



Review On Educational Academic Performance Analysis and Dropout Visualization by Analyzing Grades of Student

Mansi Choudhari¹, Saloni Rangari², Pratham Badge³, Pratham Chopde⁴, Atharva Paraskar⁵

^{1,2,3,4,5}UG - Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, Maharashtra, India.

Email Id: mansichoudhari11@gmail.com¹, salonirangari2011@gmail.com², pratham4160@gmail.com³, prathamchopde31@gmail.com⁴, atharavparaskar03@gmail.com⁵

Abstract

Education is a crucial aspect of a nation's development, and ensuring the success and retention of students is of paramount importance. The current studies show the need for an effective and efficient education prediction system. Education is a pivotal aspect of a country's development. It acts as a powerful tool to change the world. Education is the key to a literate society. In India, it is necessary to have an integrated web platform to analyze the academic performance and dropout rates across school, higher, and technical education. Student dropout is a significant problem for any nation. Discontinuing schooling due to financial, practical, and social reasons, as well as disappointment in examination results, is what is commonly referred to as student dropout. Educational Data Mining (EDM) techniques can help discover insights from data in educational environments, allowing tutors and researchers to predict future trends and student behavior. The use of machine learning and data mining techniques provides valuable tools for understanding the student learning environment. This literature review aims to synthesize the existing research findings on this topic and identify knowledge gaps for future research.

Keywords: Academic performance; Dropout visualization; Data analysis; Educational data mining; Grades.

1. Introduction

In a country with a population of over 1.3 billion, education is considered to be the foundation for a better future. However, the dropout rates in India, especially in the secondary, higher, and technical education levels, are a cause for alarm. The issue of student dropouts in India is a pressing concern that needs to be addressed urgently. In India, the secondary education level comprises classes 9 and 10, while higher education includes classes 11 and 12. Many students drop out during these crucial two years due to various reasons like financial constraints, lack of interest, and pressure to start earning for their families. This not only affects their academic performance but also hinders their future prospects. Research has shown that there are numerous factors that can predict student dropout rates in India. One such factor is household wealth status, as students from wealthier households are less likely to drop out of school compared to those from poorer households. The process of quitting the

educational system is complicated and influenced by many different circumstances. However, it is a phenomenon that occurs in schools, and various aspects of education may work as threat or protective factors. Reducing primary and secondary school dropout rates is one strategy to enhance the educational system and guarantee that every child receives an equal education. Dropping out is seen by society as an inability to build the human capital required to maintain a thriving economy. Multiple variables, including both cognitive and non-cognitive ones, influence academic success. Predicting and improving student outcomes requires a thorough grasp of the interactions between these variables. It has been discovered that traditional measures, such as test scores, only partially explain the variability in academic performance, highlighting the need for a comprehensive approach. Predictive modeling techniques like educational data mining and graph regularized



robust matrix factorization have been used to improve learning analytics and predict student grades in order to address this. Academic performance, or achievement, is determined by continuous assessment or cumulative grade point average (CGPA) and indicates how well a student, instructor, or institution succeeded in achieving their short- or long-term goals for learning. It is clear that higher education institutions are interested in using student academic performance modeling to enhance the effectiveness and caliber of the conventional procedures. Researchers and tutors could, for example: identify students who would benefit from an intervention if they were at risk of failing some of their courses; ascertain the value of students' response times in performance prediction; recommend activities that optimize the knowledge acquired as measured by anticipated post-test success; and provide tailored support based on each student's academic abilities.

2. Literature Review

To improve classroom performance tracking and support for struggling students using machine learning algorithms. The study assesses the effectiveness of decision tree models, which provide practical information for teachers to take safety precautions. The research aims to provide instructors with supportive measures to monitor students' performance and provide additional attention to struggling students. The relationship between grit, self-efficacy, achievement-oriented goals, and university students' academic success is examined in this paper. Grit and academic success are found to be positively correlated, with self-efficacy coming before achievement orientation targets. While avoidance goals have a detrimental impact on educational outcomes, mastery and approach goals have a good impact. Additionally, the research points to the possibility of using self-efficacy to increase or diminish the impact of particular goals on academic performance. This emphasizes the significance of grit in predicting a variety of educational outcomes and provides insightful information to professionals in the education sector and policymakers to increase the

success rates of their students.[1] The study explores the use of Educational Data Mining (EDM) techniques to predict student behavior and trends. It uses course report data from tutors and assesses five classification models using cross-validation techniques and algorithms. The models can help tutors identify at-risk students early on, enhancing their academic performance. This study investigates how Educational Data Mining (EDM) might be used to forecast students' academic success at the beginning of a semester, particularly in demanding courses like Data Structures and Programming. The study evaluates the effectiveness of a deep neural network (DNN) and conventional machine learning methods in predicting academic achievement and identifying children who are at risk. The suggested DNN model beats alternative techniques, predicting performance in data structure courses with an amazing 89% accuracy. To promote student achievement and lower dropout rates, educational institutions can benefit from the practical consequences of this research.[2] Mesfin Tadese, Alex Yeshaneh, and Getaneh Baye Mulu's 2022 study aimed to identify academic performance determinants among 659 university students in Southern Ethiopia. A self-administered survey was used for data collection, and a p-value of ≤ 0.05 was considered statistically significant. The study found that academic performance was significantly influenced by smoking, age, and field of study.[3] The paper analyzes how to forecast student success in Massive Open Online Courses (MOOCs) using ensemble learning techniques. Data from 480 students are analyzed using three methods: bagging, boosting, and stacking, covering 17 features. The stacking ensemble classifier achieves an 88% accuracy and an AUC of 0.85, outperforming existing techniques. This study advances the field of educational data mining and has the potential to enhance instructional strategies and online learning results.[4] The research investigates the application of convolutional neural networks, discriminant analysis, and clustering to educational data mining to forecast students' academic achievement. It presents an optimization method for figuring out



clustering numbers in K-means algorithms and assesses its performance with discriminant analysis. Labeled data is trained and tested using convolutional neural networks, which produce predictive models for tracking performance in the future. By enhancing student results and instructional strategies, the research advances intelligent technology in education.[6] The study focuses on predicting models for student performance based on previous academic grades to address the important topic of student success in higher education. Recognizing that imbalanced datasets can produce biased findings, the study intends to examine previous research and offer a cutting-edge method for managing imbalanced classification in higher education settings. The research investigates several approaches—data-level, algorithm-level, and hybrid approaches—for resolving unbalanced classification through an extensive review of the literature from 2015 to 2021. The results show that SMOTE oversampling is widely used at the data level, but they also point out that hybrid and feature selection techniques are not applied often enough to improve predictive model generalization. The advantages and disadvantages of the suggested approaches are examined, providing insightful information regarding them. [7] Predicting college students' academic performance in demanding courses is the focus of the paper, especially in undergraduate programs with high dropout rates. To identify students who are at risk of failing, it analyzes student data using Educational Data Mining (EDM) techniques and builds predictive models. The study assesses several machine learning algorithms, such as deep neural networks (DNN), choice trees, random forests, gradient boosting, logistic regression, assist vector classifiers, and K-nearest neighbors, using a public 4-year college dataset. The suggested DNN model predicts performance in data-driven publications with an astounding 89% accuracy rate. The paper explores the impact of grading systems on academic motivation, highlighting a gap in existing literature. It compares multi-interval grades to pass/fail systems with narrative evaluations, revealing that

traditional grading systems may have negative effects, such as increased anxiety and avoidance of challenging coursework. Narrative evaluations, however, are seen as a promising alternative, supporting students' psychological needs and fostering motivation through feedback, trust-building, and peer collaboration. [8] To obtain insights from conventional course report data, this study investigates the application of Educational Data Mining (EDM) techniques. In order to determine optimal performance, the study developed and evaluated five classification models using two cross-validation techniques and five algorithms. Tutors can improve academic performance by using the models, which are based on time segmentation and course performance attributes from course reports, as early indicators to identify students who may be at risk. By following the guidelines, educators and practitioners can improve their teaching strategies by using the data that is already available to forecast future trends and behaviors among their students. [9] The Random Forest algorithm, a type of machine learning, is used by the Student Career Suggestion System to assist students in selecting the best career path. It assesses technical proficiency, participation in sports, and academic achievement to offer tailored career advice. This automated method provides thorough evaluations, allowing recruiters and students to make educated decisions. With an emphasis on technical advancement and career prediction, the system gives students the ability to plan and make informed decisions about their future pursuits. This program supports academic and professional growth while raising awareness of career options. [10] A study by Garg, Chowdhury, and Sheikh in 2023 found that 74% of 18-year-olds in India drop out of school before 12th grade. Factors like caste, wealth, institution type, and regional differences significantly influence dropout rates. The study emphasizes the need for improved school infrastructure and education quality. [11] Dr. S.Y. Swadi's review of research on school dropouts and school environment highlights the limited impact of initiatives aimed at reducing high school dropout



rates. The paper emphasizes the need for more empirical research to understand how social workers can collaborate with stakeholders to improve special education and involve them in decision-making regarding special needs. [12] Cem Kirazoğlu's study explored the reasons behind secondary school dropouts, focusing on the perspectives of school administrators and counselors. The study involved semi-structured interviews with administrators and counselors from 19 Istanbul schools, gathered through undergraduate students and the researcher. The aim was to understand the perspectives of these key school institutions on the issue and identify risk factors such as students, family, teachers, the educational system, and primary school applications. [13] The paper presents SDA-Vis, a visualization system designed to explain student dropout rates using various academic, social, and economic factors. It provides insights into feature-perturbed student versions, enabling decision-makers to interpret situations and implement corrective actions. SDA-Vis, developed under domain experts, uses linked views to identify variable alterations and synthesize non-dropout scenarios. Case studies from a Latin American university show its efficacy in identifying at-risk students and proposing targeted interventions, demonstrating its potential for improving educational outcomes. [15] Using ensemble learning techniques, the paper explores the prediction of student success in Massive Open Online Courses (MOOCs). Three methods (bagging, boosting, and stacking) are used to analyze data from 480 students, covering 17 features. Compared to existing methods, the stacking ensemble classifier achieves 88% accuracy and an AUC of 0.85. The field of educational data mining is advanced by this work, which may improve online learning outcomes and instructional tactics. [16]

3. Review Methodology

A systematic review is conducted using a research methodology that must be unbiased and ensure comprehensiveness to evaluate all existing studies

in the relevant subject area. The chosen approach highlights the rigorous and consistent process of the systematic review. Studies on student dropout rates and academic performance are pertinent across various disciplines. Previous research is categorized into two primary groupings to provide context for our methodology. Identifying research inquiries is paramount for a reviewer, thus, we have specifically tackled the following crucial questions in our review:

- What are the principal factors influencing the analysis of student academic performance and the visualization of dropout trends in educational environments?
- Which interventions have proven effective in enhancing student retention rates based on academic performance metrics?
- In what ways does visualizing dropout rates impact academic achievement concerning subject grades and scores?
- What are the common methods and instruments utilized for assessing student academic performance?
- How can educators employ subject grades and scores to pinpoint students at risk and prevent dropouts?

Our intention was to integrate the prevailing understanding in the sector and underscore any prospective research deficiencies and upcoming trajectories by meticulously investigating these areas. This approach ensures a comprehensive investigation and paves the way for further advancements in our understanding of Academic Performance analysis and dropout trends.

3.1. Data Collection Methods

The study aims to gather data on students' academic experiences at a public university and featuring national characteristics. The university dataset comprised 4266 anonymized student records, containing 12 academic variables, including course grades, carryover courses from the first year, and grades in the data structure course. Institutional records and databases are used to gather



comprehensive data, including academic performance, grades, and attendance records. The objective was to construct predictive models for academic achievement in the data structure course. Data preprocessing involved scaling, encoding features, discretization, cleaning, and handling imbalanced datasets. To address the imbalance, various resampling techniques like SMOTE, random over-sampling, ADASYN, and SMOTE-ENN were applied. Additionally, xAPI-edu datasets were utilized, sourced from an online learning management system, containing 17 features and 480 students. These datasets encompassed students' learning behavioral traits, academic backgrounds, and demographic information. For binary classification, the Class attribute was transformed into numerical data types, with low grades classified as 0 and medium to high grades as 1. The datasets were typically provided in tabular format, with the UFES dataset available in CSV and the INEP dataset in XLSX. Tools like Pandas in Python were employed to manipulate and analyze the tabular data, offering functions to convert data into Data Frames for analysis. Demographic information, such as age, gender, ethnicity, socioeconomic status, and place of residence, is collected to provide context and insights into students' backgrounds. This helps researchers explore how factors like socioeconomic status or geographic location might impact academic performance and dropout risk.

3.2. Data Analysis Techniques

The paper includes deep artificial neural networks, decision trees, random forest, gradient enhancement, logistic regression, support vector machine, and K-nearest neighbor in oversampling methods such as SVM-SMOTE and Borderline-SMOTE which are used for future research is also done which was discussed. This increases the robustness and accuracy of the data analysis.[2] The study examined the connections between academic performance, smoking, faculty, age, and other factors using bivariable and multivariable data analysis techniques. The study employed various statistical techniques such as multivariable logistic regression, chi-square tests, SPSS version 25, and

structured questionnaires to investigate the association between variables and academic performance. These techniques gave insightful information for the goals of the study.[3] Group learning techniques such as stacking, boosting, and bagging were used in the study to determine student performance. These methods increase prediction accuracy by taking advantage of the combined power of multiple models. To ensure that all items are on the same scale for effective sampling, item measurement algorithms, standard scalar methods and support vector classifier (SVC) were used as supervised machine learning algorithms as classification algorithms used for data preprocessing phase. By using higher-order features to find optimal decision boundaries, SVC excels in solving complex classification tasks.[4] The systematic literature review applied machine learning techniques to predict and improve students' overall performance. A decision tree model was used to extract classification rules from student grades, and final GPA results were predicted. This approach identified key structures and relationships, which enabled the development of predictive models that could predict students' learning trajectories and support their educational journeys.[5] The paper uses a combination of differential, clustering, and convolutional neural network methods to analyze and predict student learning outcomes. These methods are widely used by researchers in the field to assess student performance in educational content searches. To better analyze learning outcomes, some researchers also use clustering and other data mining techniques such as k-means algorithm, entropy, unsupervised machine learning methods etc. This review highlights the properties of the analytical approach that combines cluster analysis with convolutional neural networks.[6] Analytical methods included post hoc Tukey test, backward elimination, and ANOVA analysis. To ensure that the data were consistent with the hypotheses, correlations were analyzed, using ANOVA to assess how universities' levels of motivational autonomy differed differently. Traditionally of post-cleaning methods were used to



improve the model. These pathways were examined more closely for differences in university motivation.[8] The study used five Scikit-learning classification algorithms, including Decision Tree CART, Extra Trees Classifier, Random Forest Classifier, Logistic Regression, and C-Support Vector Classification, as well as two methods for classifying the data sets three seasons including six predictive qualities. Conventional skit-learning classification techniques were used for advanced analysis, such as decision tree, logistic regression, ensemble classifiers, and support vector machines.[9] The paper uses a random forest algorithm to predict career paths based on students' academic achievement and technical skills. It discusses content-based approaches to policy recommendations, individual factors such as age and gender, and pedagogical data mining strategies. The study also discusses how to grade final prediction models using techniques such as optimal equal width binning and synthetic minority oversampling. These approaches contribute to a broader understanding of performance prediction and academic achievement models in educational settings.[10] The data analysis techniques employed in the study involved several steps. Semi-structured interviews were conducted to collect data. These interviews provided a flexible framework for obtaining detailed feedback from participants. A non-probability sampling method was used to select participants for the study. This approach involved selecting individuals based on specific criteria rather than randomization, allowing for targeted data collection. After the data were gathered, they were divided into two source groups—school administrators and counselors. Finally, the collected data were analyzed to examine the reasons for students dropping out. Through this analysis, common themes such as academic failure, truancy, and discipline problems were identified, providing valuable insight into the factors contributing to high dropout rates.[13] The study used neural networks to predict students dropping out and found high accuracy. Data cleaning, formatting, selection, and extraction were performed for comparability. The

data set included applicant and course progress data. Data sets were developed for training, and new data sets were created to address data security concerns. This step enhanced the reliability and validity of the predictive modeling approach.[14] The paper uses a counterfactual-based approach to analyze student dropouts and proposes strategies to prevent attrition. It comes with SDA-Vis, a visual analytics program that helps students understand symbols and information that contradict reality. The paper also highlights machine learning models such as random forests, neural networks, support vector machines, deep learning, and logistic regression that were used in previous research to analyze the patterns of student dropouts and create independent ones receives input reports.[15] The study used Python, Pandas, and Bokeh libraries for data processing and visualization. Pandas converted tabulated data into DataFrames using functions like `pandas.read_excel` and `pandas.read_csv`. The data were then used to estimate dropout rates and identify trends in higher education institutions (HEIs). This chart identified trends and areas for intervention. The inclusion of these libraries improved data processing and visualization and enhanced the analytical power of the study.[16]

4. Result and Finding

Neural Networks The deep neural network (DNN) model outperformed other models such as decision trees, logistic regression, support vector classifiers, and K-nearest neighbors, predicting student success in data structures learning with an accuracy of 89%. The use of the SMOTE method as an oversampling method improved the predictive ability of the DNN model, resulting in an accuracy, F1 score, and sensitivity of 89% for predicting students' academic achievement.[2] The study shows that 66% of students performed well academically, with significant correlations with parameters such as age, faculty, smoking habits etc. Students aged between 20 and 24 from medical/health faculties performed well. Students who did not smoke had three times as many good grades. Family psychosocial variables such as weight loss and family background also influence academic achievement. Behavioral



characteristics such as physical activity and smoking also influenced academic success.[3] The stacking ensemble classifier had an 88% accuracy rate with an AUC of 0.85, outperforming base classifiers and other ensemble classifiers in prediction accuracy, according to the study. An increased AUC rate of 0.85 was obtained by combining the stacking approach with the Extremely Randomized Trees ensemble learning technique. Standard scaling, nominal feature conversion, and binary categorization were among the preprocessing techniques used. The framework of genetic algorithms played a pivotal role in the classifier's optimization.[4] Systematic reviews revealed machine learning strategies used to predict students at risk of dropping out, thereby increasing overall student performance. Among these methods, decision tree structures played an important role in deriving classification rules from students' grades in previous courses to determine their final GPA. By analyzing the patterns in these grades, the decision trees provided valuable insights into factors affecting student learning outcomes, facilitating active interventions to support struggling students and reinforced overall educational success.[5] The paper uses clustering, differentiation, and convolutional neural network techniques to analyze and predict student learning outcomes. It covers data preprocessing, clustering, discrimination, analysis, and training data for convolutional neural networks. The model achieves good predictive accuracy, allowing schools to conduct appropriate assessments. Data analytics techniques transform unstructured educational data into valuable information for predicting student performance and evaluating academic achievement.[6] The study highlights the importance of addressing distributional imbalances in predicting student outcomes. Various approaches have been proposed to address this issue, such as sampling methods, feature selection, cost-effectiveness learning, and mixed methods. The findings provide insights into current approaches to addressing distributional imbalances in predicting student outcomes and pave the way for future research and practical

applications. [7] The study findings highlight the significant differences in the effects of academic grades and narrative assessments on student motivation. Although grades were associated with higher levels of anxiety and a desire to avoid difficult courses, narrative research appeared to support students' basic psychological needs and motivation through action feedback. Universities using narrative assessment significantly increased intrinsic autonomous motivation compared to students using traditional multiinterval grading. Participants in universities using narrative and hybrid assessment methods the role exhibited particularly strong academic motivation, as seen in the increased levels of autonomous motivation and decreased controlled motivation.[8] To characterize student achievement, the study developed and tested five classification models using traditional course report data. These models improved academic achievement by better identifying at-risk children. After testing several classification algorithms, the best-performing logistic regression classifier was found to predict student academic success. This study extends the ability to predict student academic success using data mining techniques from traditional report data on.[9] The proposed student career prediction system assesses student achievement through grading, technical skills, athletic activities, and extracurricular activities. It employs the Random Forest algorithm to make job recommendations, assisting students in identifying their strengths and limitations for future career growth. Technology automates student information administration in educational institutions, reducing manual workload for teachers and school administrators. The system's output predicts job routes based on academic performance, sports participation, and talent assessments.[10] The neural network improved the prediction of dropout after the first semester by 95%, outperforming decision trees and logit models. However, as the study progressed, the prediction accuracy of dropout improved to 95% after the third month. The prediction accuracy of neural networks was high, and the advantages of translational and descriptive



aspects over traditional methods were retained.[14] The paper used ensemble learning methods, such as stacking, boosting, and bagging, to predict student performance. These methods use the combined strength of multiple models to improve prediction accuracy. In addition, feature scaling algorithms and the standard scaler method were used for data preprocessing, ensuring that all features were on the same scale for effective modeling. As part of the classification process, a supervised machine learning algorithm called Support Vector Classifier (SVC) was used. SVC excels at handling complex classification tasks by determining optimal decision

boundaries in high-dimensional feature spaces.[15] The purpose of the study was to create a computational tool for data visualization that could effectively profile students and explain the underlying causes of dropout in Higher Education Institutions (HEIs). The study's goal was to create a dashboard that would allow for continuous data updates over time and provide a platform for exploratory analysis of relevant factors associated with university dropout is show in Table 1.[16]

Table 1 Comparison of Previous Works in Predicting Students' Performance Using Academic Grades.

Ref No.	Algorithm	Dataset	Grades	Attributes	Accuracy
[1]	Path Analysis	258 UG students	Yes	Gender, age, nationality, student status, Cumulative Grade Point Average (CGPA), Achievement Goal Orientation (AGO)	85% -90%
[2]	Deep Neural Network (DNN), Decision Tree, Logistic Regression, Support Vector Classifier (SVC), K-Nearest Neighbor (KNN)	14 years (2006-2020) of data of 500 records and 16 features	Yes	subjects, marks, carryover(backlog), target	71.09% - 93%
[3]	Bivariable and Multivariable Logistic Regression Analysis	615 Ethiopian Students	Yes	Age, gender, CGPA, study hours, attendance	85% - 95%
[4]	Genetic Algorithm	480 students and 17 features	Yes	Nationality, Educational level, Grade level, Section ID, Semester	84% - 87%
[5]	Decision Tree (DT), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), K-nearest neighbors (KNN)	6 Distinct Student Performance Dataset	Yes	GPA, degrees, marks, Course level, Enrolment data	80%



[6]	Random Forests (RF), Nearest Neighbour (NN), Support Vector Machines (SVM), Logistic Regression (LR), Naïve Bayes (NB), k-nearest neighbour (KNN)	1854 Turkish Language students 2019–2020	Yes	Midterm exam grades, Final exam grades, Faculty, Department	70% –75%
[7]	SMOTE (Synthetic Minority Oversampling Technique) ENN (Extended Nearest Neighbor) GWO (Grey Wolf Optimization) RF (Random Forest)	Several Student-grade-related	Yes	Student ID, C/GPA, dropout status	80%
[8]	ANOVA Backward Elimination	392 UG student	Yes	Age, Year of study, Field of study, High school GPA, Autonomous motivation to attend university (AMS), Engagement in specific learning techniques (SRQ-L)	95%
[9]	Decision Tree CART Extra Trees Classifier Random Forest Classifier Logistic Regression Classifier C-Support Vector Classification (C-SVC)	1077 student records and 43 features	Yes	Gender, year, GPA, credits, attendance, marks, course status	85% - 90%
[10]	Multilayer Perceptron (MLP) Neural Network	12 Datasets of 5,168 students (2002 – 2016) includes 327,144 observations.	Yes	Student ID, Success (indicating dropout, success, or enrollment), and End at (date)	91% - 98%
[11]	Descriptive, Predictive, Diagnostic, Prescriptive analysis	UFES and INEP (2010) dataset	Yes	Student ID, course enrollment, and subjects taken.	90%



5. Challenges and Future Scope

The education sector faces challenges and opportunities. A key challenge is to understand the impact of factors other than cognitive variables on academic achievement. This will help teachers develop strategies to better support students. Also, academic achievement can be enhanced by developing grit and self-efficacy. Expanding research to student populations can help develop more inclusive educational strategies. Addressing these challenges and embracing future opportunities can improve educational practice and ensure equitable academic success for all students. [1] Addressing an unbalanced dataset presented a significant challenge for the research, negatively affecting the models' performance. One possible way to tackle this problem in the future would be to add data from more semesters to the dataset. This augmentation provides a more thorough representation of the underlying patterns, with the goal of improving the accuracy and reliability of the model. Furthermore, in contrast to the algorithms used in the study, future research could examine the efficacy of alternative oversampling strategies like SVM-SMOTE and Borderline-SMOTE. Through the assessment of these methods' efficacy, scientists can obtain further knowledge to enhance model training and effectively tackle problems related to class imbalance. These initiatives have the potential to improve the methodology of the field and increase the resilience of prediction models in various scenarios. [2] Social desirability bias, uncontrolled confounders, and under- and overreporting were among the difficulties the study encountered. Direct causal inferences were limited due to the cross-sectional methodology. Longitudinal studies, intervention initiatives, and investigations into variables such as study habits, mental health, and socioeconomic background should all be part of future research. These channels may offer perceptions into methods for enhancing results and guide focused activities to encourage academic achievement. This would improve field knowledge and guide focused initiatives. [3] The research offers several exciting new avenues for

analysis and application. First, the discovery of sophisticated evolutionary patterns or hybrid methods can greatly improve the optimization process, leading to effective and efficient solutions. Second, through multiple resources and data sources upon incorporation, group learning models can be more predictive to increase their accuracy and variability. Finally, beyond the scope of this work, examining the application of group learning to other academic tasks or real-world settings may provide insightful information and practical applications. [4] Universities face various challenges in assessing student achievement, providing high standards, tracking student progress effectively. E-learning systems face similar challenges, such as high attrition rate, lack of uniform assessment standards there, the problem of predicting the unique needs of each student and future students. It may be possible to go beyond traditional assessment methods by using machine learning algorithms to establish early signs that objects the prophecy of the destruction. Ultimately, the use of predictive analytics can create an inclusive and productive learning environment by engaging at-risk students and adapting interventions to promote their success on the snow. These efforts promise to improve student performance and enrich the educational experience, as well as transform the shift toward data-driven approaches to education. [5] The main obstacles facing academic data analysis include the difficulty of estimating appropriate group sizes and the potential for bias introduced by the initial random selection of groups. Using a variety of methodologies and sophisticated research tools, scholars can overcome traditional boundaries, enhance data-driven decision-making and uncover new approaches and applications. These projects have the potential to spur creative and transformative growth outside of various disciplines other than academia. [6] Problems associated with predicting student grades include unbalanced characteristics of data sets, which prevent more accurate predictive models, and inappropriate use of mixed choices of factors and methods, which limits the generalizability of models. In addition, notable



attention has been paid to the use of hybrid methods for multi-class classification to improve the accuracy of the prediction model in predicting student grades. Researchers can make progress in these areas to develop more robust and useful predictive models, which will help improve decision-making and educational outcomes. [7] Predictive models for education present a number of issues, including limitations associated with the use of data from a single subject and the need to test models across different curriculums and learning styles. Harnessing the wealth of information generated by virtual learning environments may enable scholars to increase the accuracy and scalability of predictive models across learning environments. This method of inquiry has the potential to inform educational research has made significant progress and guided interventions for individualized learning. [9] The difficulties that student career prediction systems confront are overcoming the constraints that come with manual methods and overcoming the complexity of evaluating various aspects of student performance. In the future, there will be a chance to improve career advice by adding more variables, like personal traits, and using sophisticated data analysis methods to increase precision. In addition, the future scope calls for automating the management of student data, which will relieve instructors and administrative staff of some of their manual labor and promote more streamlined and effective educational systems. Educational institutions can enhance their entire student experience and empower learners with individualized career insights by adopting these developments and optimizing their student guidance processes. [10] Despite initiatives like national policies on education, Right to Education Act and Sarva Shiksha Abhiyan, the education sector still faces challenges. Minority groups such as SC, ST, Muslims and women have unequal access, and income inequality leads to discrimination in education. The New Education Policy 2020 aims to provide equality, affordability, quality and accountability will be at the forefront of education

and these issues will be addressed to promote social and economic development. [11] High school dropout rates are a major challenge, requiring government and NGOs to work together to develop targeted strategies. Future initiatives include free education programs, funding, combating training culture, improving school systems, ensuring equal access to education. By addressing these challenges, stakeholders in the project can create an inclusive educational environment, reduce dropout rates and promote academic success for all students. [12] The study encountered several significant obstacles, such as possible biases originating from data collection personnel and a scarcity of thorough research on the subject. To obtain a more comprehensive understanding of the factors influencing school dropout rates, complementary research involving educators, young people, and families has a bright future ahead of it. In addition, the study intends to evaluate the educational system's operations critically and provide reform recommendations based on scientific knowledge. Through tackling these issues and broadening the scope of the research to incorporate a range of stakeholders, the study aims to support well-informed policymaking and promote constructive modifications in the educational environment. [13] Compared to traditional methods such as decision trees and logistic regression, the use of neural networks to predict student dropout presents interpretive issues. Subsequent research will focus on improving neural network interpretation capabilities without compromising their high prediction accuracy. In addition, more research is possible on the use of constrained input variables in predictive modeling to address data security issues. Further research could examine the impact of including data from multiple sessions in predicting dropout rates. Researchers can address these issues and explore these possible future directions to improve the effectiveness and usefulness of predictive models in addressing dropout rates while maintaining confidentiality and interpretable data. [14] Among the many obstacles the study must overcome are the limitations of the current system



and the need for automated prediction of student performance. Later projects may include reviewing high school transcripts to develop appropriate career counseling for students. Furthermore, improving designs to evaluate counterfactual effects on specific student populations—for example, by integrating gender-based assessments into curriculum—has the potential to inform development and provide support systems. Subsequent research will focus on early trends affecting university dropout rates, taking into account risk factors related to family dynamics and demographics. By addressing these barriers and exploring these potential directions in the future, scientists can improve predictive models and intervention strategies to drive student achievement and retention high. [15] Challenges stem from the unprocessed data that educational institutions collect as they attempt to understand and reduce student dropouts. On the other hand, these computer tools, which can be used by other academic institutions to estimate dropout rates and analyze student data, present opportunities for the future. Strong assessment plays an important role in an educational setting, as evidenced by the fact that improving data analysis and quality can significantly improve different aspects of the program Stakeholders can be deterred species currently available and seize future opportunities to develop well-defined methods. [16]

6. Observations

The study examines school dropouts in India using a retrospective approach and the Cox proportional hazard model. It found that 74% of individuals aged 18 and above discontinue their education before completing the 12th standard. Factors such as caste division, wealth quintile, institution type, and regional disparities influence dropout rates. Factors contributing to dropout risk include lack of interest in education, distance from school, academic struggles, and financial constraints. Disinterest and academic challenges are linked to educational quality, while financial constraints and distance from schooling are associated with public-school delivery deficiencies. Early marriages are

particularly impactful for the female population, contributing to school attrition. The study emphasizes the need for improved school infrastructure and quality, affordable, and accessible education to improve enrollment [1]. The paper highlights the current state of research on school dropout and the school environment, highlighting the limited impact of interventions aimed at curbing high school dropout rates. The issue is a serious concern for any country, encompassing financial constraints, practical challenges, and dissatisfaction with the social system and examination results. The paper focuses on the foundational aspects of being a school social worker, emphasizing roles such as eliminating dropouts, building relationships, conducting assessments, collaborating with multidisciplinary teams, and assisting children and adolescents in overcoming academic obstacles. It calls for the inclusion of social workers in decision-making processes related to special education needs and for more empirical research to explore how social workers can effectively collaborate with stakeholders to achieve positive outcomes for students with special needs [2]. The study suggests enhancing research on secondary school dropouts by involving diverse stakeholders, including students, parents, and community members. It recommends longitudinal studies to capture dynamic insights and explore contextual factors like socio-economic conditions and cultural influences. The study also suggests leveraging technology for data collection and integrating qualitative and quantitative metrics. Comparative analysis across different school types is suggested as a strategic approach. Targeted intervention strategies and collaboration with educational policymakers are also recommended. Ethnographic studies and observational approaches can uncover implicit factors contributing to dropout, while long-term impact assessment ensures thorough evaluation of intervention effectiveness. Cross-cultural comparative studies provide a global perspective on secondary school dropout issues [3]. This research highlights the importance of education in economic development and community problem-solving. A



study in Southern Ethiopia found that academic performance among university students is influenced by factors such as smoking, age, and study field. Students aged 20-24 and in medical/health faculties showed better academic performance. Non-smokers were three times more likely to achieve higher grades. The study suggests that reducing or discontinuing smoking is crucial for academic success, especially among targeted groups. It also suggests inviting older students to share their experiences and reasoning methods to enhance the academic environment. The findings suggest practical interventions to improve educational outcomes in Southern Ethiopia [4]. The study demonstrates the potential of Educational Data Mining (EDM) techniques in predicting future student trends and behaviors. It uses five algorithms and two cross-validation methods to develop and evaluate classification models, focusing on time segmentation and course performance attributes from historical reports. These models can identify students at risk and provide targeted interventions for improved academic performance. The study encourages practitioners to revive old data to gain valuable insights for upcoming academic years. Overall, EDM techniques can empower tutors and researchers in fostering student success in educational settings [5].

Conclusion

The purpose of this review was to view the trends in composition studies within the past years and the problems faced by students for academic development and resource Quality. This evaluation paper examines the challenges and trends in composition research, specializing in pupil instructional development and aid fine. It highlights the complexity of the education device, and the demanding situations students face in identifying their hobbies and reaching educational success. One key problem is pupil dropout, prompted via monetary, realistic, and social elements. Understanding dropout rates is important for policymaking and student guide. The evaluation emphasizes the importance of students' overall performance in shaping their destiny trajectories. It

additionally discusses the significance of addressing imbalanced datasets in pupil grade prediction and the capacity of machine mastering algorithms like deep synthetic neural networks, selection trees, and logistic regression. The overview also highlights the efficacy of neural networks in forecasting pupil fulfillment and dropout quotes, emphasizing the price of knowledge complicated academic records. Educational records mining strategies, along with deep learning and survival evaluation, offer valuable insights into pupil overall performance and dropout styles. Visualization gear like SDA-Vis and Performance Vis are essential for studying student overall performance and exploring dropout risks. Visual analytic gear like VIS4ML and ViCE can interpret gadget gaining knowledge of fashions and generate counterfactual reasons for model choices. Combining predictive modeling strategies with visualization equipment can decorate information of scholar academic overall performance and guide powerful intervention techniques. It is clear from the research reviewed that the government cannot get the exact ideology about the weaker point of the education system of the country because of that students face problems in identifying the area of interest for future outcomes. Students Discontinuing schooling due to financial, practical, and social reasons, as well as disappointment in examination results, is what is commonly referred to as student dropout. The review helps to get the exact ratio of dropout students so that the government can make policies for student support and guidance. Students' performance is the key element for selecting area of interest for future aspects.

Acknowledgement

It is our pleasure to acknowledge our sincere thanks with a deep sense of gratitude towards our project guide "Dr. Nitesh A. Ghodichor", Assistant Professor, Computer Technology Department for his continuous knowledge and support in conducting this dissertation work. He has wholeheartedly helped us in this endeavor at all stages of this work. We are thankful to Dr. S. A. Dhale, Principal, Priyadarshini College of Engineering, Nagpur, for providing the facilities at



the institute. We thank Dr. N.M. Thakare, Professor and Head of Computer Technology Department of Priyadarshini College of Engineering, Nagpur. We here with express our immense thanks to “Dr. Nitesh A. Ghodichor”, Assistant Professor, Assistance Professor, Computer Technology Department, Priyadarshini College of Engineering, Nagpur for giving us suggestions and co-ordination with us from time to time as the project in charge. We also take the opportunity to thank all who have directly or indirectly extended help and encouragement in executing this project.

Reference

- [1]. Amal Alhadabi & Aryn C. Karpinski (2020) Grit, self-efficacy, achievement orientation goals, and academic performance in University students, *International Journal of Adolescence and Youth*, 25:1, 519-535, DOI: 10.1080/02673843.2019.1679202
- [2]. Nabil, Aya & Seyam, Mohammed & Abou-Elfetouh, Ahmed. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access*. PP. 1- 1. 10.1109/ACCESS.2021.3119596.
- [3]. Tadese M, Yeshaneh A, Mulu GB. Determinants of good academic performance among university students in Ethiopia: a cross-sectional study. *BMC Med Educ*. 2022 May 23;22(1):395. doi: 10.1186/s12909-022-03461-0. PMID: 35606767; PMCID: PMC9125903.
- [4]. Chin-Wei Teoh, Sin-Ban Ho, Khairi Shazwan Dollmat, and Ian Chai. 2022. An Evolutionary Algorithm-Based Optimization Ensemble Learning Model for Predicting Academic Performance. In *Proceedings of the 2022 11th International Conference on Software and Computer Applications (ICSCA '22)*. Association for Computing Machinery, New York, NY, USA, 102–107. <https://doi.org/10.1145/3524304.3524320>.
- [5]. Albreiki, B., Zaki, N., & Alashwal, H. (2021, September 16). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*, 11(9), 552. <https://doi.org/10.3390/educsci11090552>
- [6]. G. Feng, M. Fan and Y. Chen, "Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining," in *IEEE Access*, vol. 10, pp. 19558-19571, 2022, doi: 10.1109/ACCESS.2022.3151652.
- [7]. Abdul Bujang, S. D., Selamat, A., Krejcar, O., Mohamed, F., Cheng, L. K., Chiu, P. C., & Fujita, H. (2023). Imbalanced Classification Methods for Student Grade Prediction: A Systematic Literature Review. *IEEE Access*, 11, 1970–1989. <https://doi.org/10.1109/access.2022.3225404>
- [8]. Chamberlin, Kelsey & Yasué, Mai & Chiang, I-Chant. (2018). The impact of grades on student motivation. *Active Learning in Higher Education*. 24. 10.1177/1469787418819728.
- [9]. Bucos, Marian & Drăgulescu, B.. (2018). Predicting student success using data generated in traditional educational environments. *TEM Journal*. 7. 617-625. 10.18421/TEM73-19.
- [10]. Sadasivam, R., Paramasivam, S., Prakash raj, N., & Saravanan, M. (2022). Students career prediction. *International Journal of Health Sciences*, 6(S5), 1357–1365. <https://doi.org/10.53730/ijhs.v6nS5.8883>
- [11]. Garg, Mausam & Chowdhury, Poulomi & Sheikh, Ilias. (2023). Determinants of school dropouts in India: a study through survival analysis approach. *Journal of Social and Economic Development*. 10.1007/s40847-023-00249-w.
- [12]. Mr. Basavaraj Biradar, Dr. S.Y. SWADI, Reviews on school dropouts and research gap. *International Journal of Research and Analytical Reviews (IJRAR)*.
- [13]. Kirazoğlu, Cem. (2009). The investigation of school-dropout at the secondary level of



- formal education: the stated reasons by the school administrators and school counselors: a preliminary study. *Procedia - Social and Behavioral Sciences*. 1. 905-914. 10.1016/j.sbspro.2009.01.161.
- [14]. Buchhorn, J., & Wigger, B. U. (2021). Predicting Student Dropout: A Replication Study Based on Neural Networks. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3929194>
- [15]. Garcia-Zanabria, G.; Gutierrez-Pachas, D.A.; Camara-Chavez, G.; Poco, J.; Gomez-Nieto, E. SDA-Vis: A Visualization System for Student Dropout Analysis Based on Counterfactual Exploration. *Appl. Sci.* 2022, 12, 5785. <https://doi.org/10.3390/app12125785>
- [16]. Ladeira, Pedro & de Lima, Leandro & Krohling, Renato. (2022). A visualization tool for data analysis on higher education dropout: a case study at UFES.