



Predictive Modelling And Data Analysis Of Student Performance In Higher Education Institution Of Goa

Sameer Patil

Research Scholar

Shri J.J.T University Rajasthan ,INDIA

Dr. Murlidhar Dhanawade

Professor & Director

Sterling Institute of Management Studies Navi Mumbai,INDIA

Abstract :- This study focuses on "Predictive Modeling and Data Analysis of Student Performance in Higher Education Institutions of Goa," with the goal of identifying trends, important performance indicators, and providing actionable insights for improving academic outcomes. The study employs advanced data analysis tools and predictive modeling to assess the relationship between demographic, socioeconomic, academic, and institutional variables and student achievement.

The goals include examining patterns in student achievement, identifying significant factors impacting performance, and evaluating the effectiveness of institutional support and extracurricular activities. Predictive models will be built to forecast academic success, and the impact of attendance, family history, and peer influence will be thoroughly investigated.

The hypotheses aim to investigate significant correlations between academic achievement and variables such as socioeconomic position, peer influence, and institutional assistance.

In addition, the study assesses the usefulness of predictive models in properly projecting outcomes and the effects of family background and extracurricular engagement on academic achievement.

The study uses a comprehensive methodology that includes data gathering via structured questionnaires, statistical analysis, and machine learning techniques. By bridging the gap between raw data and strategic decision making, the study aims to equip Goan educators and policymakers to execute data-driven interventions that improve student performance and institutional effectiveness.

This paper contributes to the emerging field of educational data analytics by providing a detailed knowledge of the interactions between many factors and academic accomplishment, paving the path for more informed and equitable educational methods.

Keywords:- Student performance, higher education, predictive modelling, data analysis, Goa, academic success, institutional support, socio-economic factors, machine learning, educational analytics.

Introduction :- Higher education plays a critical role in developing individuals' intellectual, professional, and economic paths. Higher education institutions in Goa serve a varied student population with a variety of socioeconomic backgrounds, academic objectives, and career goals. Understanding the elements that influence student success in these institutions is crucial for promoting academic excellence and guaranteeing equal access to opportunities.

The advent of data-driven decision-making in education has transformed the way academic institutions view student achievement. Using data analysis and predictive modeling, it is now possible to discover major predictors of academic success, reveal hidden trends, and devise targeted interventions to help students succeed. This strategy is consistent with the overarching goals of improving institutional effectiveness and optimizing resource allocation while also meeting individual student needs.

The introduction of data-driven decision-making in education has altered how academic institutions perceive student accomplishment. Using data analysis and predictive modeling, it is now possible to identify important drivers of academic achievement, uncover hidden trends, and develop focused interventions to assist students succeed. This technique is in line with the overarching goals of increasing institutional effectiveness and optimizing resource allocation while also fulfilling individual student requirements.

The study also aims to assess the influence of emerging technologies such as artificial intelligence (AI) in improving student learning and performance. By investigating the relationship between numerous factors and academic accomplishment, this project hopes to bridge the gap between theoretical knowledge and practical implementations, ultimately helping to better higher education outcomes in Goa.

This study intends to pave the path for a future in which educational decisions are guided by actionable insights, allowing institutions to empower students to reach their academic and professional potential.

Review of Literature: -

According to Pankajini Sahu et al. (2022) in their study "Predictive Analysis of Students Performance Evaluation in Higher Education: A Machine Learning Approach," a comprehensive and relevant performance review procedure should be initiated at the start of the academic year. The number of colleges has grown recently, underscoring the importance of better approach performance in international competitiveness. Institutions might use performance evaluation to develop future projects. Every lecturer is expected to set annual goals for each area.[1] In-depth performance evaluations give instructors direction and constructive feedback to help them develop. Using a range of Machine Learning classifiers and ensemble approaches, we have carefully analyzed, assessed, and predicted the impacts of online education systems in this article. This study's main goal is to clarify the importance of higher education in assessing students' performance.

Education is one of the many fields that have been profoundly impacted by the development of predictive models.

Predictive modeling has been essential in improving student results in a number of educational institutions in recent years, according to Francis Makombe et al. (2020) in their paper "A Predictive Model for the Determination of Academic Performance in Private Higher Education Institutions." Their study presents a computational predictive model that uses artificial neural networks to forecast whether a student will pass or fail. The suggested model stands out from the rest of the literature since it assesses the effectiveness of five other algorithms in addition to applying neural network-based forecasting techniques.

In terms of forecasting academic performance, experimental data show that Artificial Neural Networks perform better than eXtreme Gradient Boosting (XGBoost), Linear Regression, Support Vector Machine (SVM), Naïve Bayes, and Random Forest.

In their paper titled "Predicting academic success in higher education: literature review and best practices," Alyahyan, E., et al. (2020) claimed that Student achievement is a significant factor in educational institutions and is often a key indicator of the general performance of the students. Early identification of at-risk pupils and rapid implementation of intervention strategies can significantly enhance their academic performance. In recent years, machine learning techniques have been widely used in educational predictive analysis. Although

there are numerous success examples in the literature, these approaches are primarily used by educators who possess expertise in computer science, or more specifically, artificial intelligence. Effective use of data mining techniques requires making several important decisions, such as defining student achievement, selecting relevant student attributes, and selecting the most appropriate machine learning approach for the given circumstance.

By providing a methodical, structured guidance for teachers interested in using data mining techniques to forecast student progress, this study seeks to close this gap. It compiles the most recent developments into a methodical framework by thoroughly reviewing the body of existing literature and outlining the important choices and constraints involved. This study aims to increase educators' access to data mining tools and help them realize their full potential in the field of education by offering concise explanations and rationales for these decisions.

In their study "Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques," Raza Hasan et al. (2020) claimed that innovation and technology enable HEI to employ a variety of learning systems, including video learning. The efficiency of this type of learning can be determined by examining the digital footprints created by these online interactions. Flipped teaching combined with video-based learning can help students perform better academically. 772 students enrolled in e-commerce and e-commerce technology courses at a HEI served as examples for this study. Using data mining and video learning analytics, the study sought to forecast students' total performance at the end of the semester. Eight distinct classification algorithms were used to analyze data from mobile applications, learning management systems, and student information systems. Additionally, preprocessing and data transformation methods were used to minimize the features. Moreover, genetic search and principle component analysis were carried out to further reduce the features. Additionally, the CN2 Rule Inducer and multivariate projection can be used to assist faculty in interpreting the rules to gain insights into student interactions. The results showed that Random Forest accurately predicted successful students at the end of the class with an accuracy of 88.3% with an equal width and information gain ratio

Problem Statement :-

In recent years, predictive modeling and data analysis have emerged as revolutionary tools in higher education, allowing institutions to improve student performance, lower dropout rates, and optimize academic outcomes. While global and national studies have looked into numerous techniques and models, there is still a considerable vacuum in understanding how these approaches might be implemented in unique regional contexts, such as Goa.

Higher education institutions in Goa have distinct obstacles, such as various student populations, varying levels of academic aptitude, and socioeconomic issues influencing performance. Despite the increased availability of educational data, institutions frequently lack specific frameworks for assessing it in order to forecast academic outcomes and plan focused interventions. Furthermore, existing research focuses on wider, generalized datasets, which limits the findings' relevance to Goa's unique educational context.

This study aims to close these gaps by using predictive modeling and data analysis approaches to assess factors influencing student performance in Goa's higher education institutions. By identifying significant predictors, the project hopes to deliver actionable insights that can help inform decision-making, improve academic support systems, and contribute to the establishment of a data-driven educational ecosystem in the area.

Objective:-

- 1) To assess the relationship between students' desired careers and their perceptions of their academic performance.
- 2) To examine how peer pressure and family expectations affect students' job decisions.
- 3) To evaluate how students' career decisions are influenced by career counselling and coaching.
- 4) To investigate how students' career goals are affected by extracurricular activities.
- 5) To investigate the need for further assistance (such as internships or counseling) in order to make well-informed career decisions.

Hypothesis :-

H₀: Students' academic achievement and their desired careers do not significantly correlate.

H₁: Students' academic achievement and their desired careers are significantly correlated.

H₀: Peer pressure and family expectations have little influence on students' job decisions.

H₁: Peer pressure and family expectations have a big impact on students' job decisions.

H₀: Students' professional decision-making is not substantially impacted by career counseling and guidance.

H₁: Students' professional decision-making is greatly influenced by career counseling and guidance.

H₀: Students' professional goals are not substantially impacted by their involvement in extracurricular activities.

H₁: Students' professional goals are greatly impacted by their involvement in extracurricular activities.

H₀: Students' professional decision-making process is not substantially impacted by extra help (such as internships or career counseling).

H₁: Students' professional decision-making process is greatly impacted by additional support (such as internships and career counseling).

Research Methodology:- To evaluate the study's goals and test the hypotheses effectively, a methodical research technique is required. This comprehensive research strategy is intended to address the specified objectives and theories.

Sample Selection: -

Population: The population of Goa is made up of students enrolled in higher education.

Sample Size: The study's sample will consist of 125 students. To ensure that the various streams (including Bsc, Bcom, BA, BCA, and BBA) are fairly represented, stratified random sampling will be used as the sample technique. Each stratum's students will be selected at random.

Data Collection Tool: A structured questionnaire will be used to collect data.

Topics like students' academic achievement, professional goals, the impact of extracurricular activities, the role of career counseling, the influence of peer pressure and family expectations, and the influence of career aspirations are all specifically addressed in the questionnaire.

Analyzing and Interpreting Data:

The following are the standard methods for analyzing and interpreting data based on the presumptions:

Testing of Hypothesis:

1. H₀: Students' academic achievement and their desired careers do not significantly correlate.

H₁: Students' academic achievement and their desired careers are significantly correlated.

Questions 1:-

a. How would you rate your overall academic performance in the past?

- b. What is your current GPA or percentage?
- c. How important is academic performance in achieving your career goals?

Normality Test: - A normality test is conducted to assess whether a dataset follows a normal distribution or exhibits a bell-shaped curve. This evaluation is crucial, as many statistical tests assume that the data is normally distributed, ensuring the validity and reliability of their results.

Shapiro-Wilk normality test

data: data\$How.important.is.academic.performance.in.achieving.your.career.goals.

W = 0.7984, p-value = 8.168e-12

is less than the standard significance level of 0.05.

Non-parametric test:

Kruskal-Wallis rank sum test

data: data\$How.important.is.academic.performance.in.achieving.your.career.goals by
data\$How.would.you.rate.your.overall.academic.performance.in.the.past.

Kruskal-Wallis chi-squared = 9.6425, df = 2, p-value = 0.008057

9.6425 is the Kruskal-Wallis chi-squared value.

This test statistic was computed using the data ranks across groups.

Degrees of freedom: df = 2.

(df = number of groups - 1) This indicates that there are three groups.

0.008057 is the p-value.

The likelihood of seeing the test statistic (or something more severe) under the null hypothesis is indicated by the p-value.

- Since p-value (0.008057) < 0.05, we reject the null hypothesis at a 5% significance level.
- This indicates that at least one of the groups is significantly different from the others in terms of the variable being tested.

2. Question 2:Ho: Peer pressure and family expectations have little influence on students' job decisions.

H_i: Peer pressure and family expectations have a big impact on students' job decisions.

Questions are :-

- What is the highest education level of your parents/guardians?
- How much does your family influence your academic choices?
- How much does peer influence affect your academic performance?

- Do you collaborate with peers for academic projects or assignments?

Normality test :-

Shapiro-Wilk normality test

data: data\$How.much.does.your.family.influence.your.academic.choices.

W = 0.81399, p-value = 2.788e-11

- Since **p-value (2.788e-11)** is much smaller than **0.05**, we reject the null hypothesis (**H₀**: Data is normally distributed).
- This means **your data is NOT normally distributed.**

Non Parametric Test

Wilcoxon rank sum test with continuity correction

data: data\$How.much.does.peer.influence.affect.your.academic.performance. and
data\$How.much.does.your.family.influence.your.academic.choices.

W = 7078, p-value = 0.1674

alternative hypothesis: true location shift is not equal to 0

- Since **p-value (0.1674)** is greater than **0.05**, we fail to reject the null hypothesis.

3. H₀: Students' professional decision-making is not substantially impacted by career counseling and guidance.

H₁: Students' professional decision-making is greatly influenced by career counseling and guidance

Question :

- Career Counselling
- Mentoring Program
- Access to study materials/resources
- How often do you participate in extracurricular activities?

Shapiro-Wilk normality test

data: data\$Mentorship.programs.numeric

W = 0.80131, p-value = 1.022e-11

So Non Parametric Test applicable here :- Kruskal-Wallis chi-squared = 34.742, df = 3, p-value = 1.381e-07

Since the p-value is much lower than the significance level (typically 0.05 or 0.01), reject the null hypothesis and conclude that there is a significant difference among the groups. This suggests that at least one group differs significantly from the others.

4.H₀: Students' professional goals are not substantially impacted by their involvement in extracurricular activities.

H₁: Students' professional goals are greatly impacted by their involvement in extracurricular activities.

Questions :-

- How often do you participate in extracurricular activities?
- Do you feel that extracurricular activities help in improving your academic performance?
- Mentorship programs
- Gender

Shapiro-Wilk normality test:-

data: data\$Do.you.feel.that.extracurricular.activities.help.in.improving.your.academic.performance.

W = 0.82266, p-value = 5.683e-11

Since the **p-value is significantly lower than the conventional threshold (0.05)**, we reject the **null hypothesis**, which assumes that the data follows a normal distribution. This indicates that the variable is not normally distributed.

Kruskal-Wallis rank sum test.

data: data\$Gender by data\$Do.you.feel.that.extracurricular.activities.help.in.improving.your.academic.performance.

Kruskal-Wallis chi-squared = 2.3877, df = 4, p-value = 0.6648

Since **p > 0.05**, we fail to reject the **null hypothesis**, which states that there is no significant difference in how different genders perceive the impact of extracurricular activities on academic performance.

5.H₀: Students' professional decision-making process is not substantially impacted by extra help (such as internships or career counseling).

H₁: Students' professional decision-making process is greatly impacted by additional support (such as internships and career counseling).

Questions :-

- Career counselling:
- How often do you participate in extracurricular activities?
- Mentorship
- Do you feel that extracurricular activities help in improving your academic performance?
- Does your institution provide support for mental health and well-being?

Normality Test :-

Shapiro-Wilk normality test

data: data\$Career.counseling.numeric

W = 0.75445, p-value = 3.53e-13

Since the Shapiro-Wilk normality test gave a $W = 0.75445$ with a p-value of $3.53e-13$, this indicates that the variable Career.counseling.numeric is not normally distributed. When normality assumptions are violated, you should consider non-parametric tests instead of parametric tests.

data: data\$Career.counseling by
data\$Do.you.feel.that.extracurricular.activities.help.in.improving.your.academic.performance.

Kruskal-Wallis chi-squared = 21.971, df = 4, p-value = 0.0002031

Since p-value = 0.0002031 is less than 0.05, we reject the null hypothesis.

Result : - The following findings were derived from the above-mentioned hypotheses:

- There is a strong association between students' academic success and their preferred careers, supporting the alternative theory.
- The null hypothesis, which contends that parental expectations and peer pressure have no bearing on students' career choices, is upheld.
- Support for the alternative hypothesis demonstrates how important career counseling and help are to students' decision-making over their careers.
- The null hypothesis is upheld, suggesting that students' participation in extracurricular activities has no discernible effect on their aspirations for their careers.
- By showing that extra support systems like internships and career counseling have a big impact on students' professional decision-making, the alternative hypothesis is validated.

While highlighting the limited relevance of extracurricular activities and peer influence on professional aspirations, this analysis also emphasizes the significance of academic success and formal career assistance in influencing students' career choices.

Conclusion :-

The results of this study offer important new information about the variables affecting students' choice of career. The findings highlight the significance of excellent academic performance in determining future career trajectories by showing a significant correlation between academic achievement and students' job goals. Furthermore, professional decision-making heavily relies on career counseling and organized support networks like internships, underscoring the necessity for educational institutions to improve their career counseling offerings.

On the other hand, students' profession choices are not much influenced by peer pressure or family expectations, indicating that contemporary students place greater weight on their own interests, career guidance, and market trends. In a similar vein, extracurricular activities have little bearing on career goals, although being advantageous for personal growth.

In order to assist students in making informed career decisions, these findings highlight the need for educational institutions to improve internship opportunities, career counseling services, and academic support. The 1

longterm effects of academic achievement and career counseling on students' success in the workplace can be further investigated in future studies.

Implication and Recommendation :-

Implications :-

- Improved Academic Support: Because academic success is strongly linked to career goals, educational institutions should concentrate on offering solid academic foundations, individualized learning opportunities, and focused interventions for students who are having academic difficulties.
- The function of career counseling The necessity for well-structured career advice programs that match students' goals with industry expectations is highlighted by the substantial influence of career counseling and other support systems, such as internships.
- Limited Influence of Peer and Family Pressure: Students are becoming more independent in their professional choices, and as a result, career counseling approaches must change to emphasize individual strengths and market realities over traditional family-driven preferences. Peer pressure and family expectations have little effect on career decisions.
- Extracurricular Activities and Career Goals – Extracurricular activities help people develop their skills and personalities, but they have little direct impact on career goals. This implies that in order to close this gap, educational institutions must incorporate extracurricular activities that are focused on careers, like exposure to the market and leadership development.

Recommendations

1. Enhancing Career Guidance Services: Educational institutions ought to make investments in career counseling divisions manned by qualified experts who can offer tailored advice according to students' interests, abilities, and employment patterns.
2. Growing Programs for Industry Exposure and Internships The curriculum should incorporate industry partnerships, mentorship programs, and internships to give students practical experience and help them make informed career decisions.
3. Academic Excellence Programs: To help students match their performance with their professional goals, schools and institutions should provide academic improvement programs, skill-based learning, and remedial sessions.
4. Customizing Extracurricular Activities: To increase their influence on students' job choices, educational institutions should include career-relevant extracurricular activities including industry projects, entrepreneurial endeavors, and skill-building contests.
5. Policy Reforms in Educational Institutions: By combining AI-driven career counseling tools, job market analytics, and professional development workshops, higher education policymakers may bridge the gap between academia and employment by giving priority to defined career pathways.

By putting these suggestions into practice, educational institutions may improve students' decisionmaking and make sure they get the professional and academic assistance they need to successfully accomplish their career objectives.

Scope of Further research :- The study provides valuable insights into the factors influencing students' career decision-making; however, several areas warrant further exploration:

- Effects of Career Counseling Over Time Future studies can look at how career counseling and guidance programs affect students' professional achievement and career paths over the long run. Deeper insights would be obtained by a longitudinal research that followed students from higher secondary school through their careers.
- Impact of Socioeconomic Background: More research can look into how socioeconomic characteristics, like parental education levels, financial limitations, and resource availability, affect students' profession choices, especially in different educational environments.

- **AI Integration in Career Counseling** Research may now examine the effects of AI-powered career counseling tools on students' decisionmaking, offering tailored suggestions based on academic achievement, job market trends, and talent evaluations.
- **Effectiveness of Experiential Learning Programs:** Future research can evaluate how internships, apprenticeships, and industry partnerships help close the knowledge gap between academic study and career readiness by gauging how well they prepare students for the workforce.
- **Comparative Evaluation of Various Educational Systems**One way to examine differences in students' career decisionmaking tendencies is to compare several educational boards (state, national, and international curriculum).
- It is possible to compare various educational boards (state, national, and international curricula) in order to examine differences in students' career decision-making styles.
- **Psychological Aspects Affecting Career Decisions**In order to comprehend how these elements interact with academic achievement and professional goals, research might examine the psychological aspects of career selections, such as self-efficacy, motivation, personality traits, and resilience.
- **Effects of Skill-Based Education on Professional Results** Future research can evaluate the effects of skill-oriented learning on employability, entrepreneurial success, and professional development, given the increasing focus on vocational and skill-based education.
- **The function of online certifications and digital learning**As MOOCs, professional certifications, and online learning platforms gain popularity, more research may examine how well they affect job choices and improve employability.

By tackling these topics, future studies can offer a more thorough comprehension of the changing dynamics of job choice, assisting academic institutions, decisionmakers, and industry participants in creating more efficient student support networks.

References:

- 1) Sahu, P. and et.all (2022). Predictive analysis of students' performance evaluation in higher education: A machine learning approach. *International Journal of Advanced Research in Computer and Communication Engineering*, 11(5).
- 2) Makombe, F., & Lall, M. (2020). A predictive model for the determination of academic performance in private higher education institutions. *International Journal of Advanced Computer Science and Applications*, 11(9), 415.
- 3) Alyahyan, E., & et.all (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(3). <https://doi.org/10.1186/s41239-020-0177-6>
- 4) Hasan, R., Palaniappan and et.all (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences*, 10(11), 3894. <https://doi.org/10.3390/app10113894>