



# STUDENT CAREER PREDICTION USING MACHINE LEARNING

<sup>1</sup>Sanskruti Khot, <sup>2</sup>Ritika Salvi, <sup>3</sup>Vaishnavi Gurav, <sup>4</sup>Mayuri Revale, <sup>5</sup>Rashmi More

<sup>1</sup>Bachelor of Engineering, <sup>2</sup>Bachelor of Engineering, <sup>3</sup>Bachelor of Engineering, <sup>4</sup>Bachelor of Engineering, <sup>5</sup>Assistant professor of Mumbai University

<sup>1</sup>Dept of Computer Engineering,  
<sup>1</sup>RM CET Ambav, Maharashtra, India

**Abstract:** - Every Student goes through multiple phases like peer pressure, parents pressure, etc. However, many students are dropping out of their next level of education due to various reasons. A proposed method utilizes machine learning concepts such as Decision Tree, Random Forest, Support Vector Machine, and AdaBoost. to predict students future career paths. The system, implemented using Python, assists students in choosing a course based on their personality traits, interests, and capacity to take the course. The study found that students are confused about their career options, potentially leading to wrong career selections and reduced human resource productivity. The system aims to help students make informed decisions about their future career paths.

**Index Terms** - Student Career Prediction, Machine Learning, Random Forest.

## I. INTRODUCTION

In today's world, there are lots of different jobs to pick from, and this can make it tough for students to decide what they want to do. But if they want to be successful and achieve their goals, it's super important for them to start planning early. That means they should think often about what they're good at, what they like doing, and how they're doing in school or any other activities they're involved in. By doing this, they can figure out what they're interested in and how they're doing on their way to reaching their goals. This helps them get better at what they're good at and make sure they're on the right track to reaching their dreams. So, it's really important for students to think ahead and take control of their future careers. Right now, many students feel confused when they have to pick a career because there are so many things to consider. It's important for them to look at themselves and figure out what they're good at. Unfortunately, some just pick randomly, which can make them feel really frustrated and down.

Hence, this system aims to assist students in selecting suitable careers. It will be a web application that gathers necessary data from students to predict suitable career paths. This system will provide quizzes for career prediction and help students identify their strengths, enabling them to plan their schedules to achieve their goals and improve their prospects. Using this system, students will be able to identify their strengths and suitable career paths. Accordingly, they can plan their schedules to achieve their goals and improve their prospects. It is a web-based machine learning model built in Python using libraries such as Scikit-learn, Pandas, NLTK, and frameworks such as Django. The database used for the model was obtained from Kaggle. With numerous new career options emerging in different fields, it's challenging for students in tenth or twelfth grade to choose one. They might be confused because they don't know their talents, have many interests, or feel pressure from peers and society. Making the wrong choice can lead to unhappiness and

stress. To address this, students need proper counseling based on their academic performance, interests, and their ability to work in a specific field. While there are websites and apps that help students find suitable careers, most only consider personality traits or interests, which can yield inconsistent results. Few take into account whether the student would actually excel in that field.

## II. LITERATURE SURVEY

[1] **Student Career Interest Prediction Using Machine Learning:** In this paper, the author has proposed how students often feel uncertain about their future careers because there are so many choices nowadays. They believe that predicting a student's career interests could help solve this problem. This prediction is based on things like their academic background and what they're interested in. The paper emphasizes how crucial it is to research this topic because it could greatly benefit students. By understanding what careers students are likely to be interested in, educators and counselors can offer better guidance. This could lead to students making more informed decisions about their futures. So, the paper suggests using machine learning to predict students' career interests, taking into account various factors like their past academic performance and personal interests. Ultimately, the goal is to help students navigate the complex world of career choices more confidently.

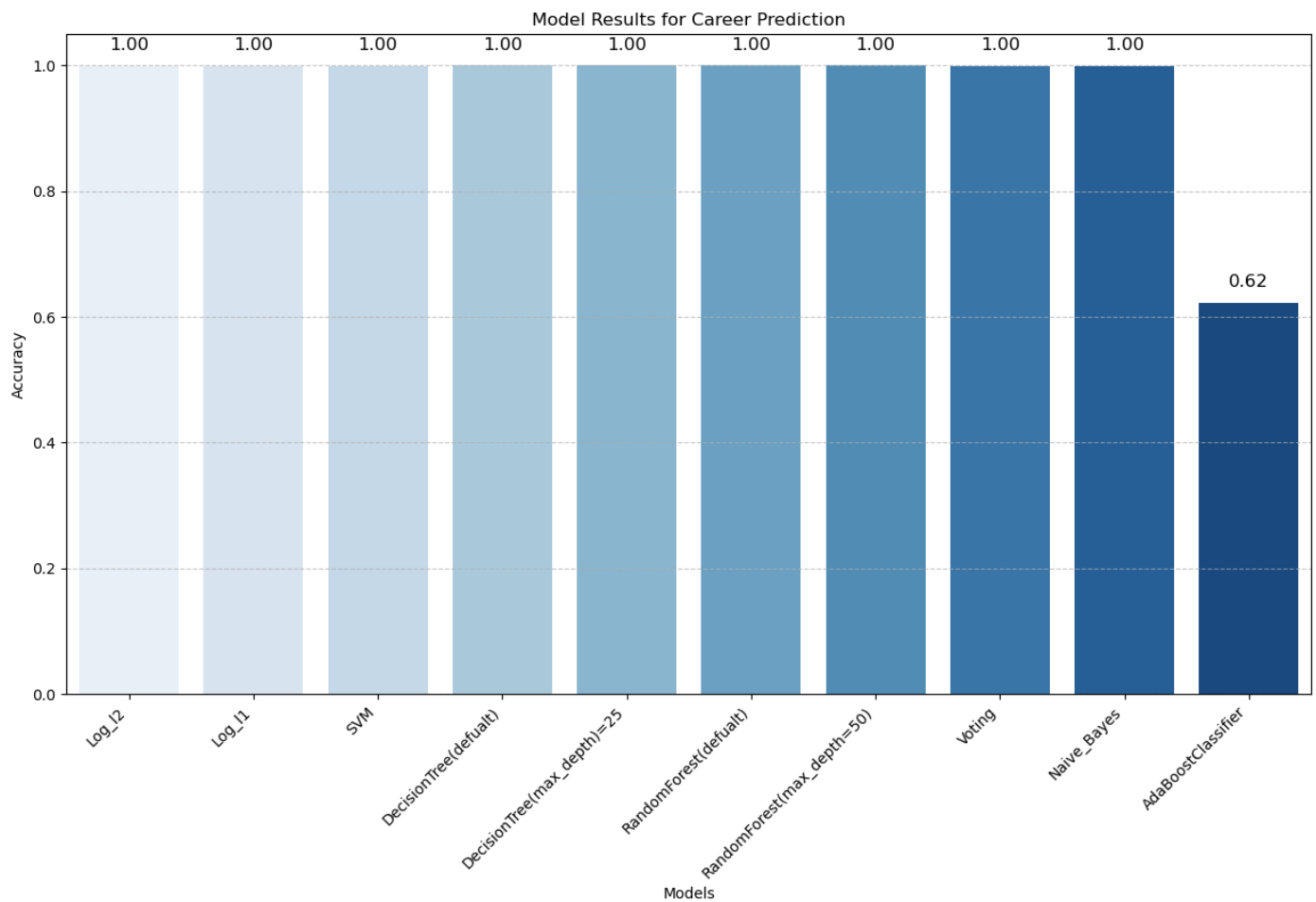
[2] **Student Future Prediction Using Machine Learning:** In this paper author has proposed This paper talks about how important it is for students to choose the right career path because there are so many options available now, and many students feel confused. It mentions that about 40% of students experience this confusion. The paper also discusses how picking the wrong career can negatively affect a person's productivity. To address these issues, the author suggests creating a web application to help high school students choose their careers. This application would consider things like their personality traits, interests, and abilities. The goal is to provide personalized guidance to students so they can make informed decisions about their futures. By using machine learning, the application could analyze various factors and recommend suitable career options for each student. This way, students can feel more confident about their choices and avoid making decisions that might not be right for them in the long run.

[3] **Career Prediction System Based On Student Preferences:** In this survey paper, the author discusses a system designed to predict careers based on what students prefer. They created a content-based recommender system that collects information about students to recommend job domains that match their skills, interests, and academic performance. The paper also compares different algorithms like SVM, Decision tree, and Random Forest to see which one is the most accurate for this task. The goal is to help students find suitable career paths by analyzing their preferences and academic background. By using machine learning techniques, the system can offer personalized recommendations tailored to each student's profile. This way, students can explore career options that align with their strengths and interests, ultimately assisting them in making informed decisions about their future careers.

**III. COMPARISON**

Classifiers	Accuracy
Log_11	99.9799
Log_12	99.9798
SVM	99.9831
Decision Tree(default)	99.9943
Decision Tree(max_depth)=25	99.9898
Random Forest	99.9943
Random Forest(Mx_depth)=50	99.9899
Voting	99.9879
Naïve_Bayes	99.9799
AdaBoost	62.3078

**Table 3.1.** Classification



**Fig. 1.** Accuracy Graph

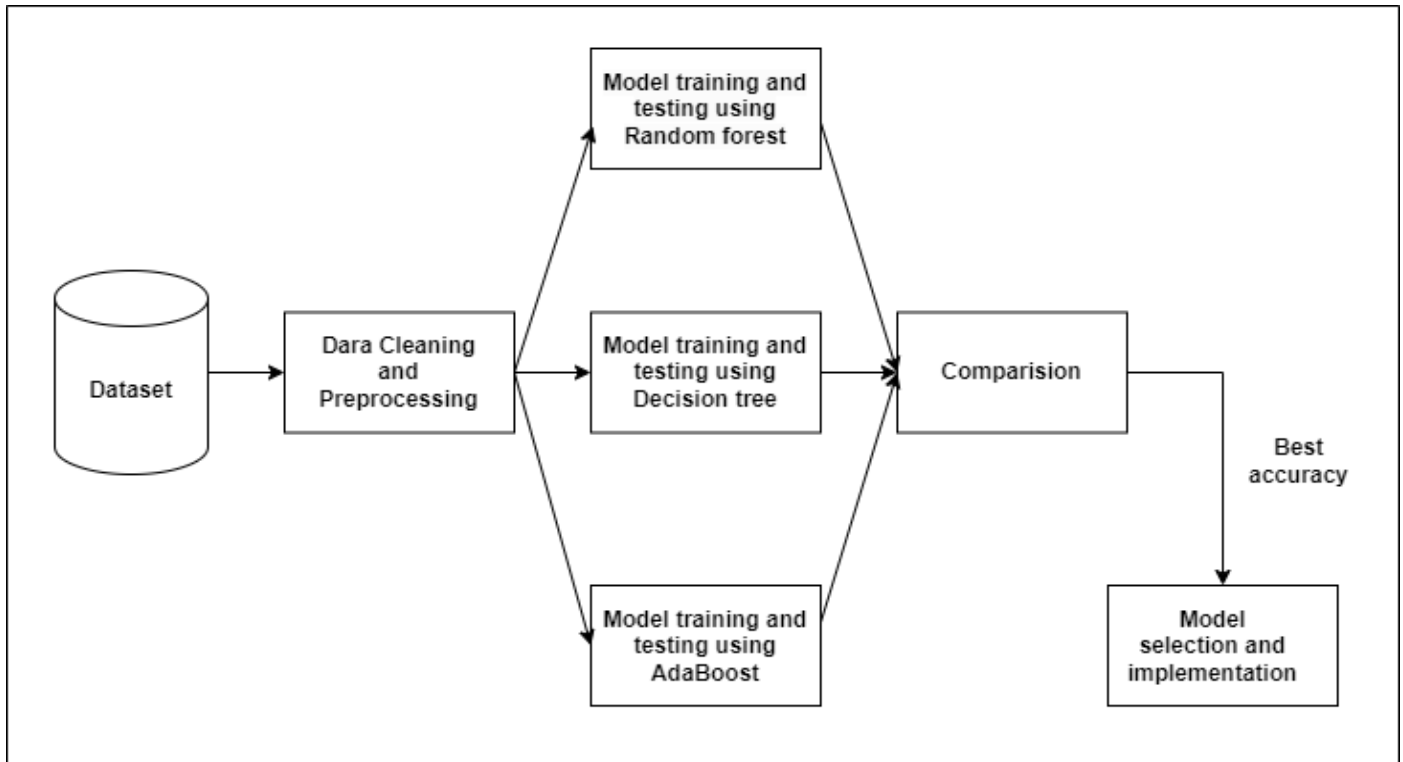
#### IV. PROPOSED SYSTEM

The proposed system utilizes various machine learning techniques such as Decision Trees, Random Forest, AdaBoost, and Support Vector Machines (SVM). Among these techniques, Random Forest stands out as it's an ensemble learning method capable of handling classification, regression, and other tasks effectively. Random Forest works by creating multiple decision trees and combining their predictions to produce more accurate results. It's particularly adept at handling missing data and maintaining high accuracy, even with large datasets. student career prediction using machine learning involves collecting data about students, cleaning and preprocessing that data, selecting relevant features, training a machine learning model, evaluating its performance, and ultimately using it to predict students' future career paths based on their characteristics.

##### IV.I: Objective

The primary objective of the Student Career Prediction Using Machine Learning:

- To suggest career options and predict outcomes while collecting accurate information about learning styles and methodologies.
- To predict the learning style combinations of students and suggest a field or domain using algorithms like k-means, SVM, and decision trees.



**Fig. 2.** Block Diagram of Proposed System

## IV.II Module

The system is proposed to have the following modules:

- a. Data Cleaning & Preprocessing
- b. Model Training & Testing
- c. Comparison
- d. Model Selection and Implementation

### IV.II.I Data Cleaning & Preprocessing

Cleaning all these data and replacing them with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in preprocessing of data. In machine learning, we preprocess data to make it suitable for our algorithms to work with.

### IV.II.II Model Training & Testing

In order to assess the performance of machine learning models like Random Forest, Decision Tree, and AdaBoost, a common practice is to split the available dataset into two parts: a training set, which typically comprises 80% of the data, and a testing set, which consists of the remaining 20%. The models are then trained on the training set and evaluated on the testing set to measure their accuracy. This process allows for validating the models' ability to generalize to unseen data. By comparing the predicted outcomes with the actual labels in the testing set, the accuracy of each model can be calculated, providing insights into their effectiveness in making accurate predictions. This approach helps in selecting the most suitable model for the given task based on its performance on unseen data.

#### IV.II.III Comparison

In machine learning, when comparing different algorithms, you typically train each algorithm on your dataset and evaluate their performance using a metric like accuracy. The predicted values from each algorithm are then compared to the actual values to assess how well each algorithm has learned from the data. The algorithm that achieves the highest accuracy on the dataset is considered the best model for that particular task. This process helps in selecting the most effective algorithm for solving a specific problem based on its predictive performance.

#### IV.II.IV Model Selection and Implementation

The model with the highest accuracy among Random Forest, Decision Tree, and AdaBoost will be chosen for the project implementation. This model selection ensures that the chosen algorithm performs the best in making accurate predictions, thereby maximizing the project's success rate. By implementing the most accurate model, the project aims to achieve optimal outcomes and effectively address the problem at hand

### V. CONCLUSION

This system will help students to use it to predict the suitable course. System helps to Minimize the failure ratio and to take acceptable action for their career and facilitate the students, as it Will guide them to take an appropriate decision while choosing the stream as her career. This System will be helpful for very individual career guidance.

### REFERENCES

- [1] Prof. Priyanka Shahane, Prutha Rinke, Taniksha Datar, Soham Badjate, "Student Career Interest Prediction using Machine Learning", IRJET, November 2022.
- [2] Dileep Chaudhary, Harsh Prajapati, Rajan Rathod, Parth Patel, Rajiv Kumar Gurjwar, "Student Future Prediction Using Machine Learning", ResearchGate, July 2020.
- [3] Shivakoti Udayasri, Ch Varsha, "Career Prediction System Based On Student Preferences", IRJMETS, July 2017.
- [4] Jie Xu, Kyeong Ho Moon, Mihaela van der Schaar, "A Machine Learning Approach from Tracking and Predicting student Performance in Degree Programs", *IEEE*, July 2017.
- [5] Yulius Lie, Bens Pardamean, "Information System Model of Succession Planning and Career Path", *IEEE*, March 2016
- [6] Bo Guo, Rui Zhang, "Predicting Students Performance in Educational Data Mining", *International Symposium on Educational Technology*, August 2016.
- [7] Ali Daud, Naif Radi Aljohani, "Predicting Student Performance using Advanced Learning Analytics", International Wide Web Conference Committee, July 2016
- [8] A Roshani, PR Deshmukh, "An incremental ensemble of classifiers as a technique for prediction of student's career choice", *IEEE Networks and Soft Computing*, September 2015.
- [9] Roshani Ade, Dr. P. R. Deshmukh, "An incremental ensemble of classifiers as a technique for prediction of student's career choice", First International, December 2015.
- [10] Sudheep Elayidom, Dr. sumam Mary Idikkula, "Applying Data mining using Statistical Techniques for Career Selection", *International Journal of Recent Trends in Engineering*, May 2009.