTION

Making higher education affordable has a significant impact on ensuring the nations' economic prosperity and represents a central focus of the government when making education policies. Yet student loan debt in the India has blown past the billion-dollar mark, exceeding Indians combined credit card and auto loan debts . As the cost in college education has skyrocketed over the past few decades, prolonged graduation time has become a crucial contributing factor to the ever growing student loan debt. In fact, recent studies show that more than 580 public four-year institutions in the United States have on-time graduation rates at or above 50 percent for their full-time students. To make college more prohibitory, it is crucial to ensure that more students graduate on time through early conciliation on students whose performance will be strained to meet the graduation criteria of the degree program on time.

II. EXISTING CONCEPT

Machine learning for enlightenment has attain much notice in fresh years. A affluent amount of literature anchor on predicting student performance in solving problems or achieved courses. Many machine learning techniques, such as decision trees, artificial neural networks, matrix factorization, collaborative filters and probabilistic graphical models, have been applied to develop prediction algorithms[1] [2]. Most of this work ignores temporal / sequential effect that students improve their knowledge over time and treats the prediction as a one-time task. To grab the secular/sequential effect into account, `a three mode tensor factorization (on student/problem/time) mastery was developed for predicting student performance in crack problems in ITSs and a similarity-based algorithm was proposed to issue predictions of student grades in courses only when a certain confidence level is reached. However, due to the aforementioned notable differences of predicting student performance in these methods are not pertinent in our mounting[7]. Our escalating prediction algorithm uses the ensemble learning manner, in distinct, the Exponentially Weighted Average Forecaster (EWAF)[6] as a production block to qualify progressive prediction of student performance and online modernize of the predictor as the current student data is sustained. The prime dissention from the standard EWAF algorithm is that an ensemble predictor has ingress to multiple base predictors (experts) as well as the foregoing-term ensemble predictor, whose output condense the outputs of all premature-term base predictors (experts) whereas the conventional EWAF algorithm has spasm to all experts directly[4].

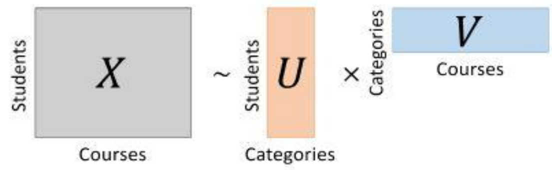
A. Algorithm 1 Ensemble-based Progressive Prediction (EPP)

1: Initialization: $L(ht) = 0, L(ft) = 0, \forall t$.
2: for each student i do
3: Observe backgrounds θi, student group gi
4: for term $t = 1$ to T do ◃ *Prediction Phase*
5: Observe performance state xti
6: Extract relevant state $\tilde{}xti$
7: Receive prediction $\hat{}yt{-}1$
i from $ft{-}1$
8: Base predictor $h \in Ht$ predicts zt
$h;i = ht(\theta i, \tilde{}xti)$
9: Ensemble predictor ft predicts
10: $\hat{}yti = ft(\hat{}yt{-}1i, \{zth;i\}h/vt{-}1i, wti)$
11: end for
12: Observe true label yi.
13: for term $t = 1$ to T do ◃ *Update Phase*
14: Compute prediction loss $l(\hat{}yti, yi)$ and $l(zti;ht, yi)$
15: Update $Li(ht/gi) \leftarrow Li{-}1(ht/gi) + l(zti;ht, yi)$
16: $Li(ft{-}1/gi) \leftarrow Li{-}1(ft{-}1/gi) + l(\hat{}yti, yi)$
17: Update weights $wti{+}1$ and vt
18: end

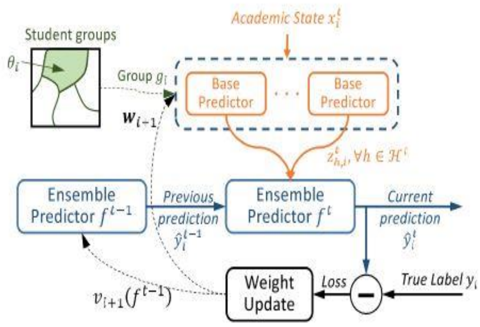
B. Limitations

1. Requires huge set of data to train the predictor.
2. 2. Requires a bilayered structure.
3. Factorization of matrices are complex to have a maximized prediction

C. Matrix Factorization Illustration



D. Overview of Existing System



III. PROPOSED CONCEPT

The Proposed concept follows a classifier constructing technique i.e. Naive bayes technique for having independency among the values. Naive bayes is a group of algorithms combined together for a common concept or a goal. As our aim is predicting the results or performance scores mere accurately, we call it as a probability model in which this naive bayes classification technique can be used as a supervised learning model[5].

Another technique which is used along with the naïve bayes classification technique is the Regression technique[8]. Regression is a statistical model estimating variables. The various dependencies and independencies among the different types of variables are analysed by this regression model. The major reason for simultaneously using this two techniques is to prevent the unwanted variable illusions or unreal relationships in the variables during the prediction process. This issue rises in the regression phase were it is solved by the naïve bayes phenomenon.

For regression there should be sufficient amount of data to be used which makes it testable for later actions. Assumptions of process are the evidence for knowing the data generating process, this makes the prediction of a student's performance more accurately predictable than the other methods.

The aim is to predict the final cumulative GPA of a student in a certain area at the end of each term or semester. Specifically, at the end of each term , the predictor outputs a prediction GPA given student backgrounds and the evolving performance . However, since the cumulative GPA is a function of all trail grades, namely the contaminated average of all course grades.

A below chart of a basic regression model will merely be the model for the proposed methods prediction process where the x-axis holds for term or semester and the y-axis holds for the marks or the CGPA based on the designation variables given to the predictor[3]. In this way we can easily classify various student performances in each term or term by term.

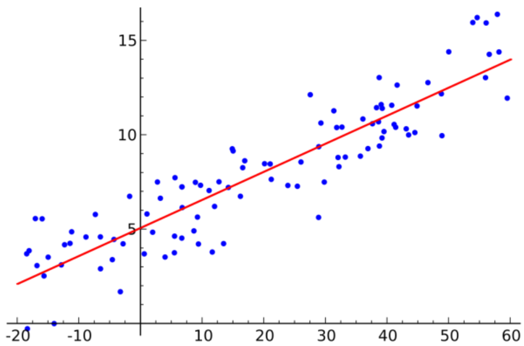


Fig 1. A sample of a Regression chart

Basic Algorithm of Naive Bayes

TRAIN

```
$samples = [[60], [61], [62], [63], [65]];
$targets = [3.1, 3.6, 3.8, 4, 4.1];

$regression = new LeastSquares();
$regression->train($samples, $targets);
```

PREDICT

```
$regression->predict([64]);
```

REGRESSION

```
$samples = [[73676, 1996], [77006, 1998], [10565, 2000], [146088, 1995], [15000, 2001], [65940, 2000], [9300, 2000], [93739, 1996], [153260, 1994], [17764, 2002], [57000, 1998], [15000, 2000]];
$targets = [2000, 2750, 15500, 960, 4400, 8800, 7100, 2550, 1025, 5900, 4600, 4400];

$regression = new LeastSquares();
$regression->train($samples, $targets);
$regression->predict([60000, 1996])
```

INTERCEPT AND CO-EFFICIENTS

```
$regression->getIntercept();
// return -7.9635135135131

$regression->getCoefficients();
// return [array(1) {[0]=>float(0.18783783783783)}]
```

IV.CONCLUSION

In this paper, we proposed a novel method for predicting students' future performance in degree programs given their current and past performance. A regression based naïve bayes model is developed for the prediction process. this performance prediction technique can be helpful to elective courses and use the prediction results to recommend courses to students.

V. REFERENCES

[1]. The White House, "Making college affordable," https://www.whitehouse.gov/issues/education/ highereducation/making-college-affordable, 2016.

[2]. Complete College America, "Four-year myth: Making college more affordable," http://completecollege.org/wp-content/uploads/2014/11/4-Year-Myth .pdf, 2014.

[3]. H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis-a general method for cognitive model evaluation and improvement," in International Conference on Intelligent Tutoring Systems. Springer, 2006, pp. 164-175.

[4]. M. Feng, N. Heffernan, and K. Koedinger, "Addressing the assessment challenge with an online system that tutors as it assesses," User Modeling and User-Adapted Interaction, vol. 19, no. 3, pp. 243-266, 2009.

[5]. H.-F. Yu, H.-Y. Lo, H.-P. Hsieh, J.-K. Lou, T. G. McKenzie, J.-W. Chou, P.-H. Chung, C.-H. Ho, C.-F. Chang, Y.-H. Wei et al., "Feature engineering and classifier ensemble for kdd cup 2010," in Proceedings of the KDD Cup 2010 Workshop, 2010, pp. 1-16.

[6]. Z. A. Pardos and N. T. Heffernan, "Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset," Journal of Machine Learning Research W & CP, 2010.

[7]. Y. Meier, J. Xu, O. Atan, and M. van der Schaar, "Personalized grade prediction: A data mining approach," in Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015, pp. 907-912.

[8]. C. G. Brinton and M. Chiang, "Mooc performance prediction via clickstream data and social learning networks," in 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 2299-2307.

Cite this article as :

C. Selvi, R. Shalini, V. Navaneethan, L. Santhiya, "A Machine Learning Approach for Predicting Student Performance", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 462-465, March-April 2019. Available at doi :
https://doi.org/10.32628/CSEIT1952106
Journal URL : http://ijsrcseit.com/CSEIT1952106