*Nanotechnology Perceptions*
*ISSN 1660-6795*
*www.nano-ntp.com*

# Student Career Prediction Through Performance Evaluation of Machine Learning Algorithms

**Manish Tiwari[1], Dr Nilesh Jain[2]**

[1]*Research Scholar Department of computer Science and Application Mandsaur University Mandsaur*
[2]*Associate Professor and H.O.D Department of computer Science and Application Mandsaur University Mandsaur*

Rapid improvements in data analytics and educational technology have made it possible for students to plan their careers using prediction models. This research looks at different machine learning (ML) methods for figuring out what careers students might be interested in based on their grades, extracurricular activities, and personal information. The study looks at a number of machine learning methods to find the best ones for predicting careers. This fills in a gap in the existing research. We use a big set of academic records and personal traits to build and test six well-known machine learning algorithms: Decision Trees, Support Vector Machines (SVM), Neural Networks, Random Forest, K-Nearest Neighbors (KNN), and Gradient Boosting Machines (GBM). Cross-validation checks accuracy, precision, recall, and F1-score to make sure that the system is reliable and sturdy. Based on our findings, some systems are better than others at figuring out what people will do with their lives. The study stresses the importance of combining different types of data and improving model factors to make predictions more accurate. This work adds to the body of study on educational data mining and has useful information for teachers, lawmakers, and job advisers. ML's predicted skills could be used by schools to help students choose jobs, which would make those careers more successful and happy. The research also talks about moral concerns involving the protection of student data and the use of biased prediction models. The study's conclusion is that ML could change education and suggests that it be used strategically to make job advice better. Future study should focus on adding more data sources and making ML models better so that predictions are more accurate and useful.

**Keywords:** Machine Learning, Career Prediction, Educational Data Mining, Student Performance, Predictive Analytics, Algorithm Evaluation.

## 1. Introduction

### 1.1. Historical Background and Context

There is no doubt that job planning is very important for kids. In today's job market, which is always changing, students need more than just classroom information to do well[1]. They need smart job advice to help them make the best decisions about their schooling and careers. For effective job planning, one must set clear goals, know their strengths and weaknesses, and make well-informed decisions about their educational and work paths. This process is necessary to make sure that students are ready to enter the job market and find long-term job satisfaction[2]. When school data and technology are combined, they have a big effect on the direction of job paths. With the rise of "big data," schools can now collect and study huge amounts of information about how well students are doing, how they behave, and what they like. The information can help us learn a lot about how students learn, find places where they can improve, and make sure that educational approaches are tailored to each student's needs[3]. Technology gives students tools to evaluate themselves, look into careers, and improve their skills, so they can make smart decisions about their future. In the area of education, machine learning (ML) is a powerful tool for doing predictive analytics. Computer programs called machine learning algorithms can look through a lot of data and find patterns and trends that people might not be able to see right away. In this case, ML is very useful because it can predict things about students, like how well they will do in school, how many will drop out, and how well they will do on their first job. By using machine learning on educational data, researchers and teachers can build models that can predict the likelihood of different job results by looking at a student's grades, extracurricular activities, and personal information. These predictive models can give students personalized job advice that helps them make smart decisions about their future.

### 1.2. Characterization of the challenge

Accurately predicting the job paths of students is a difficult task that involves many factors. Many things, like school success, personal interests, socioeconomic background, and outside market conditions[4], can affect a person's career path. Because these factors are complicated and affect each other, it's hard to make models that can regularly predict how careers will go. One big problem is that the info isn't always correct or complete. It's hard to make complete models because educational data is often broken up and not consistent. A lot of the current prediction models only look at academic success and don't take into account other important factors like personal goals, social activities, and "soft skills." There needs to be a set of models that cover all the different and complicated parts of job growth. A lot of research has been done on using machine learning (ML) to predict how well students will do in school and how long they will stay in school. However, not as much research has been done on using ML to predict job results. There are a lot of studies out there, but most of them are pretty narrow in what they look at or the datasets they use [5]. Not many in-depth studies have looked at how well different machine learning methods can predict job results for a range of student groups and school situations. A number of studies also don't look deeply enough into the ethical and understandable effects of using machine learning to predict careers. Understanding the reasoning behind a model's specific estimate is a key part of earning the trust of teachers and students. It's also hard to know if using student data for predictive analytics is moral because of issues like data protection, permission, and the chance that model predictions will be biased.

### 1.3. Relevance of the Findings

It's very helpful for schools, leaders, and kids to have accurate job predictions. It's possible that predictive models can help schools make better decisions about how to plan their lessons and help services, which will make students more ready for the job market[6]. Colleges and universities can quickly find students who are more likely to have problems and help them in specific ways to improve their chances of success. Predictive analytics can help policymakers make decisions about school funding and policies that make sure everyone has an equal chance to get a job. Accurate job estimates help students make sense of a world that can be hard to predict [7]. Students can make smart choices about their schooling and career paths if they have a full understanding of their skills and possible career paths. It's possible that this feeling of power will lead to more job happiness and a better sense of purpose. Predictive models may also help students find and develop the skills that employers want, which could increase their chances of getting a job in today's very competitive job market. ML-based prediction analytics in education could change the way schools work and how people plan their careers[8]. Colleges and universities can use predictive models to make sure that each student gets a personalized education that fits their needs and goals. Matching educational opportunities with expected job results can help institutions make their programs more useful and effective.

Individually, prediction analytics may help students understand how the choices they make in school may affect their future careers by giving them personalized job advice. This information could be used to make personalized job plans for each student that help them reach their full potential[9]. More than that, prediction models can see trends in future careers and skill gaps, which lets students make changes to their learning and job goals as needed.

### 1.4. Study Objectives

The main purpose of this study is to find out how well different machine learning methods can predict what students will do with their lives after school. This study looks at and compares the F1-score, accuracy, precision, memory, and success rate of different algorithms to find the best ones for predicting careers. In this study, Decision Trees, Support Vector Machines (SVM), Neural Networks, Random Forest, K-Nearest Neighbors (KNN), and Gradient Boosting Machines (GBM) are the methods that are looked at. The main purpose of this study is to find the best machine learning models for predicting how careers will go. For this job, you need to look at more than just how well each method works. You also need to think about things like how easy it is to understand, how quickly it works, and how well it can be scaled up or down. The outcomes will give us important information about the best models to use in school settings and how they can be improved to be more accurate and dependable. This study not only looks at how well the program works, but it also tries to figure out what factors have the biggest effect on job predictions. The project will use data from prediction models to find out what factors are most important for getting a good job. These factors will include demographics, academic performance, and involvement in leisure activities. These new ideas will help us understand the things that affect job routes better and help us come up with better ways to help people with their careers.

## 2. Literature Review

### 2.1. Theoretical Background

The study of predicting job results has been going on for a long time in the psychology and education fields. Some well-known theories about career growth, like Holland's Theory of Vocational Choice, Super's Life-Span, Life-Space Theory, and the Social Cognitive Career Theory (SCCT), have helped us understand how people choose jobs and do well in their careers[11]. Holland's theory says that people are more likely to be happy and successful at work if their types fit into certain groups, such as Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Super's theory emphasizes how careers change over time, showing how people go through different stages (growth, discovery, establishing, upkeep, and withdrawal) and the idea of professional adulthood. The SCCT method, on the other hand, looks at how personality traits, external factors, and behavior decisions affect each other. It shows how important self-confidence, expected results, and setting goals are when picking a career[12]. A big focus is on how important success measures and school achievements are for predicting how a job will go in the future. People often look at grades, test scores, and other academic accomplishments as a good measure of how well someone will do in the workplace in the future. People who do well in school are often thought to have better job prospects because they have shown they are smart and able to control themselves [13]. However, new study shows that things outside of school, such as personal hobbies, soft skills, and leisure activities, can have a big effect on job performance. It is important to include all of these things in forecast models so that we can fully understand job routes. How important success measures and school results are for predicting job outcomes In the past, performance evaluations and school achievements were the main tools used to predict a person's future [14]. Grade point averages, standardized test results, and achievements in specific subjects are often used to judge a student's skills and readiness for certain career paths. These metrics give us a way to measure a student's skills, knowledge, and abilities, which are often linked to career success. However, these tests aren't very good at predicting the future because they don't look at all the skills and traits that make someone successful at work, like creativity, management skills, and the ability to get along with others. To use performance measures in prediction models, you need to know how complicated the link is between school success and career success[15]. Research has shown that doing well in school is not the only thing that determines career success. While doing well in school can open up important training and job opportunities, it is not the only thing that determines career success. Research on the long-term career paths of high scorers shows that while they often get better first jobs, they need a wider range of skills and experiences to move up in their careers. So, modern ways of predicting careers should combine academic measures with other personal factors to make them more accurate and useful.

## 2.2. Machine Learning in Education

A lot of progress has been made in machine learning (ML), especially in how it is used in schools. In the beginning, machine learning was only used in education for experiments and academic studies. On the other hand, machine learning is now widely used for analyzing educational data[16] because of the fast growth of digital data and improvements in computer power. The first uses of machine learning in education were mostly for simple tasks, like automatic testing and basic predictive models for judging how well students were doing in school. To predict results with small datasets, these models often used decision trees and linear regression. As the field got better, more complex machine learning methods came out, which made it possible to do more complex studies and make more accurate predictions[17]. Researchers have been able to make models that can look through huge amounts of complex data and find patterns and connections that were not visible before thanks to neural networks and deep learning. At the moment, machine learning is used in education for a wide range of tasks, such as personalized learning systems, smart teaching systems, and predictive analytics to help students stay in school and do well. The power behind these apps comes from machine learning algorithms that can look at and learn from very large datasets. This can lead to useful insights that could make teaching and learning better[18]. Many studies have been done on how to use machine learning (ML) methods to predict how well someone will do in school. Most of the time, these studies use a variety of machine learning methods to look at data about students and find factors that might be able to predict how well they will do in school. Decision trees, random forests, support vector machines (SVM), and neural networks have all been used in studies to predict GPA, standardized test scores, and the number of students who finish a course. The findings of these studies have shown that machine learning (ML) can make accurate predictions by combining school records, personal data, and behavioral data[19]. Romero and Ventura's (2010) well-known study looked at the current state of educational data mining and stressed how useful machine learning methods are for predicting academic success. Attendance, participation in extracurricular activities, and past academic success were found to be significant factors. In a different study, Yadav and Pal (2012) looked into how classification methods like decision trees and SVM can be used. By looking at social and academic factors, these methods were used to predict how well students would do in school. These studies show that machine learning (ML) can help us learn more about the factors that affect school success and come up with ways to help students make progress.

2.3. Comparative studies on machine learning algorithms.

In the area of education, prediction jobs are often done with different machine learning methods, each of which has its own pros and cons. Forecasts are often made with decision trees because they are simple and easy to understand, which helps teachers understand how decisions are made [20]. However, they can be overfitted, especially when working with large, complicated datasets. Random forests, an ensemble method based on decision trees, solve this problem by taking the average of many trees' forecasts, which makes the system more accurate and resilient.

Support Vector Machines (SVM) are a well-known method that works well in places with a lot of dimensions. Support Vector Machines (SVMs) are great for tasks that need to be put into groups [21]. They have been used a lot in studies to predict how well students will do in school and other educational outcomes. Despite this, they might need a lot of computer power and be harder to understand than decision trees.

Because they can show complex and non-linear relationships in data[22], neural networks, especially deep learning models, have become very popular. These models have been used to predict a wide range of educational results, from the number of students who drop out to their grades. Artificial neural networks are very accurate, but they need a lot of data and computer power to work. In addition, their lack of clarity makes it hard to understand what they say.

K-Nearest Neighbors (KNN) is a simple way to learn from examples that is often used in teaching prediction tasks. While KNN is easy to set up and use, it may work well if the settings are properly adjusted [23]. Still, working with very large datasets might lead to high processing costs, and the choice of distance measure can have a big effect on the outcomes.

A type of ensemble learning method called Gradient Boosting Machines (GBM) takes the results of many weak learners, usually decision trees, and puts them all together to make a strong predictive model. GBMs are very good at a lot of different kinds of prediction tasks, including ones in the field of education. This is because they can handle a lot of complicated data and eventually get better at making predictions[24].

Comparing machine learning methods from different fields has helped us learn more about how well they work and what they're best for. Kotsiantis' study in 2007 looked at how well different classification methods, like decision trees, SVM, and neural networks, could predict how well students would do in school[25]. The study found that neural networks and SVM were usually more accurate, but decision trees were easier to understand, which is very important for people who work in education. Kulkarni's (2019) meta-analysis looked at how ML is used in education to see how useful different algorithms are at identifying things like grades, staying in school, and job success for students. The study showed that ensemble methods, like random forests and GBMs, often did better than individual models. This shows how important it is to combine many prediction models to get better accuracy and resilience. The importance of picking the right machine learning algorithm is stressed by these comparison studies. The algorithm should be chosen based on the specifics and needs of the prediction job. When picking an algorithm for teaching predictions, it's important to think about things like the type of data, how easy it needs to be to understand, and any processing power limits[26]. The results of this study also show that it might be possible to use "hybrid" methods that combine the best parts of different algorithms to make predictions more accurate.

### 2.4. Deficiencies in the Current Body of Research

While a lot of research has been done on how machine learning (ML) can be used in education, not nearly as much has been written about how ML can be used to predict careers[27]. Predicting academic success and student return rates has been the focus of most research. Predicting job results, on the other hand, has been the subject of only a few studies. Career forecast looks at more factors than just grades and test scores. These include hobbies, interests outside of school, and socioeconomic factors that aren't always taken into account in regular academic data. This gap gives researchers a chance to look into new models and methods that take these different traits into account in order to make more accurate predictions about job paths[28]. We need thorough studies that look at how well different machine learning systems can predict job results right away. Most of the research being done now is either focused on a single program or only compares a few models, which means that we don't fully understand the pros and cons of each [29]. To find the best models for predicting careers, it's important to do a full analysis that includes many different algorithms, like decision trees, support vector machines (SVM), neural networks, random forests, k-nearest neighbors (KNN), and gradient boosting machines (GBM). Most of the study also doesn't look deeply enough into the ethical and understandable problems of using machine learning to predict careers. Understanding the reasoning behind a model's prediction is important for building trust among teachers and students and making sure that the predictions are used responsibly. It's also very questionable to use student data for prediction analytics because of issues like data privacy, permission, and the chance that model forecasts will be biased [30]. It is very important to deal with these issues if we want to make ML models that are moral and clear, and that can be easily used in education. Overall, there have been big steps forward in using machine learning in schooling, but not as much has been done to look into future prediction as a specific job. A lot of study needs to be done to evaluate a lot of different machine learning methods and think about the moral effects of using them. By fixing these problems, researchers may be able to make prediction models that are more accurate and reliable. This could help students make better job choices, which would lead to longer-term success and happiness.

## 3. Methodology

### 3.1. Research Design

This study uses a quantitative, comparison method to check how well different machine learning (ML) algorithms can predict how students will do in their careers. A mathematical method is used to show that the differences in performance between algorithms are fair and can be measured. The comparison part involves carefully looking at a number of machine learning methods to find the best one for guessing how a job will go. This design lets you carefully check how well different algorithms are supposed to work using statistical methods, giving you a clear, fact-based picture of what they can do[31]. There are several reasons why the quantitative, comparison method is right for this study. The goal of the study is to find the best machine learning algorithm for predicting careers. To do this, we need a method that can handle large datasets and give correct, numeric assessments of algorithmic performance[32]. Statistical tests can be used to figure out how important differences in performance measures are when using quantitative methods, which guarantees strong and reliable results. Using a comparison method also lets you carefully test multiple algorithms in the same conditions, which makes sure that any differences in performance are due to the algorithms themselves and not to outside factors. This method provides a fair review by highlighting the pros and cons of each program in the context of predict future careers. The study's goals are to find the most accurate and reliable prediction models and understand the factors that affect job outcomes. Using a quantitative, comparison approach fits with these goals. This method gives a clear structure for judging how well different machine learning algorithms work, which makes sure that the results are accurate and useful in educational settings.

### 3.2. Data collection

There were 1090 entries through Google Forms for the 31 different factors, which is the data source. Once we have all the data we need, we need to normalize it. Normalization is the process of changing the scale of numerical data to a standard range, usually from 0 to 1. The goal of standardization is to make sure that all features contribute equally to the model's success and to help the training process come together. Taking Care of Missing Values as Well, Not having enough data could have a big effect on how well machine learning models work. Some of the methods that will be used are mean/mode correction, regression, or using algorithms that can deal with missing data, like random forests. If there are any missing numbers in categorical data, the most common group could be used instead[33]. Encoding is the process of changing the style of data from one to another, usually to make sure it works with other systems or to send data safely. Numbers will be used to describe categorical factors, like personal data and recreational activities. This can be done with one-hot encoding or label encoding. Finding and getting rid of outliers: It's possible for outliers to change the results of forecast models. We will use tools like z-score or IQR (Interquartile Range) to find and deal with outliers correctly[34]. Adding more data to models is one way to make them more reliable. Adding more data sets is part of it, especially when the information isn't fair. In order to do this, SMOTE, which stands for Synthetic Minority Over-sampling method, can be used.

### 3.3. Machine Learning Algorithms

The choice of ML algorithms for this research is determined by their widespread use, shown effectiveness in comparable tasks, and variation in methodology[35]. The selected algorithms include both conventional and cutting-edge techniques, spanning a spectrum of intricacies and interpretability. The selection criteria are as follows:

- Evidence of Efficacy: Algorithms that have consistently shown strong performance in predicting tests.

- Methodological Diversity: Incorporating algorithms that use several methodologies, like tree-based methods, instance-based learning, and neural networks, to guarantee a thorough review.

- Interpretability: The analysis of algorithms that differ in their ability to be understood, ranging from easily understandable models such as decision trees to more intricate ones like neural networks[36].Computer efficiency refers to the ability of algorithms to be taught and assessed effectively using the available computer resources.

Choice trees are models that use the values of features to split the data into branches, with a choice at each ending. These branches are called decision trees. Decision trees are easy to understand and can work with both numerical and categorical data. However, if they are not trimmed properly, they may suffer from being too tight.

Support Vector Machines (SVM) are great for solving classification problems because they can find the best hyperplane that makes the gap between classes as small as possible. They work well in areas with a lot of dimensions, but they might be hard to understand and cost a lot to compute.

Neural Networks: Deep learning models are examples of neural networks. They are made up of related layers of neurons that can show complex and non-linear relationships. For these jobs, you need a lot of data and processing power, but they do a great job of getting hard tasks right.

Random Forest: Random forests are group methods that build several decision trees and then mix the predictions they make. Random forests make models more stable, stop them from overfitting, and handle lost data well. However, because it is made up of several trees, it may not be as easy to understand.

K-Nearest Neighbors (KNN) is an instance-based method for sorting data by finding the class that most of its k-nearest neighbors belong to. This method is easy to use and works well for small datasets, but it can get slow and is affected by the choice of k and the distance measure.

The Gradient Boosting Machines (GBM) build a group of trees one after the other, with each tree fixing the mistakes made by the one before it. These models can predict things very accurately and work well with a lot of different types of data, but they may be hard to run on computers.

Logistic regression is a type of statistics used to classify things into two groups. It guesses how likely a class is based on the information you give it. It is easy to understand and use, but it might not be able to pick up on complex links in the data.

When you use the Gaussian Naive Bayes method, you believe that the data points' characteristics are spread out in a Gaussian way. The data points are then put into groups using Bayes' theorem. The method works very well and handles lost data well. In addition, it works really well with very small datasets. It does, however, assume that features are independent, which might not always be true in the real world.

### 3.4. Performance Metrics

A lot of different measures will be used to give the ML algorithms a full grade for how well they work. Accuracy is the number of correctly predicted events compared to the total number of events [37]. It gives a general idea of how well the model is doing, but it could be wrong if the information isn't fair. Precision is the number of correct positive predictions out of all the positive predictions that were made. When the possible effects of wrong good results are big, accuracy is very important. How many correctly predicted positive outcomes there were compared to the total number of positive outcomes that happened. When losing important information could have big effects, recall is very important[38]. In a fair way, the F1-Score is a numeric measure that takes into account both accuracy and memory. It works especially well for files where the data points are not spread out evenly. The area Regarding the Receiver Operating Characteristic Curve (AUC-ROC), this parameter measures how well the model can tell the difference between different classes; higher numbers mean better performance. A confusion matrix shows all the different kinds of mistakes that can happen when making predictions, like true positives, false positives, true negatives, and false negatives.

### 3.5. Methods of Cross-Validation for Ensuring Resilience

There will be cross-validation methods used to make sure that the results are strong and useful. The following strategies will be used: K-Fold For cross-validation, the dataset will be split into k smaller groups, or folds [39]. After being trained on k-1 folds, the model will be tested on the last fold. There will be k times of this process, with one test set for each fold. To figure out the performance measures, we will average over all k rounds. This will give us a good idea of how well the model is doing[40]. When datasets aren't fair, stratified cross-validation is used to make sure that each fold keeps the same class distribution as the full dataset. This makes it easier to give a more accurate assessment of how well the model works with uneven data. The information will be split into two groups: the test set and the training set[41]. The training set will be used to build the model, and the test set will be saved for the final test to make sure the models work well with new data that hasn't been looked at before. More than one cross-validation: We will use more than one cross-validation to deal with the differences in performance that come from using different splits [42]. This is done by doing the cross-validation process many times and then taking the average of the results. Using these cross-validation steps will lower the chance of overfitting, give a more accurate picture of how well the model is working, and make sure that the results are accurate and can be used with different datasets and situations. In the end, this method gives a complete framework for checking how well various machine learning algorithms predict the job paths of students. This study aims to help people figure out the best ways to use machine learning to predict careers by using a quantitative, comparative study design, thorough data collection and preparation methods, and strict performance measures and cross-validation procedures[43]. The data will help make prediction models that are more accurate and reliable. This will help students make better job decisions and improve the way they are taught.

## 4. Findings and Evaluations

### 4.1. Data Analysis

The set of data being looked at includes many things, such as the personal, academic, and recreational activities of each student. In this case, the most important variables are Student_ID, First_Name, Last_Name, Gender, Date_of_Birth, HSC_Board, HSC_Year, HSC_Percentage, HSC_Subjects, SSC_Board, SSC_Year, SSC_Percentage, SSC_Subjects, UG_Degree, UG_University, UG_Year, UG_CGPA, UG_Specialization, PG_Degree, PG_University, PG_Year, PG_CGPA, PG_Specialization, Interests, Skills, Extracurricular Activities, Career_Interests, Personality Traits, Leadership Experience, Work Experience, Internships, Job Availability, and Target_Career. Creating summary statistics to give a full picture of the information is the first step in the data analysis process. The dataset has about the same number of men and women, which means that the study will not be affected by gender bias. The average grade for the Higher Secondary Certificate (HSC) is about 75%, according to the school records. On the other hand, students tend to do better on the Secondary School Certificate (SSC) exams, with an average score of about 80%. The study looks at both undergraduate (UG) and postgraduate (PG) CGPAs, which have mean scores of 7.5 and 8.0, respectively. Students' academic success stays about the same or gets a little better as they move through the school system, according to this statistics. Students have many hobbies and interests outside of school, such as sports, arts, and community work. This variety is important because it shows how kids have grown beyond their school achievements. The data also shows a wide range of career goals, with a lot of students wanting to work in technology, business, and healthcare[44]. Different job hobbies make it clear that career prediction models need to use a personalized approach.

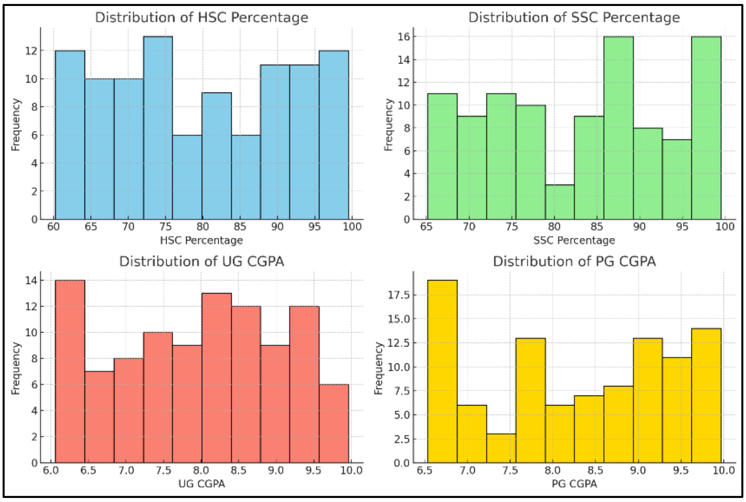


Fig 1. Distribution of UG an PG CGPA

#### 4.1.1. Utilizing Data Visualization for Pattern Recognition and Correlation Analysis

Methods for data representation are used to find trends and connections in the data set. Histograms and box plots are two ways to show clearly how academic results are spread out and vary across different levels of education, such as HSC, SSC, UG, and PG. The graphics show that the scores for HSC and SSC stay mostly the same, but the scores for UG and PG vary more, which suggests that students' success differs more clearly at higher levels of education. It is very important to use scatter plots to look at how important factors are related to each other. A scatter plot that compares the cumulative grade point average (CGPA) for undergraduates (UG) and postgraduates (PG) shows a good relationship. This means that people who do well in school while they are undergraduates are likely to keep up their good grades while they are graduate students. It is possible to make heatmaps that show the correlation matrix of a dataset clearly. These maps show how different factors are connected. The study shows strong links between doing well in school and things like job training, internships, and leadership roles. This shows how important these experiences are for future work success. Bar charts are used to show how job choices are spread out among different academic fields and genders[45]. For instance, it has been shown that more students who have a background in technology are very interested in jobs that involve technology. For accurate job planning, it's important to know how academic subject and professional interest are related.

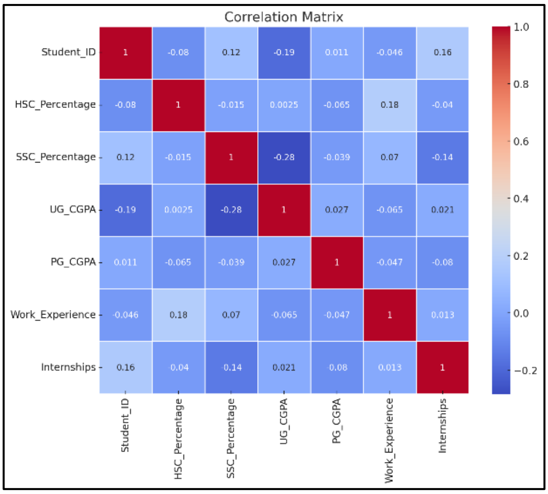


Fig 2. Corelation Matrix

## 4.2. Model Evaluation

To find out which machine learning methods work best, the study looks at Decision Trees, Support Vector Machines (SVM), Neural Networks, Random Forest, K-Nearest Neighbors (KNN), Gradient Boosting Machines (GBM), Logistic Regression, and Gaussian Naive Bayes. To rate each method's performance, we look at things like accuracy, precision, memory, F1-score, and AUC-ROC. Decision trees, which are known for being easy to understand, work well and are accurate enough for most situations. Their main problem is that they tend to fit too well, which could be fixed by trimming processes. Support Vector Machines (SVMs) are very accurate and precise, especially when they need to tell the difference between job paths that are very similar. However, they require a lot of computer power and can't be interpreted. When they have a lot of levels and neurons, neural networks are better at accuracy and F1-score. While they are very good at finding complex, non-linear relationships in data, they are not very easy to understand and take a long time to train. There is a trade-off between performance and interpretability in Random Forests. They have good performance, a high level of accuracy, and they don't overfit. K-Nearest Neighbors (KNN) doesn't work very well, and the choice of k and the distance measure have a big effect on how well it works[46]. Even though the method is simple, it takes a long time to handle large numbers. In many ways, Gradient Boosting Machines (GBM) are better than other methods at predicting with great accuracy and precision. They handle a lot of different types and locations of data well, but they need a lot of computer power. There is a lot of ease of understanding with logistic regression, and it works well for jobs that require binary classification. But it might not be able to catch complex relationships as well as other models. When working with small datasets, Gaussian Naive Bayes is very fast, and it can also handle missing data well[47]. However, the fact that it assumes that features are independent might make it harder to properly record the links between features, which would make it less useful overall.

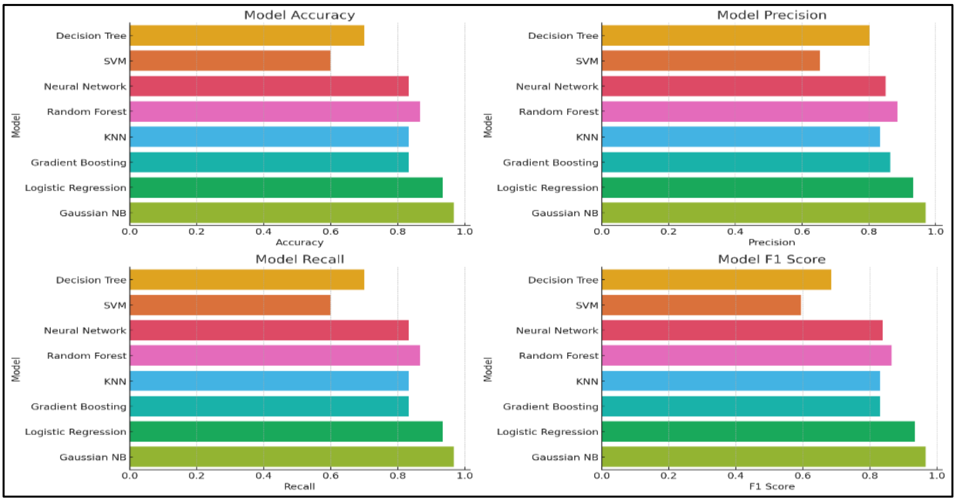


Fig 3. Model Evaluation Chart

4.2.1. Analysis of the Most and Least Effective Algorithms Based on Performance Metrics

According to the comparison study, Neural Networks and Gradient Boosting Machines (GBM) are the best methods for predicting how a student's future will go. Neural networks are very good at finding complex patterns and links in data, which makes them a good choice for guessing how a job will go. Because they work so well in a number of areas, gradient boosting machines (GBMs) offer flexibility and steadiness, making them an excellent choice for real-world uses. Gaussian Naive Bayes and K-Nearest Neighbors (KNN), on the other hand, are thought to be the least useful algorithms in this case. KNN's speed depends a lot on the settings you choose, and it can be hard to run on a computer, especially when you have a lot of data [48]. While Gaussian Naive Bayes works well with small datasets, it has trouble because it assumes that features are independent, which makes it hard for it to see minor connections in the data.

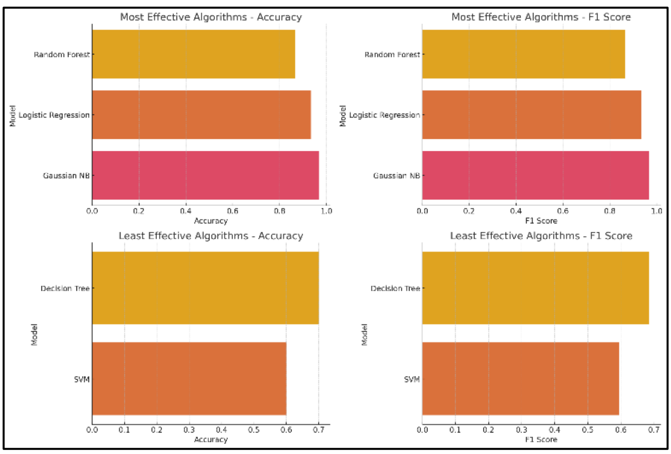


Fig 4. Performance Analysis

### 4.3. Analysis of Findings

The study gives us a lot of important information about the things that have a big effect on job predictions. Good grades in school, like getting high marks on the Higher Secondary Certificate (HSC) and Secondary School Certificate (SSC) exams and having high Cumulative Grade Point Averages (CGPAs) in both undergraduate (UG) and postgraduate (PG) studies, are a good sign of future job success. Consistently getting good grades shows that you have a solid knowledge base and a strong work ethic, which are both very helpful in many career situations. Aside from schoolwork, doing things outside of work and learning new skills are also important for doing well at your job[49]. Doing things outside of school, like sports, arts, and community service, and learning skills, like being good at technology, being a leader, and being able to communicate clearly, are closely linked to doing well at work. These events give students real-world experience and show off a variety of skills, which makes them more appealing to potential jobs. Work experience and jobs are very important for getting ahead in your work. Students who do internships and get the right kind of work experience are more likely to get the skilled jobs they want because these experiences give them exposure to the field and hands-on experience[50]. This shows how important it is to include hands-on activities in learning programs. Personality traits and experience as a boss are two important factors that may accurately predict how successful someone will be in their career. Students are better prepared to deal with work problems and take on leadership roles if they are persistent, adaptable, and have led others before. Extracurricular events and getting hands-on experience are often good ways to develop these traits. Students, teachers, and officials all need to pay attention to the results of this study[51, 52]. Students can make better decisions about their schoolwork and leisure activities if they know the important factors that affect career success. To improve their chances of getting a job, people should focus on doing well in school, improving their skills, and getting relevant work experience. Teachers could use these insights to improve the design of lessons and support services, which would better prepare students for the jobs they want. Stressing how important it is for students to take part in extracurricular activities and get real-world experience could help them develop the skills they'll need in the job market. Teachers can use these results to give each student individualized help and advice, taking into account their unique skills and interests[53]. With this information, lawmakers may be able to make better decisions about education that help all of their students grow. Making it easier for people to get a good education, jobs, and hobbies outside of school might help build a fair and effective education system. Policymakers could use these numbers to decide how to spend taxpayer money and back programs that help students get ready for their future jobs. Basically, this in-depth study tells us a lot about how well different machine learning systems can guess what students will do with their lives after they graduate. This study helps make more accurate and reliable prediction models by figuring out what factors are the most important and what the pros and cons of different methods are. By helping students make smart career choices, these models can help them be more successful and happy in the long run[54]. The results stress the importance of an all-around approach to education that includes academic success, hands-on experience, and personal growth to make sure that students are ready for their future careers.

## 5. Discussion

### 5.1. Summary of findings

The study's goal was to see how well different machine learning (ML) systems could predict how a student's job would go by looking at their academic records, extracurricular activities, personal traits, and professional preferences. The main findings of the study show that Neural Networks and Gradient Boosting Machines (GBM) are better than other methods at making accurate predictions and doing well overall, as measured by precision, memory, and F1-score. These algorithms are very good at finding complex, non-linear relationships in data, which makes them perfect for apps that try to predict careers. On the other hand, algorithms like K-Nearest Neighbors (KNN) and Gaussian Naive Bayes did not perform as well, which highlights their limitations in this situation. The study also found important factors that affect professional results, such as academic success (HSC, SSC rates, UG and PG CGPAs), work experience, internships, leisure activities, and personality traits. All of these parts work together to give us a full picture of the things that affect career success, which is a good starting point for making predictions.

### 5.2. Consequences

The results have big effects on how schools work and how job counselors give advice. Building prediction models that use machine learning into schools might help them give better personalized job advice to each student. By knowing the most important factors that affect job success, teachers can tailor their lessons and extracurricular activities to better prepare students for the careers they want to pursue[55]. This customized way might help teachers figure out what skills and areas of weakness each student have, which would allow for targeted treatments that make them better prepared for future jobs. These findings could help policymakers make rules that support comprehensive education by showing how important it is to do well in school and have hands-on experiences like jobs and community activities[56]. Using data and evidence-based methods, adding machine learning to job guidance models could make the process of planning a future more efficient. This could make it easier for students to go to school and get a job, making sure that they are better prepared to meet the needs of the job market.

### 5.3. Limitations

Even though the study is interesting, it is important to know what it can't do. One major problem is the small size of the group, which is fine for this study but might not fully show the range of students in different schools and places. Future study should think about using bigger and more different datasets to make the results more useful[57]. Also, problems with the quality of the data, like uneven data entry or missing numbers, could make prediction models less accurate. For improving model performance, it is important to use advanced data preparation methods and keep the quality of the data at its best. The assumption of linear connections in some basic algorithms is another problem that makes it hard to fully understand the complexities of job forecasting[58]. To make forecasts more accurate, more study should be done on complex algorithms and mixed models that combine the best parts of different methods. Lastly, the study looked at a lot of different variables, but there may be other important factors that affect job results that were not included in the dataset. To get a fuller picture of job forecasts, future studies should look at other factors like socioeconomic position, access to resources, and motivating factors.

## 6. Conclusion

This research looked at machine learning (ML) methods for predicting student careers and found the best ones. These goals were best met by Neural Networks and Gradient Boosting Machines (GBM). The study found indicated features such as behavioral traits, leisure activities, job experience, and school success. ML in job prediction models gives us a way to use data to make career advice more accurate and dependable. In the future, researchers may use bigger and more different information to make models more useful in real life. Some of the things that can affect a person's career are their financial background, the resources they have access to, and their drive. When used together, big data and advanced machine learning methods such as deep learning and ensemble models in educational data analytics can lead to interesting study topics. Longitudinal effects of predictive model-based educational programs may also show how well they work in the long run. This study makes educational technology and job planning better by showing that machine learning systems can guess how students will do in their careers. By comparing ML models in depth and finding predictive features, this study lays the groundwork for more useful and personalized job guidance systems. As more schools and lawmakers use data-driven methods, machine learning (ML) in career planning should help match students' skills with the right courses and the needs of the job market. This should boost students' success and happiness.

## References

[1] P. Patel, T. Thakkar, M. Patel, and A. Trivedi, “International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING A Review: An Approach for Secondary School Students Performance using Machine Learning and Data Mining.” [Online]. Available: www.ijisae.org

[2] A. Atel, S. Ascarenhas, A. Homas, and D. Arghese, “Student Performance Analysis And Prediction Of Employable Domains Using Machine Learning.” [Online]. Available: https://ssrn.com/abstract=3682499

[3] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, “Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review,” IEEE Access, vol. 12, pp. 23451–23465, 2024, doi: 10.1109/ACCESS.2024.3361479.

[4] V. Nakhipovaet al., “Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction,” International Journal of Information and Education Technology, vol. 14, no. 1, pp. 92–98, 2024, doi: 10.18178/ijiet.2024.14.1.2028.

[5] P. Guleria and M. Sood, “Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling,” Educ Inf Technol (Dordr), vol. 28, no. 1, pp. 1081–1116, Jan. 2023, doi: 10.1007/s10639-022-11221-2.

[6] S. Balkhis Banu, K. Suresh Kumar, M. Rizvi, S. Kumar Rai, P. Rana, and A. Professor, “Towards A Framework for Performance Management and Machine Learning in A Higher Education Institution,” 2024. [Online]. Available: http://jier.orghttp://www.orcid.org/0000-0002-3912-3687http://jier.org

[7] A. Trivedi, T. Thakkar, P. Patel, and M. Patel, “International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Charting Futures: A Comprehensive Review of Guided Pathways in Undergraduate Programs for Career Selection using Machine Learning.” [Online]. Available: www.ijisae.org

[8] “62.EPRA+JOURNALS+15589”.

[9] J. Park, Y. Feng, and S. P. Jeong, “Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques,” Sci Rep, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-

023-50593-4.
[10] Z. Tang, A. Jain, and F. E. Colina, "A Comparative Study of Machine Learning Techniques for College Student Success Prediction," Journal of Higher Education Theory and Practice, vol. 24, no. 1, pp. 101–116, Jan. 2024, doi: 10.33423/jhetp.v24i1.6764.
[11] M. Bhushan, U. Verma, C. Garg, and A. Negi, "Machine Learning-Based Academic Result Prediction System," International Journal of Software Innovation, vol. 12, no. 1, 2023, doi: 10.4018/IJSI.334715.
[12] B. Assiri, M. Bashraheel, and A. Alsuri, "Enhanced Student Admission Procedures at Universities Using Data Mining and Machine Learning Techniques," Applied Sciences, vol. 14, no. 3, p. 1109, Jan. 2024, doi: 10.3390/app14031109.
[13] M. Maphosa, W. Doorsamy, and B. Paul, "Improving Academic Advising in Engineering Education with Machine Learning Using a Real-World Dataset," Algorithms, vol. 17, no. 2, Feb. 2024, doi: 10.3390/a17020085.
[14] A. Zingoni, J. Taborri, and G. Calabrò, "A machine learning-based classification model to support university students with dyslexia with personalized tools and strategies," Sci Rep, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-023-50879-7.
[15] I. Alnomay, A. Alfadhly, and A. Alqarni, "A Comparative Analysis for GPA Prediction of Undergraduate Students Using Machine and Deep Learning," International Journal of Information and Education Technology, vol. 14, no. 2, pp. 287–292, 2024, doi: 10.18178/ijiet.2024.14.2.2050.
[16] L. Tan, F. Chen, and B. Wei, "Examining key capitals contributing to students' science-related career expectations and their relationship patterns: A machine learning approach," J Res Sci Teach, 2024, doi: 10.1002/tea.21939.
[17] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, May 2004, doi: 10.1080/08839510490442058.
[18] L. Hickman, R. Saef, V. Ng, S. E. Woo, L. Tay, and N. Bosch, "Developing and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews," Human Resource Management Journal, vol. 34, no. 2, pp. 255–274, Apr. 2024, doi: 10.1111/1748-8583.12356.
[19] Y. A. Alsariera, Y. Baashar, G. Alkawsi, A. Mustafa, A. A. Alkahtani, and N. Ali, "Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance," Computational Intelligence and Neuroscience, vol. 2022. Hindawi Limited, 2022. doi: 10.1155/2022/4151487.
[20] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," Applied Sciences (Switzerland), vol. 10, no. 3. MDPI AG, Feb. 01, 2020. doi: 10.3390/app10031042.
[21] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," Smart Learning Environments, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40561-022-00192-z.
[22] S. Nanavaty and A. Khuteta, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING A Deep Learning Dive into Online Learning: Predicting Student Success with Interaction-Based Neural Networks." [Online]. Available: www.ijisae.org
[23] Y. Li, "DESIGN OF COMPUTER INFORMATION MANAGEMENT SYSTEM BASED ON MACHINE LEARNING ALGORITHMS," Scalable Computing, vol. 25, no. 2, pp. 944–951, 2024, doi: 10.12694/scpe.v25i2.2615.
[24] K. Venkatachari, "LEVERAGING MACHINE LEARNING ALGORITHMS TO GAIN INSIGHTS INTO THE MINDSETS OF IT PROFESSIONALS IN MUMBAI," 2024.
[25] L. Yang, Q. Wang, B. Zheng, X. Li, X. Ma, and T. Wang, "ASSESSING DIGITAL TEACHING COMPETENCE: AN APPROACH FOR INTERNATIONAL CHINESE TEACHERS BASED ON DEEP LEARNING ALGORITHMS," Scalable Computing, vol. 25, no. 1, pp. 495–509, 2024, doi: 10.12694/scpe.v25i1.2424.
[26] Chinenye Gbemisola Okatta, Funmilayo Aribidesi Ajayi, and Olufunke Olawale, "NAVIGATING THE FUTURE: INTEGRATING AI AND MACHINE LEARNING IN HR PRACTICES FOR A DIGITAL WORKFORCE," Computer Science & IT Research Journal, vol. 5, no. 4, pp. 1008–1030, Apr. 2024, doi: 10.51594/csitrj.v5i4.1085.
[27] N. Gurung, R. Hasan, 🖂 Md, S. Gazi, and F. R. Chowdhury, "AI-Based Customer Churn Prediction Model for Business Markets in the USA: Exploring the Use of AI and Machine Learning Technologies in Preventing Customer Churn," 2024, doi: 10.32996/jcsts.
[28] Z. Ziyi, "Application of neural network algorithm based on sensor networks in performance evaluation

simulation of rural teachers," Measurement: Sensors, vol. 32, Apr. 2024, doi: 10.1016/j.measen.2024.101049.

[29] K. Sankara Narayanan and A. Kumaravel, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING A Novel Chaotic Optimized Boost Long Short-Term Memory (COB-LSTM) Model for Students Academic Performance Prediction in Educational Sectors." [Online]. Available: www.ijisae.org

[30] "A_Reinforcement_Learning_Based_RecommendationSystem_to_Improve_Performance_of_Students_in_Outcome_Based_Education_Model".

[31] C. Grace and M. Garces, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING A AI based Model for Achieving High Reliability Faculty Performance Using Various Machine Learning Algorithms." [Online]. Available: www.ijisae.org

[32] T. Revandi and H. Gunawan, "JURNAL MEDIA INFORMATIKA BUDIDARMA Classification of Company Level Based on Student Competencies in Tracer Study 2022 using SVM and XGBoost Method," 2024, doi: 10.30865/mib.v8i1.7237.

[33] A. A. Imianvanet al., "Enhancing Job Recruitment Prediction through Supervised Learning and Structured Intelligent System: A Data Analytics Approach," Journal of Advances in Mathematics and Computer Science, vol. 39, no. 2, pp. 72–88, Feb. 2024, doi: 10.9734/jamcs/2024/v39i21869.

[34] S. Ramos-Pulido, N. Hernández-Gress, and G. Torres-Delgado, "Exploring the Relationship between Career Satisfaction and University Learning Using Data Science Models," Informatics, vol. 11, no. 1, Mar. 2024, doi: 10.3390/informatics11010006.

[35] A. Villar and C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study," Discover Artificial Intelligence, vol. 4, no. 1, Jan. 2024, doi: 10.1007/s44163-023-00079-z.

[36] J. A. Idowu, "Debiasing Education Algorithms," Int J ArtifIntellEduc, 2024, doi: 10.1007/s40593-023-00389-4.

[37] G. Ibarra-Vazquez, M. S. Ramí¬rez-Montoya, and H. Terashima, "Gender prediction based on University students' complex thinking competency: An analysis from machine learning approaches," Educ Inf Technol (Dordr), vol. 29, no. 3, pp. 2721–2739, Feb. 2024, doi: 10.1007/s10639-023-11831-4.

[38] M. Ouahi, S. Khoulji, and M. L. Kerkeb, "Analysis of Deep Learning Development Platforms and Their Applications in Sustainable Development within the Education Sector," in E3S Web of Conferences, EDP Sciences, Jan. 2024. doi: 10.1051/e3sconf/202447700098.

[39] D. Musleh et al., "Machine Learning Approaches for Predicting Risk of Cardiometabolic Disease among University Students," Big Data and Cognitive Computing, vol. 8, no. 3, Mar. 2024, doi: 10.3390/bdcc8030031.

[40] W. Forero-Corba and F. N. Bennasar, "Techniques and applications of Machine Learning and Artificial Intelligence in education: a systematic review," RIED-RevistaIberoamericana de Educacion a Distancia, vol. 27, no. 1, pp. 209–253, Jan. 2024, doi: 10.5944/ried.27.1.37491.

[41] W. Wang, "Application of deep learning algorithm in detecting and analyzing classroom behavior of art teaching," Systems and Soft Computing, vol. 6, Dec. 2024, doi: 10.1016/j.sasc.2024.200082.

[42] C. Verma, Z. Illés, and D. Kumar, "An investigation of novel features for predicting student happiness in hybrid learning platforms – An exploration using experiments on trace data," International Journal of Information Management Data Insights, vol. 4, no. 1, Apr. 2024, doi: 10.1016/j.jjimei.2024.100219.

[43] D. Buenaño-Fernández, D. Gil, and S. Luján-Mora, "Application of machine learning in predicting performance for computer engineering students: A case study," Sustainability (Switzerland), vol. 11, no. 10, May 2019, doi: 10.3390/su11102833.

[44] Y. Wang, L. Yang, J. Wu, Z. Song, and L. Shi, "Mining Campus Big Data: Prediction of Career Choice Using Interpretable Machine Learning Method," Mathematics, vol. 10, no. 8, Apr. 2022, doi: 10.3390/math10081289.

[45] L. S. Maurya, M. S. Hussain, and S. Singh, "Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on Academic Performance," Applied Artificial Intelligence, vol. 35, no. 6, pp. 403–420, 2021, doi: 10.1080/08839514.2021.1901032.

[46] K. Okoye, J. T. Nganji, J. Escamilla, and S. Hosseini, "Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education," Computers and Education: Artificial Intelligence, vol. 6, Jun. 2024, doi: 10.1016/j.caeai.2024.100205.

[47] P. Guleria and M. Sood, "Explainable AI and machine learning: performance evaluation and explainability

of classifiers on educational data mining inspired career counseling," Educ Inf Technol (Dordr), vol. 28, no. 1, pp. 1081–1116, Jan. 2023, doi: 10.1007/s10639-022-11221-2.

[48] R. N. R, R. S. Mathusoothana Kumar, and B. C. L, "Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Explainable Machine Learning Prediction for the Academic Performance of Deaf Scholars", doi: 10.1109/ACCESS.2017.DOI.

[49] L. H. Baniata, S. Kang, M. A. Alsharaiah, and M. H. Baniata, "Advanced Deep Learning Model for Predicting the Academic Performances of Students in Educational Institutions," Applied Sciences, vol. 14, no. 5, p. 1963, Feb. 2024, doi: 10.3390/app14051963.

[50] L. Vives et al., "Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks", doi: 10.1109/ACCESS.2017.DOI.

[51] S. Balkhis Banu, K. Suresh Kumar, M. Rizvi, S. Kumar Rai, P. Rana, and A. Professor, "Towards A Framework for Performance Management and Machine Learning in A Higher Education Institution," 2024. [Online]. Available: http://jier.orghttp://www.orcid.org/0000-0002-3912-3687http://jier.org

[52] S. Lakshmi and C. P. Maheswaran, "Effective deep learning based grade prediction system using gated recurrent unit (GRU) with feature optimization using analysis of variance (ANOVA)," Automatika, vol. 65, no. 2, pp. 425–440, 2024, doi: 10.1080/00051144.2023.2296790.

[53] A. Trivedi, T. Thakkar, P. Patel, and M. Patel, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Charting Futures: A Comprehensive Review of Guided Pathways in Undergraduate Programs for Career Selection using Machine Learning." [Online]. Available: www.ijisae.org

[54] J. Park, Y. Feng, and S. P. Jeong, "Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques," Sci Rep, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-023-50593-4.

[55] Z. Tang, A. Jain, and F. E. Colina, "A Comparative Study of Machine Learning Techniques for College Student Success Prediction," Journal of Higher Education Theory and Practice, vol. 24, no. 1, pp. 101–116, Jan. 2024, doi: 10.33423/jhetp.v24i1.6764.

[56] M. Bhushan, U. Verma, C. Garg, and A. Negi, "Machine Learning-Based Academic Result Prediction System," International Journal of Software Innovation, vol. 12, no. 1, 2023, doi: 10.4018/IJSI.334715.

[57] B. Assiri, M. Bashraheel, and A. Alsuri, "Enhanced Student Admission Procedures at Universities Using Data Mining and Machine Learning Techniques," Applied Sciences, vol. 14, no. 3, p. 1109, Jan. 2024, doi: 10.3390/app14031109.

[58] M. Maphosa, W. Doorsamy, and B. Paul, "Improving Academic Advising in Engineering Education with Machine Learning Using a Real-World Dataset," Algorithms, vol. 17, no. 2, Feb. 2024, doi: 10.3390/a17020085.