

Performance Prediction in Blended Courses

P.Divya¹, M.Uha Sree², N.Hitesh³, P.Ramulu⁴, K.Rama Krishna⁵

^{1,2,3,4}UG Scholar, Department of Computer Science & Engineering, Gudlavalleru Engineering College, AP, India

⁵Assistant Professor, Department of Computer Science & Engineering, Gudlavalleru Engineering College, AP, India

Abstract - Blended courses that mix in-person instruction with online platforms are increasingly common in secondary education. These platforms record a rich amount of data on students' study habits and social interactions. The main objective of the paper is to inform a student about his performance condition on particular course based on the test values. There are several parameters which effect the student performance level like gender, father's occupation and marks secured in prior years. The research is mainly aimed at predicting whether the student performance condition on particular course is a satisfactory or not. It is also mainly useful to larger institutions like Universities for predicting the performance of students. Machine learning techniques can be utilized for students' grades prediction in different courses. Such techniques would help students to improve their performance based on predicted grades and would enable instructors to identify such individuals who might need assistance in the courses.

Key Words: Course, performance, SVM, Accuracy, Random Forest .

1. INTRODUCTION

Mixed courses that blend face to face guidance with online stages are progressively basic in auxiliary instruction. That is fundamental so as to help in danger understudies and guarantee their maintenance, giving the brilliant learning assets and experience, and improving the foundation's positioning and notoriety. These stages record a rich measure of information on understudies' examination propensities and social collaborations. Notwithstanding, that may be hard to be accomplished for startup to fair sized colleges, particularly those which are had practical experience in graduate and post graduate projects, and have little understudies' records for examination. Without getting any help on the course area and its intricacy, it might demotivate an understudy and can be the reason to pull back the course. In this way, the primary point of this task is to demonstrate the chance of preparing and displaying a little dataset size and the practicality of making an expectation model with sound exactness rate. There are a few parameters which impact the understudy execution level like father's occupation and imprints made sure about in earlier years. This examination investigates too the chance of distinguishing the key markers in the little dataset, which will be used in making the forecast model, utilizing grouping calculations. Best pointers were taken care of into numerous AI calculations to assess them for the most precise model.

The principle results of this examination have demonstrated the proficiency of delivering a satisfactory grouping's exactness and unwavering quality test rates.

2. RELATED WORK

Diverse AI calculations are accessible to perform determining of understudies positions and exhibitions. The issue raises when we picked calculation. The calculations of various kinds are recorded underneath. Information Mining Techniques For Predicting Student Performance Predicting Student Performances to forestall or avoid any and all risks against understudy disappointments or dropouts is enormous these days. Understudy disappointment and dropout is a significant issue nowadays. There can be various factors affecting understudy dropouts. Data mining can be used as a convincing technique to recognize and anticipate these dropouts. In this paper, a gathering method for desire is been discussed. Choice tree classifiers are used here and methods for fathoming the class lopsidedness issue is expansion.

There are a few kinds of measures for assessing the achievement of models. In any case, the assessment of each model relies intensely upon the area and framework's objectives. For our framework, we will probably foresee understudies' GPA and settle on choices if an understudy needs to strive to finish the course. These choices function admirably when our forecasts are precise. To accomplish it, we need to analyze the expectation GPA against the real GPA for the understudies courses pair.

Support Vector Machine and K-Nearest Neighbour for Student execution expectation. This work showed two expectation models for the estimation of understudies execution in last assessment. The work utilized the well known dataset given by the University of Minho in Portugal, which relate to the execution in math subject and it contains 395 data tests. Foreseeing the exhibition of understudies can be useful in maintaining a strategic distance from potential hazard, second exercises, or picking an understudy that is fit for a particular task. The need to explore better models to achieve better execution can't be overemphasized. By far most of earlier work on the proportionate dataset used K-Nearest Neighbour algorithm and achieved low results, while Support Vector Machine count was every so often used, which happens to be an astoundingly pervasive and incredible forecast procedure system.

3. METHODOLOGY

AI is a strategy for information examination that computerizes logical model structure. It is a part of man-made brainpower dependent on the possibility that frameworks can gain from information, distinguish examples and settle on choices with negligible human intercession. There are four sorts of AI calculations that are at present being utilized. Figure 1, shows these four kinds of AI calculations. Administered learning includes order and relapse issues. It is utilized for the most part for prescient investigation as it constructs a model from information, this information additionally incorporates the results or reactions. Model is prepared utilizing marked information.

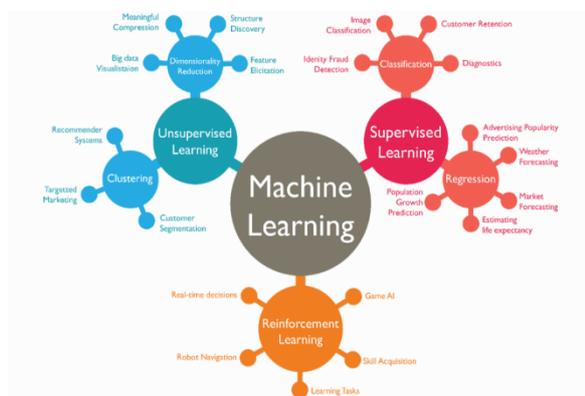


Fig-1: Types of machine learning algorithms

In this model, six AI calculations are utilized in the model. These six calculations are K - Nearest Neighbours (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR) and Random Forest (RF).

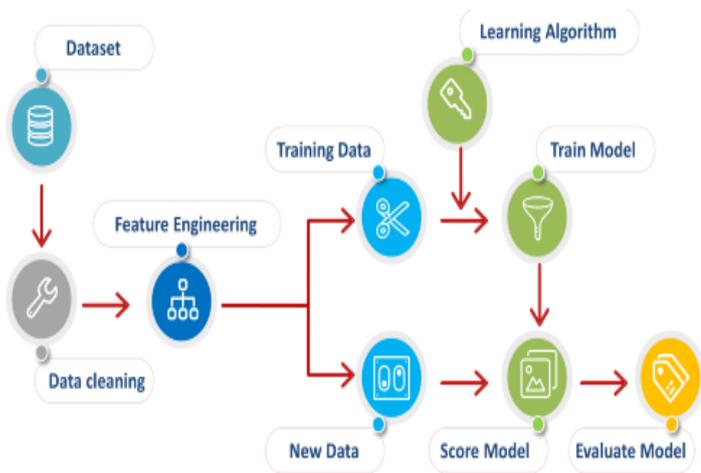


Fig-2: Architecture Model

Support Vector Machine (SVM)

SVM is one of the standard arrangement of directed AI model utilized in grouping. Given a two-class preparing test the point of a help vector machine is to locate the best most noteworthy edge isolating hyper plane between the two classes. For better speculation hyper plane ought not lies nearer to the information focuses have a place with the different class. Hyper plane ought to be chosen which is a long way from the information focuses from every classification.

Accuracy Measures

SVM, Decision Tree, Random Forest, KNN, calculations are utilized in this examination work. Tests are performed utilizing inside cross-approval 10-folds. Exactness, disarray grid and ROC (Receiver Operating Curve)measures are utilized for the arrangement of this work.

In this venture, we have utilized the Support Vector Machine calculation. Bolster vector machine or SVM calculation depends on the idea of choice planes, where hyper planes are utilized to group a lot of given items.

As should be obvious in Figure 4, we have two arrangements of information. These datasets can be isolated effectively with the assistance of a line, called a choice limit.

Favourable circumstances of Support Vector Machine Algorithm

- Accuracy
- Works very well with restricted datasets
- Kernel SVM contains a non-straight change capacity to change over the confused non-directly distinct information into directly distinguishable information.

Problem Statement

Use Machine Learning to predict student performance using previous history and Student Data.

Dataset: Student Dataset

Stage 1: Load Pandas library and the dataset utilizing Pandas

Stage 2: Define the highlights and the objective

Stage 3: Split the dataset into train and test utilizing sklearn before building the SVM calculation model.

Stage 4: Import the help vector classifier capacity or SVC work from Sklearn SVM module. Assemble the Support Vector Machine model with the assistance of the SVC work.

Stage 5: Predict values utilizing the SVM calculation model.

Stage 6: Evaluate the Support Vector Machine model.

Once the model trains itself using training data, testing data was used for predicting the responses and checking the accuracy, and lastly the model was evaluated. This process was followed for all 6 machine learning algorithms used in this project. Experiments were performed and results were obtained.

SVM is one of the standard arrangement of managed AI model utilized in characterization. Given a two-class preparing test the point of a help vector machine is to locate the best most elevated edge isolating hyper plane between the two classes. For better speculation hyper plane ought not lies nearer to the information focuses have a place with the different class. Hyper plane ought to be chosen which is a long way from the information focuses from every classification.

Random Forest is an adaptable, simple to utilize AI calculation that produces, even without hyper-parameter tuning, an extraordinary outcome more often than not. It is additionally one of the most utilized calculations, since it's straightforwardness and the way that it very well may be utilized for both grouping and relapse errands. In this post, you will realize, how the irregular woods calculation works and a few other significant things about it.

Statistical techniques and data visualization

Correlation graph: Correlation values range between -1 and 1.

There are two key components of a correlation value:

- magnitude – The larger the magnitude (closer to 1 or -1), the stronger the correlation
- sign – If negative, there is an inverse correlation. If positive, there is a regular correlation.

The dataset is used to demonstrate each plot. This dataset describes whether or not each patient will have an onset of diabetes within five years. As such it is a classification problem.

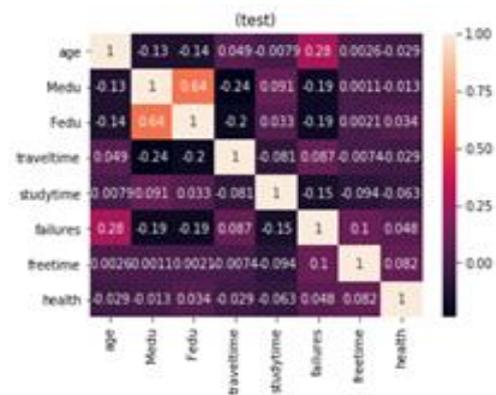


Fig-3: Correlation Graph

4. ANALYSIS

In this exploratory examination, a portion of the AI calculations were utilized. These calculations are NB, KNN, SVM, DT and RF. Every one of these calculations were applied on the dataset. Information was isolated into two bits, preparing information and testing information, both these segments comprising 70% and 30% information individually.

All these four calculations were applied on same dataset and results were gotten. Anticipating exactness is the fundamental assessment parameter that we utilized in this work. Exactness can be characterized utilizing condition 1. Precision is the general achievement pace of the calculation.

For SVM, by plotting AUC graph the accuracy percentage was 80%. The parameters which are given to plot the graph were label and color name. The legend() is automatically determined, when you do not pass in any extra arguments.

For Random Forest, by plotting AUC graph the accuracy percentage was 75%. The parameters which are given to plot the graph were label and color name. The legend() is automatically determined, when you do not pass in any extra arguments.

5. CONCLUSION

Prediction of student’s performance became an urgent desire in most of educational entities and institutes. In this study, systematic efforts are made in designing a system which results in the prediction of student performance on particular course. During this work, six machine learning classification algorithms are studied and evaluated on various measures.

In this paper, we used five popular machine learning algorithms for predictive analytics. These algorithms include SVM, KNN, DT, LR, RF and NB. Predictions were made on Student dataset consisting 1000 records. Some of the

attributes were selected for training and testing the predictive model. From the experimental results obtained, it can be seen that SVM gives highest accuracy for predicting performance. These algorithm provide 80% accuracy which is highest as compared to other four algorithms used in this paper. Therefore, it can be concluded that SVM is appropriated for predicting the performance of the student on particular course.

6.FUTURE SCOPE

Test results decide the amplexness of the planned framework with an accomplished exactness of 0.80 utilizing the Support Vector Machine calculation. In future, the structured framework with the pre-owned AI order calculations can be utilized to anticipate understudy execution in advanced education. The work can be expanded and improved for the presentation in online courses including some other AI calculations.

7.REFERENCES

[1] M.Ahmadzadeh, D. Elliman, and C. Higgins, "An analysis of patterns of debugging among novice computer science students," *ACM SIGCSE Bull.*, vol. 37, pp. 84–88, 2005.

[2] B. Amnueypornsakul, S. Bhat, and P. Chinprutthiwong, "Predicting attrition along the way: The UIUC model," in *Proc. EMNLP Workshop Anal. Large Scale Soc. Interact. MOOCs*, 2014, pp. 55–59.

[3] J.M. L. Andres, R. S. Baker, D. Gasevic, G. Siemens, S. A. Crossley, and S. Joksimovic, "Studying MOOC completion at scale using the MOOC replication framework," in *Proc. 8th Int. Conf. Learn. Analytics Knowl.*, 2018, pp. 71–78..

[4] Bandura and R. H. Walters, *Social Learning Theory*, vol. 1. Englewood Cliffs, NJ, USA: Prentice-Hall, 1977.

[5] P. Blikstein, "Using learning analytics to assess students' behavior in open-ended programming tasks," in *Proc. 1st Int. Conf. Learn. Analytics Knowl.*, 2011, pp. 110–116.

[6] Information retrieval. *SIAM review* 37, 4, 573–595.

[7] CALDERS, T. AND PECHENIZKIY, M. 2012. Introduction to the special section on educational data mining. *ACM SIGKDD Explorations Newsletter* 13, 2, 3–6.

[8] CALLENDER, C. AND FELDMAN, R. 2009. Part-time undergraduates in higher education: A literature review. Prepared for HECSU to inform Futuretrack: Part-time students. London, Birkbeck, University of London.

[9] U. K. Pandey, and S. Pal, "A Data mining view on class room teaching language", (*IJCSI*) *International Journal of*

Computer Science Issue, Vol. 8, Issue 2, pp. 277-282, ISSN:1694-0814, 2011.

[10] *International Journal of Advanced Computer and Mathematical Sciences*. Bi Publication-BioIT Journals, 2010.

[11] Cristobal Romero, Sebastian Ventura, "Educational Data Mining: A Review of the State of the Art", *IEEE Transactions on system, man and cybernetics-Part C: Application and Reviews*, Vol.40, No.6, pp 601-618, Nov 2010.

[12] U. Fayadd, Piatesky, G. Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", *AAAI Press / The MIT Press*, Massachusetts Institute Of Technology. ISBN 0-262 56097-6, 1996.

[13] DEERWESTER, S., DUMAIS, S. T., FURNAS, G. W., LANDAUER, T. K., AND HARSHMAN, R. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science* 41, 6, 391.