e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703034 |

# Predictive Analytics in Education: Identifying At-Risk Students through Probability Analysis

**S.Gokulraj[1], K.P.Janani[2], M.Muhammed Hashiq[3], M.Kaviya[4]**

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

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

**ABSTRACT:** Educational institutions place significant emphasis on the academic performance of their students. Early identification of students at risk of academic failure allows for rapid intervention and improved learning outcomes. The aim is to develop a model that not only predicts risk but also provides a probability associated with that prediction using readily available data points such as attendance data, study hours and test scores. We use SMO (Sequential Minimal Optimization) algorithm, Support Vector Machine (SVM) technique for building models. The SMO algorithm is particularly well suited to this approach because it efficiently processes large data sets and is capable of generating probabilistic outputs. Our results show that the proposed model accurately identifies at-risk students and provides informative probability scores. Standardized performance measures are used to assess the accuracy of the model, which indicates its effectiveness in predicting academic risk. Through experimentation, the effectiveness of this hybrid method is demonstrated and its potential to significantly improve the training data analysis.

**KEYWORDS**: Academic Risk Prediction, SMO Algorithm, Support Vector Machine (SVM), Probabilistic Output, Training Data Analysis.

## I.INTRODUCTION

The incorporation of modern technology in education is a result of the on going pursuit of improving learning experiences and optimizing academic outcomes. The use of machine learning algorithms to forecast and evaluate student performance is one such exciting frontier. This emerging subject aims to develop prediction models that can predict a student's academic progress by utilizing the large quantity of educational data that is already available. Through the utilization of machine learning, educators and administrators can acquire significant insights into the variables impacting student performance. This can facilitate the development of tailored support strategies and prompt interventions. By enabling a more flexible and efficient educational system and enabling personalized learning, the nexus of education and technology has the potential to completely transform the way we approach learning.

The ever-changing environment of education has led to the integration of cutting-edge technology, especially when viewed through the prisms of machine learning and predictive analytics, in an effort to get insights into student performance. Large-scale datasets that have been integrated into educational systems have created new avenues for studying and predicting student results. A growing area at the nexus of technology and education is called "student performance prediction," which uses sophisticated algorithms to evaluate past data, spot trends, and build prediction models. With the help of these models, educators and administrators can proactively identify potential academic difficulties and provide tailored tactics and timely interventions to support each student's unique learning path.

## II.LITERATURE REVIEW

Hana Abdullah Monash [1] makes the basic point that picking candidates who are probably going to succeed scholastically in advanced education organizations requires a confirmations framework based on authentic and dependable affirmations models. This study centers around how information mining methods can be utilized to anticipate candidates' scholarly accomplishment at college to help colleges in going with affirmations choices. The recommended strategy was approved utilizing an informational index of 2,039 understudies selected at a Saudi state

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703034 |

funded college's Software engineering and Data School somewhere in the range of 2016 and 2019. The discoveries demonstrate that certain pre-confirmation factors (secondary school grade normal, score on the Academic Accomplishment Affirmation Test, and score on the Overall Inclination Test) can be utilized to foresee a candidate's initial college accomplishment before confirmation. Additionally, research indicates that the score on the Scholastic Achievement Admission test is the most reliable pre-admission evaluation tool for a prospective applicant. Along these lines, in affirmations processes, this score should be given more weight. Moreover, we found that the Counterfeit Brain Organization strategy beats the other three arrangement techniques that were inspected (Choice Trees, Backing Vector Machines, and Guileless Bayes) with an exactness pace of more than 79%. Pre-affirmation necessities, understudy execution, execution forecast, instructive information mining, and information mining draws near.

Elman Alayna [2] poses the case that understudy accomplishment is significant in light of the fact that it's habitually used as a presentation marker for instructive establishments. Preventive methodologies related to early ID of in danger understudies can fundamentally build their accomplishment. As of late, there has been a great deal of utilization of AI strategies for expectation. Albeit the writing is loaded with examples of overcoming adversity, these techniques are for the most part accessible to teachers who are capable in "software engineering," or all the more precisely, "man-made consciousness." The facts confirm that making various decisions is vital for the powerful and proficient execution of information mining methods. Determining which machine learning approach is best suited for a given circumstance, defining student success, and choosing which student characteristics to emphasize are examples of these decisions. The objective of this task is to give educators who need to utilize information mining strategies to gauge understudy achievement a nitty gritty arrangement of rules. To do this, the writing has been inspected, and the cutting edge has been collected into a purposeful technique wherein possible decisions and boundaries are entirely examined, upheld by reasons, and made sense of.

Dipta Das [3] puts forward the viewpoint that understudy execution is urgent in delivering the greatest alumni who will be accountable for the country's social and financial turn of events. The work market is moreover worried about understudy accomplishment since the employing of late alumni depends on their scholastic standing. Deciding the reason for an understudy's exhibition fluctuation, hence, offers significant data for creating strategies and projects. Various specialists in different countries endeavor to decide the reason utilizing different information mining procedures. They didn't, be that as it may, all work with students from Bangladesh. A model was presented in this study to anticipate outcomes and identify the most important variables affecting Bangladeshi students' academic performance. In this review, a strategy that can perceive which understudies need additional support is proposed. To pick significant highlights, an assortment of element determination methods was applied, including co-connection, chi-square, and Euclidean distance. Furthermore, showing the consequences of element choice examinations utilizing the classifier calculations of fake brain organizations, choice trees, Credulous Bayes, and K-Closest Neighbour.

Faizaan Rouser [4] suggests that recognizing in danger understudies at scholarly organizations at the earliest opportunity is a huge trouble. Any educational establishment's goal is to create a learning environment that maximizes student performance and quickly identifies at-risk students. Toward the finish of a semester, disappointment rates will decline because of early distinguishing proof of in danger understudies. The motivation behind this review is to analyze a few characterization procedures, for example, capability based, tree-based, managed based, languid based, and Bayes-based calculations, to recognize in danger understudies from the get-go in the semester in light of engaging information. Before the semester closes, understudies who are recognized as being in danger of fizzling may get specific advising and mentoring. In the current review, grouping calculations are additionally contrasted with deference with their TP rate, FP rate, precision, review, Measure, exactness, kappa measurements, and model development time. A veritable dataset was utilized for the investigations, and the outcomes exhibit the cutthroat presentation of different calculations.

Lassa ad K. Smyrna [5] et al. propose that the world is seeing an unrivaled flood in the development of living structures because of the headway of data and correspondence innovation (ICT). The social, financial, and social aspects of life are continually affected by ICT. Various countries are endeavoring to execute the data society

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703034 |

effectively. Through strategies and foundation to propel information obtaining and make astute social orders. Because of the speedy headway of shrewd gadgets and specialized networks, various Learning, The Board Framework (LMS) arrangements have been accessible in the scholarly space. Various researchers have zeroed in their endeavors on making new ways to deal with manage these issues in this climate. The primary influx of scientists zeroed in on versatile learning and proposed frameworks as replies to these issues, as. They assert that adaptive education systems may significantly aid students' academic progress through the utilization of individualized environments, such as teaching strategies, resources, and evaluations. Nonetheless, these frameworks keep on confronting a lot of trouble with respect to the type of the educational techniques and point matter. To further develop understudy grouping execution and result expectation, a Stacked Speculation for Disappointment Expectation (SGFP) is utilized in this review.

## III. PROPOSED WORK

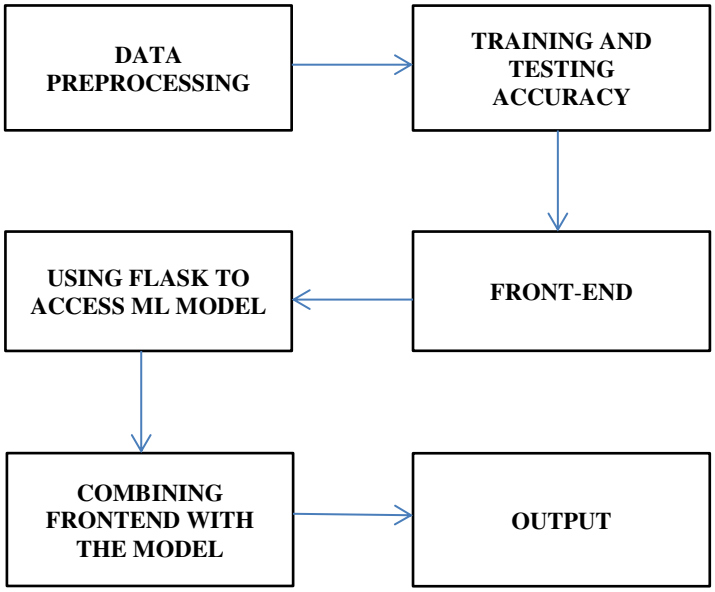


**Figure 1:** Workflow of Methodology

### DATA PREPROCESSING

Using data preprocessing techniques, we were able to build a risk prediction model by studying student data. The task involved addressing missing values, outlier detection, and scaling properties to ensure machine learning algorithms are smoothly scaled. We also addressed potential inconsistencies by standardizing assessment formats and establishing a threshold for classifying students into at-risk groups based on historical performance data. A clean and standardized data set was utilized to construct a precise model that predicted students' likelihood of academic failure.

### TRAINING AND TESTING ACCURACY

Training accuracy is assessed by the model's ability to understand certain data and recognize patterns and trends inherent in it. However, test accuracy is the measure of a model's generalization capabilities and performance with unrevealed information. Higher training accuracy means that the model fits the training data well, while better testing accuracy indicates its effectiveness in predicting new cases, which increases its reliability in identifying risk groups.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703034 |

**FRONT-END**

Employing HTML, CSS and JavaScript, the user interface is designed to seamlessly integrate with predictive algorithms, providing educators and administrators with an easy-to-use platform to access information quickly and efficiently. Responsive design and user-centered features such as interactive dashboards and real-time updates allow stakeholders to make informed decisions to intervene and support students, ultimately creating a learning environment conducive to success.

**USING FLASK TO ACCESS ML MODEL**

Flask routes are responsible for managing the data, preparing it for the model, and initiating predictions. The outcome of the risk assessment can be showcased on the web interface, providing educators with valuable data-driven insights to intervene and assist students facing challenges. This method presents a user-friendly and easily accessible way to implement ML models in educational environments, enabling timely identification and tailored support for students at risk.

**ALGORITHM DETAILS**

Sequential Minimal Optimization (SMO) is a popular supervised machine learning model used for classification and prediction of unknown data. Several researchers claim that SMO is a very accurate text classification technique. It is also widely used in emotion classification. The model can be trained to classify new information into positive and negative reviews using pre-labeled data. This is exactly how minimal sequential optimization works. This is the model we train on the dataset so that it can analyze and classify the unknown data into classes in the training set. Sequential minimum optimization is a linear learning method. It finds the optimal hyperplane to separate the two classes. Since it is a supervised classification model, it tries to maximize the distance between the nearest training point and each class to get a better classification result on the test data. The function evaluation procedure is as follows:

• It takes a labeled data sample and draws a line separating the two classes. This line is called the decision limit. Only those training data points closest to the decision boundary are considered in the solution. The data points are called support vectors. For example, the classification of movie reviews can be characterized either positively or negatively.

• Whenever new data is required for classification, it is positioned to the left or right of the decision boundary. This category will be categorized according to the side in which data is entered. Our data can be classified with the highest level of precision by breaking down these two classes so that a decision boundary separates them as much space as possible.

**IV.RESULT ANALYSIS**

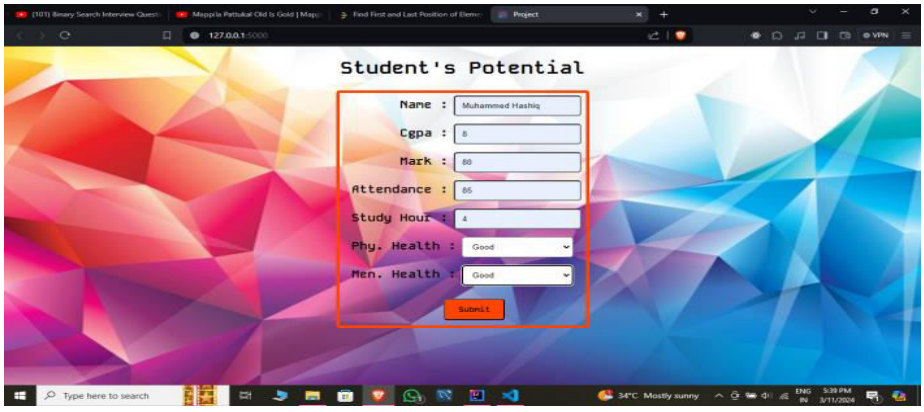


**Figure 2:** Login Page

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703034 |

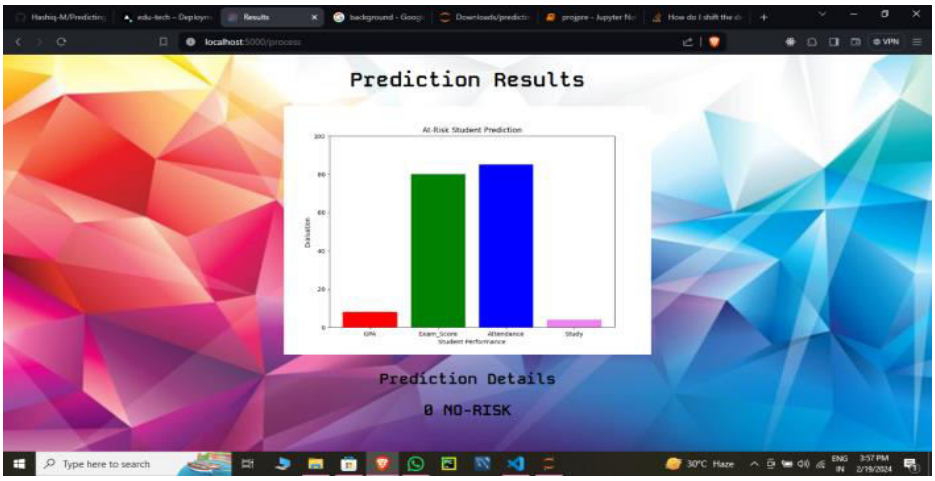


**Figure 3:** Student At No Risk

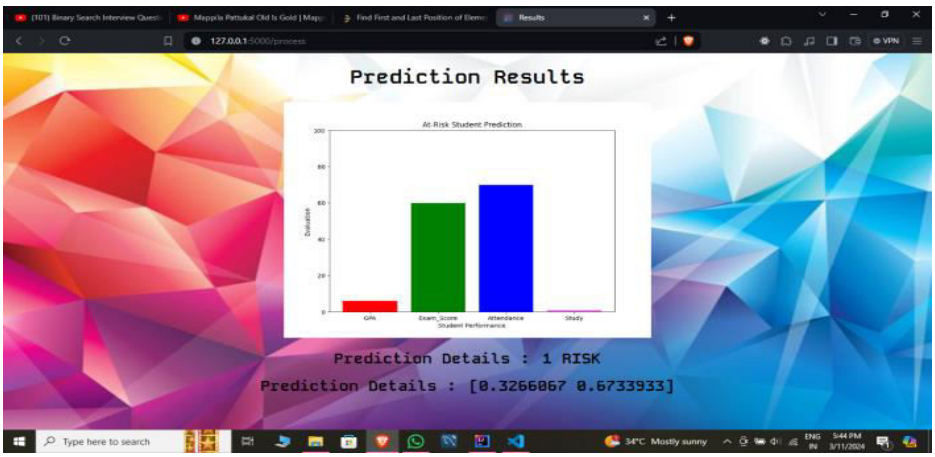


**Figure 4:** Student At Risk

Based on the given evaluation criteria, the Support Vector Machine with Sequential Minimal Optimization (SMO) algorithm performs better in terms of accuracy, computational efficiency, and scalability. With a recall of 0.95 and a high accuracy of 0.97, respectively, SMO shows that it can catch a substantial proportion of the relevant cases and has a low false positive rate. As a result, SMO obtains an impressive F-score of 0.95, which is a harmonic mean of recall and accuracy that successfully balances the two metrics. Furthermore, SMO has an accuracy of 0.95, demonstrating a high level of overall accuracy in classification tests.

## V. CONCLUSION AND FUTURE WORK

In conclusion, this work contributes to educational analysis by creating and implementing a predictive model of student academic performance. The hybrid model successfully predicts student outcomes using Sequential Minimal Optimization (SMO) classification to manage nonlinear interactions. The predictive accuracy of the model is improved by optionally including relevant information such as attendance, past academic records, and socioeconomic indicators. Systematic testing confirms the robustness and reliability of the model and ensures its applicability in different educational contexts. Future research could focus on improving model performance and addressing emerging issues in educational analytics. Exploring advanced ensemble methods or using deep learning architectures can yield more accurate predictions. In addition, adding more details, such as students' extracurricular activities, can provide a more comprehensive view of the factors affecting academic performance.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703034 |

## REFERENCES

[1] Adeniyi, D. A., & Odukoya, J. A."Predicting academic performance of students using machine learning algorithms: A case study of Covenant University, Nigeria". International Journal of Advanced Computer Science and Applications, 10(9), 111-117,2019.
[2] Amiri, S., & Hosseini, S. M."Predicting at-risk students using machine learning algorithms: A case study in higher education". Journal of Computers in Education, 7(1), 1-16,2020
[3] Chen, Y., Liu, D., & Yu, T."Predicting at-risk students using machine learning techniques: A comparative study". Journal of Educational Computing Research, 57(8), 2131-2153,2019.
[4] Gong, Y., & Liu, M."Predicting student dropout risk in higher education: A machine learning approach". Journal of Computers in Education, 7(2), 187-201,2020.
[5] Jha, S., & Mishra, S."Predicting academic performance and identifying at-risk students using machine learning techniques". Journal of Educational Technology Systems, 49(4), 431-451,2021.
[6] Kotsiantis, S. B."Predicting students' academic performance: Comparing artificial neural networks, decision trees, and linear regression". Expert Systems with Applications, 129, 263-275,2019.
[7] Nourani, M., Birjandi, P., & Papi, M."Predicting at-risk students in language learning courses using machine learning techniques". Computer Assisted Language Learning, 33(5-6), 631-654,2020.
[8] Sarwar, F., & Khan, R. A."Predicting students at risk of failing: A comparison of machine learning techniques". Journal of Educational Computing Research, 58(8), 1935-1955,2020.
[9] Al-Radaideh, Q. A., Alsmadi, I. M., & Al-Ayyoub, M."Early prediction of at-risk students using educational data mining". International Journal of Information and Education Technology, 10(1), 20-27,2020.
[10] Chawla, S., & Sengar, N."Predicting academic success of engineering students using machine learning algorithms". International Journal of Engineering Research and Applications, 9(7), 80-85,2019.
[11] Dauda, M. A., & Adeyanju, I. A."Predicting at-risk students in a Nigerian university using machine learning techniques". International Journal of Computer Applications, 181(46), 40-46,2018.
[12] Elbir, A. M., & Cebeci, Z."Predicting at-risk students in higher education using machine learning: A case study in Turkey". International Journal of Educational Technology in Higher Education, 17(1), 1-18,2020.
[13] Khan, M. A., & Salam, A."Predicting at-risk students using machine learning techniques: A case study of a public university in Pakistan". International Journal of Advanced Computer Science and Applications, 11(7), 156-162,2020.
[14] Li, X., & Zhang, Q."Predicting at-risk students using machine learning techniques in Chinese higher education". Journal of Educational Technology Development and Exchange, 13(1), 33-42,2020.

Impact Factor
7.521

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

## IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com