A Hybrid Data Mining Approach To Predict Undergraduate Student Academic Performance Predictions

Rashmi V. Varade[1] , Dr. Blessy Thankachan[2]

[1,2]Jaipur National University

Abstract: -

For students, professors, and academic administrators, predicting student success is a critical issue. Teachers may find it useful to use the findings of a prediction model to determine the level of their present pupils and take proactive initiatives in their teaching tactics. It might also be useful in implementing appropriate academic initiatives at the programme level, as well as analysing student weaknesses.

The rapid advancement of technology in recent years has made it possible to collect massive amounts of data, such as student data, alumni data, course data, and much more, all of which are stored in databases. We need to be able to extract important information from databases.

Data mining, often known as "Knowledge Discovery in Databases," is a technique for extracting information from databases. Many tools and algorithms, such as decision trees, KNN, Random forest Naive Bayes, support vector machine, and others, may be used in data mining in an academic setting to enhance creating techniques for discovering academic databases, which is referred to as 'Educational Data Mining. So we use Hybridisation technique in which the genetic algorithm is used with SVM algorithm which gives highest accuracy than other algorithms.

Keywords: Genetic Algorithm, Student Graduation, Hybrid Optimization.

INTRODUCTION: -

Graduation is an important element of any student's education, and one of the key goals for students seeking education, including diploma students, is to graduate.

Each student who has committed to complete all graduation requirements can graduate.

Students must meet a variety of standards, which vary depending on the degree of education and further education. Graduation is beneficial not just to students, but also to lecturers, guardians, and institutions in general. It is critical for pupils to be carefully evaluated and authorised as soon as possible by all parties involved. Implementing and enforcing standards relating to student graduation is the most fundamental thing that can be done. All parties involved will benefit from policies relating to student graduation. With the existence of the

application, it is envisaged that the optimal implementation and prediction of students would result in a beneficial outcome for all parties involved in student graduation.

Then, by collecting and arranging diverse data, we need a strategy for forecasting or categorising student data, and the process of assessing and making judgments becomes critical. The issue was that many students were unable to graduate on time owing to a variety of issues that posed a challenge for a postsecondary school. In a higher school, students who graduate on time are extremely beneficial in terms of accreditation and quality assurance. [1]

Decision Tree (DT), Naive Bayes (NB), Neural Network (NN), and Support Vector Machine are some of the data analysis approaches that have been used by researchers (SVM).

SVM, DT, and NN are the approaches that have been used to check the data of postgraduate students, and SVM has produced the best results. [2] The optimum subset of the original subset is chosen via feature selection. Excessive, unnecessary, or uninteresting data is removed via feature selection. [3] It is a strategy for overcoming the challenge of small sample classification) and finding a reasonable answer, using statistical learning theory and data sets to examine all sorts of objects. However, selecting the optimal SVM parameters is quite challenging. SVM is still restricted since it requires pre-setting of parameters and can result in under fitting and overfitting. To solve the problem of choosing SVM parameters, an algorithm is required. [4] The search for genetic algorithms is iterative, with each iteration assessing, choosing, and reconnecting strings in the population until they reach one of many termination criteria. The goal of GA is to enhance the quality and accuracy of models.

Students should be able to graduate on time using the GA-SVM approach. [5]

LITERATURE REVIEW: -

Hybrid Prediction Models-Prediction models have improved in effectiveness and efficiency as a result of the hybridization of diverse methods. Here we combine SVM with genetic algorithm for improvement of performance accuracy. [6] A thorough assessment of supervised machine learning algorithms was undertaken and used to predict students' test performance. They looked at a variety of indicators, including factors, to predict students' projected final grades as well as those who were at danger. Luthfia et al. [8] used Wrapper and Information Gain as feature selection strategies with a single classification algorithm: NB (Naive Bayes), DT (Decision Tree), and NN (Naive Neural Network) (Neural Network).

They improve the accuracy of the models they created by using feature selection approaches. The combination of information gain and ANN yielded a score of 79.375 for accuracy.

In their studies, Calderon et al. [8] used the ANN method with the Genetic Algorithm (GA) feature selection for 1271 data sets with 39 characteristics, the data set being the property of a university in southern Peru. The Genetic Algorithm adjustment settings are as follows: population size is 25, crossover level is 0.6, and mutation level is 0.3. Three layers (1 input layer, 1 hidden layer, and 1 output layer) are applied to the ANN algorithm parameters.

The experimental findings led to the selection of 14 characteristics for use in prediction.

When compared to predictions made with merely an ANN classifier, the accuracy value rose by 8%. The bulk of feature selection algorithms, on the other hand, aim for feature selection solutions that fall somewhere between sub-optimal and almost-optimal; rather than searching worldwide, these searches are carried out locally during the search process.

As a result, using this approach, getting an almost ideal answer to the optimal solution is challenging. [9] The goal of this study is to use GA to increase prediction accuracy in the context of students' academic achievement, as well as to solve difficulties with existing feature selection. Because GA can find answers everywhere in the search space because to its global search capabilities, feature selection using a genetic algorithm can give good results in a fair amount of time.[10]

B. Genetic Algorithm: - Among the various evolutionary algorithms based on biological concepts, the Genetic Algorithm (GA) is one of the most widely used approaches for data reduction and feature selection. For those situations, GA is one of the competent data reduction, feature selection, and global optimization algorithms. It operates by establishing optimum operators on its selection, crossover, and mutation processes to govern some people. [11]

METHODOLOGY: -

Dataset:-In this study we used undergraduate student's dataset of 600 students in which 20 attributes are taken and 70% of the data were used as the training set, and the remaining 30% was used for testing purpose.

Previous research that used GPA criteria were reviewed, and the outcomes from student graduation were inconsequential. The most accurate research on student graduation has been done utilising decision trees (DT), KNN, and nave Bayes (NB) approaches. Other studies have also been done using these three methods and adding linear regression to the data.

NB. Obtain the highest level of precision

It will be examined utilising hybrid optimization employing a support vector engine based on evolutionary algorithms, based on studies connected to student graduation.

A genetic algorithm model is combined with a support vector machine in the hybrid optimization activity. Existing parameters have a significant impact on the SVM classification's accuracy. In bigger training sets, GA is utilised to identify better parameter combinations in SVM train classifiers.

RESULTS AND DISCUSSION: -

Accuracy, precision, recall, and F-score were among the assessment measures used to assess each prediction model's performance.

True-Positive (TP), False-Positive (FP), True-Negative (TN), and False-Positive (FP) prediction models were developed (FP).

We found that Hybrid approach gives highest accuracy by using SVM with GA.

Algorithm	Accuracy	Precision	Recall	F-score
SVM	90%	90%	92%	91%
SVM+GA	96%	96%	96%	96%
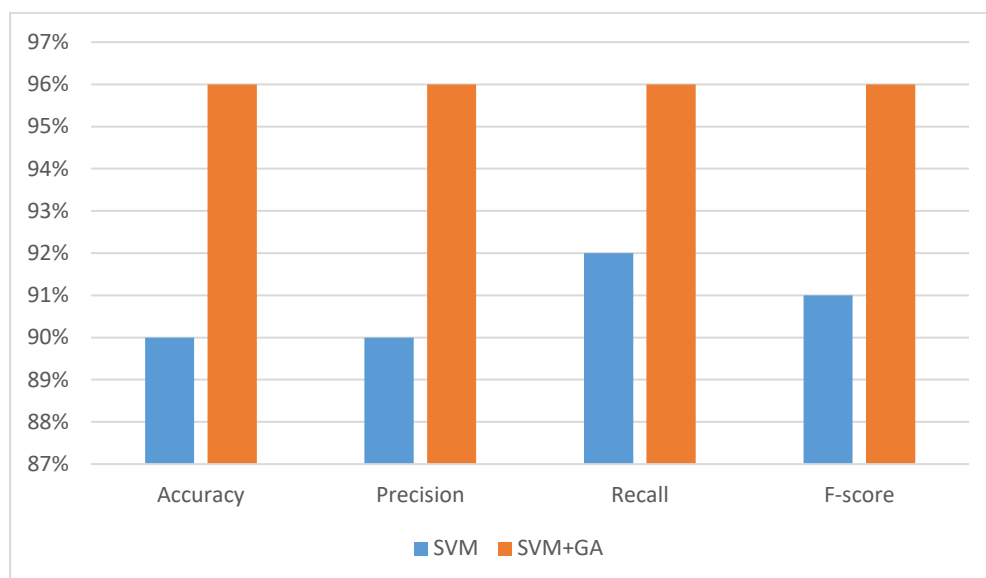
CONCLUSION: -

Not just for graduate students, but for all parties involved in education, it is critical to do further study in order to anticipate when students will graduate.

To predict student graduation, we use SVM-GA hybrid optimization.

GA is utilised in the GA-SVM technique to determine parameters that are appropriate for the SVM method's classification.

The GA-SVM classification has a greater accuracy than the SVM classification, according to the findings of the experiments.

As a result, students can graduate on time and as much as feasible

REFERENCES: -

[1] Ridwansyah1; Ganda Wijaya2; Jajang Jaya Purnama3" "HYBRID OPTIMIZATION METHOD BASED ON GENETIC ALGORITHM FOR GRADUATES STUDENTS" Jurnal PILAR Nusa Mandiri Vol. 16, No. 1 March 2020

[2] Riyanto, Verry, Abdul Hamid, and Ridwansyah. 2019. "Prediction of Student Graduation Time Using the Best Algorithm." Indonesian Journal of Artificial Intelligence and Data Mining 2(2)

[3] Jiang, Huiyan, Fengzhen Tang, and Xiyue Zhang. 2010. "Liver Cancer Identification Based on PSO-SVM Model." 11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010 (December): 2519–23.

[4] Liu, Han et al. 2019. "Effective Data Classification via Combining Neural Networks and SVM." Proceedings of the 31st Chinese Control and Decision Conference, CCDC 2019: 4006–9.

[5] Gao, Xiang Ming, Shi Feng Yang, and Yu Hu. 2010. "Leakage Forecasting for Water Supply Network Based on GA-SVM Model." Proceedings of the 2010 Symposium on Piezoelectricity, Acoustic Waves and Device Applications, SPAWDA10: 206–9.

[6] Allemar Jhone P. Delima" An Experimental Comparison of Hybrid Modified Genetic Algorithm-based Prediction Models" International Journal of Recent Technology and Engineering (IJRTE)ISSN: 2277-3878, Volume-8, Issue-1, May 2019

[7] Tomasevic, N.; Gvozdenovic, N.; Vranes, S. An overview and comparison of supervised data mining techniques for student exam performance prediction. Comput. Educ. **2019**, 143, 103676. [CrossRef]

[8] chegaray-Calderon, O.A., Barrios-Aranibar, D.: Optimal selection of factors using genetic algorithms and neural networks for the prediction of students' academic performance. In: 2015 Latin America Congress on Computational Intelligence (LA-CCI). IEEE, Curitiba (2015)

[9] Kabir, M.M., Shahjahan, M., Murase, K.: A new hybrid ant colony optimization algorithm for feature selection. Expert Syst. Appl. 39, 3747–3763 (2012)

[10] mrieh, E.A., Hamtini, T., Aljarah, I.: Preprocessing and analyzing educational data set using X-API for improving student's performance. In: IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). IEEE, Amman (2015)

[11] ahman, L., Setiawan, A.N., Permanasari, E.A.: Feature selection methods in improving accuracy of classifying students' academic performance. In: 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp. 267–271. IEEE, Yogyakarta (2017)

[12] A. B. Hassanat, V. B. S. Prasath, M. A. Abbadi, S. A. Abu-Qdari, and H.Faris, "An improved Genetic Algorithm with a new initialization mechanism based on Regression techniques," Inf., vol. 9, no. 7, 2018.