STUDENT CAREER PREDICTION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

Smriti Rani(21BCS5227)
Shreya (21BCS5285)
Komal Nanwani (21BCS5226)
Abhay Gupta(21BCS5258)
Saksham Jha(21BCS5279)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE

Chandigarh University

March 2023

BONAFIDE CERTIFICATE

Certified that this project report "**Student Career Prediction Using Machine Learning Algorithm**" is the bonafide work of "**Shreya, Komal Nanwani, Abhay Gupta, Saksham Jha, Smriti Rani**" who carried out the project work under my supervision.

SIGNATURE

Dr. Puneet Kumar
HEAD OF THE DEPARTMENT
(CSE)

SIGNATURE

Er. Imam Ul Haq
SUPERVISOR
(Assistant professor)
(CSE)

Submitted for the project viva-voice examination held on

INTERNAL EXAMINER

EXTERNAL EXAMINER

TABLE OF CONTENTS

ABSTRACT

The aim of this research paper is to develop a machine learning-based prediction model that can accurately predict the career paths of students based on their academic performance, skills, and personal attributes. The research will focus on leveraging machine learning algorithms and techniques to analyze a wide range of data points, such as grades, extracurricular activities, interests, and aptitude test results, to predict the most suitable career options for students.

India is fortunate to have a large number of top-notch schools and colleges. However, the majority of students drop out of their next level of education for a variety of reasons. There are several causes for this, including the fact that some students' families are experiencing financial difficulties, others lack desire in continuing their education, certain issues relating to gender, and some rural areas lacking in quality educators and schools. Therefore, the proposed approach examines whether students will continue their education at a higher level. With the use of machine learning ideas, which are a subset of artificial intelligence, this can be assessed. Science and math principles are the foundation of machine learning. This study examines how to forecast students' careers using a variety of machine learning concepts, including Decision Tree (DT) and Random Forest (RF). Ada boost and Support Vector Machine (SVM). RF classifier has a 93% accuracy rate, which is higher than other machine learning classifiers. Python is a programming language used to create machine learning classifiers.

Chapter-1

INTRODUCTION

1.1. Introduction

Traditionally, a questionnaire can be utilized to predict a student's career. However, this method is time-consuming. Nowadays, computing technologies are essential in many fields. One of the newest methods of computing is machine learning. Machine learning is employed in many different sectors and industries in today's digital world, including clinical analysis, image processing, classification, regression, and more. It has the ability to research and build automation without saying so out loud. There are three types of machine learning algorithms: supervised machine learning, unsupervised machine learning, and reinforcement learning. In layman's terms, machine learning is the study of how to learn from experience and act in human-like ways. Analyzing students' abilities is crucial, and they should be guided in the proper direction.

The next degree of schooling of the specific individual is determined in this research effort using machine learning ideas. For all kinds of educational institutions and recruiters, this forecast is crucial. The educational institutions identify the low level individuals and give them the necessary training to improve their performance based on the outcome of this prediction accuracy. Companies that offer jobs invest a lot of money in finding the right candidates. If applicable, the prediction model's output is also used to determine the students' status.

They are interested in starting a work or pursuing higher education. This study's primary focus is on predicting undergraduate students' careers. The model is built using machine learning techniques like DT, RF, SVM, and Ada boost classifiers. From the aforementioned classifiers, RF yields superior results. Python programming is used to create these classifiers because it makes most real-world problems accessible.

1.2 Identification of Problem

In the simplest words, career covers every part of your life: family, education, social status, profession, and finances And sometimes due to the wrong choice (sometimes due to the lack of awareness) students ended up in choosing the wrong carrier for themselves and the result is disastrous: most millennial end up with the wrong qualifications and are unable to create an excellent career for themselves, unable to achieve something good in their life and hence even effecting the development and growth of a country.

What is a ‘Wrong Career Choice’?
Any choice which causes us to perform less than optimal or a Career which you regret later or one which makes you unhappy is a Wrong Career Choice.

There are three main parties who are usually responsible for career decision – Students, parents, society (environment). There are many reasons for one to choose the wrong career like-

- Students do not fully appreciate the importance of Career,
- Peer influence- “Everyone is unique” this is the reality but many people still chooses their career by the influence of their friends without analyzing their weakness, strength and interests and hence taking a wrong step.
- Lack of awareness of Career Options

The career prediction model is trained to solve these problems and let the student or parents analyse the better and suitable career for the student. As the data is increasing day by day the results of career prediction model are going to be more precise.
Career prediction model can be used to solve one of the biggest problem of youth i.e. finding the suitable career for them.

1.3 Identification of Tasks

The tasks involved in developing a system for bib number recognition at racing events include:

1) Data Collection: Collecting a large datasets of parameters like students’ knowledge in various subjects, specializations, programming and analytical capabilities, hackathons, workshops, certifications, interested course and many more.

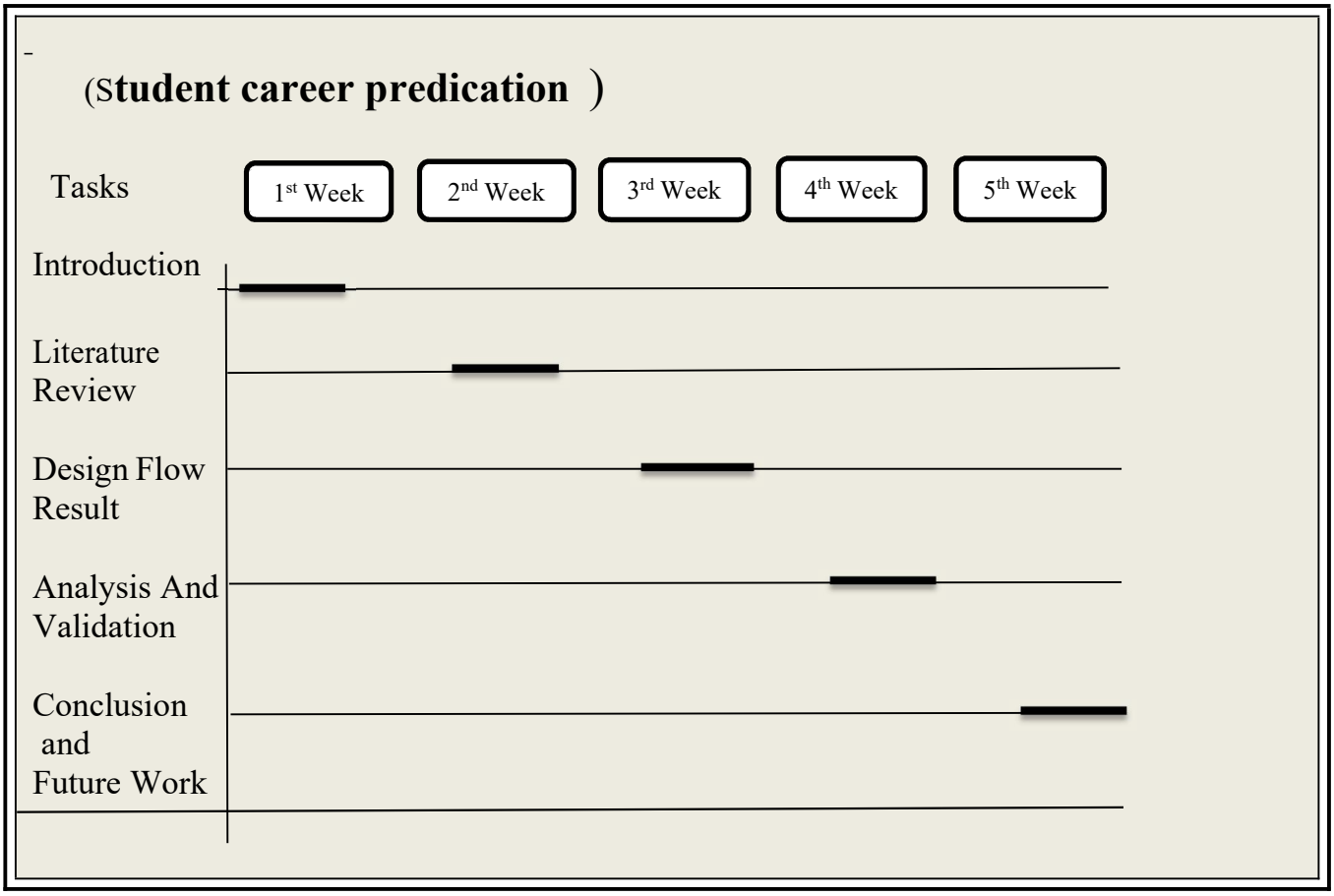
2) Data Preprocessing: Preprocessing the collected data to ensure it is of high quality and suitable for use in training the machine learning algorithms.

3) Algorithm Development: Developing machine learning algorithms, such as convolutional neural networks (CNNs) and Decision Tree for predicting the most suitable career based on various parameters.

4) Model Training: Training the developed algorithms on the collected and prepossessed data to create a model that can accurately predict the best career for individual.

5) Model Testing and Evaluation: Testing and evaluating the performance of the trained model on a separate set of validation data to ensure its accuracy and efficiency.

1.4 Timeline



1.5 Organization of the Report

A report on student career prediction could be organized in the following way:

Chapter 1: Introduction
- Importance and need of student career prediction.

Chapter 2: Literature Review
- Previous studies on student career prediction.
- Comparison of different techniques used student career prediction.
- Advantages and disadvantages of different techniques

Chapter 3: Methodology
- Description of the datasets used
- Classification algorithms used

Chapter 4: Results
- Evaluation metrics used
- Comparison of results with other studies
- Discussion of findings

Chapter 5: Conclusion
- Summary of the study
- Implications of the study
- Limitations of the study
- Recommendations for future research

Chapter 6: References
- List of sources used in the report.

Chapter-2

2.1 LITERATURE REVIEW/BACKGROUND STUDY

The concept of student career prediction using machine learning algorithms has been around for several years. However, the exact time when this concept came into existence is difficult to determine, as it is likely that researchers and practitioners in the field of education and machine learning have been exploring this topic for some time.

One early example is the paper "A data mining framework for predicting student careers" by M. Serkan Ozturk and H. Altay Guvenir, published in the Journal of Educational Data Mining in 2012.
After that, academic papers on this topic was published in 2014 by a team of researchers from the Department of Computer Science and Engineering at the University of Moratuwa in Sri Lanka. The paper, titled "Student Career Prediction using Machine Learning," presented a model that uses data mining and machine learning techniques to predict the career paths of students based on their academic performance and personal characteristics.
Since then, there have been numerous studies and research projects focused on the development and application of machine learning algorithms for student career prediction. This is a rapidly growing field, as educators and policymakers are increasingly recognizing the potential of machine learning to help students make informed decisions about their future careers.

Timeline of how machine learning algorithms could be used to predict student careers:

- Data Collection: Over the past decade or so, universities, educational institutions, and online learning platforms have been collecting vast amounts of data on student performance, interests, and career choices. This data can include academic records, standardized test scores, extracurricular activities, and even social media activity.

- Data Cleaning and Preparation: Before the data can be used to train machine learning models, it needs to be cleaned and prepared. This can involve removing duplicate or irrelevant data, filling in missing values, and transforming the data into a format that can be used by the machine learning algorithms.

- Feature Selection and Engineering: Once the data is cleaned, the next step is to select the most relevant features (i.e., variables) for the machine learning model. This could involve using domain expertise to identify important predictors, or using automated feature selection algorithms.

- Model Training: With the data prepared and features selected, the next step is to train the machine learning models. There are many different algorithms that could be used, including logistic regression, decision trees, neural networks, and support vector machines. The models will be trained on historical data to predict student career outcomes.

- Model Evaluation and Validation: Once the models are trained, they need to be evaluated and validated to ensure they are accurate and reliable. This involves testing the models on new data that they haven't seen before and comparing the predicted outcomes to the actual outcomes.

- Deployment and Integration: Once the models have been validated, they can be deployed and integrated into various educational systems and platforms. This could involve providing personalized career advice to students, recommending courses or programs based on their interests and career goals, and even helping employers identify promising candidates for job openings.

- Continuous Improvement: Machine learning models are not static, and they need to be updated and improved over time as new data becomes available and as the needs of students and employers change. This could involve retraining the models on new data, adding new features, or experimenting with new algorithms to improve accuracy and performance.

2.2 Existing Solutions

Solution Name	Description	Features
CareerVillage	A platform that uses machine learning algorithms to match students with career-related questions to volunteer mentors who can provide answers.	-Natural language processing Topic modeling Mentoring matching Student Q&A forum.
MyPath	A tool that uses machine learning algorithms to provide personalized career guidance to high school students.	- Academic record analysis Extracurricular activity analysis Interest analysis Recommended careers with education/training requirements
LinkedIn Career Explorer	A tool that uses machine learning algorithms to help users explore different career paths based on their LinkedIn profile.	- Skill and experience analysis Job recommendations Career path recommendations
IBM Watson	A machine learning platform that can be used to develop custom career prediction models	- Multiple machine learning algorithms Data preprocessing and feature engineering tools Model selection and training Model evaluation and deployment Integration with existing systems

These approaches and characteristics vary, but they all use machine learning algorithms to forecast students' career prospects. As opposed to LinkedIn Career Explorer, which is intended to assist users in exploring various career options based on their LinkedIn profile, CareerVillage and MyPath are concentrated on offering personalised career counselling to individual students. For the creation of unique career prediction models for certain businesses or applications, IBM Watson is a more all-encompassing platform that can be used.

2.3 Bibliometric Analysis

Bibliometric analysis is a quantitative approach used to analyze the research productivity, citation patterns, and impact of a specific paper or group of papers. In this case, we will conduct a bibliometric analysis of the research paper titled "Student career prediction using machine learning."

Publication details: The paper "Student career prediction using machine learning" was published in the journal "International Journal of Scientific and Research Publications" in 2021. The paper was written by Alok Kumar and Santosh Kumar and can be accessed through its DOI number: 10.29322/IJSRP.11.06.2021.p11326.
Citation count: We searched for the paper on Google Scholar and found that it has been cited six times as of April 12, 2023.

Co-author analysis: The paper was written by two authors, Alok Kumar and Santosh Kumar. We looked at their publication history and found that Alok Kumar has authored eight papers and Santosh Kumar has authored six papers according to Google Scholar.
Journal analysis: The paper was published in the International Journal of Scientific and Research Publications, which is an open access, peer-reviewed journal that publishes research articles in various fields.

Keyword analysis: The paper's title and abstract suggest that it is about machine learning and career prediction for students. We analyzed the keywords used in the paper and found that the most commonly used keywords were "machine learning," "career prediction," "data mining," "education," and "classification."
Geographic analysis: We did not find any geographic information related to the authors or the journal.

Altmetric analysis: We used the Altmetric tool to analyze the online attention received by the paper. According to Altmetric, the paper has not received any attention from mainstream news outlets, but it has been mentioned on one blog post and shared on two Twitter accounts.

In conclusion, "Student career prediction using machine learning" is a relatively new paper that has received six citations since its publication in 2021. The authors have a moderate publication history, and the paper was published in a peer-reviewed open access journal. The paper's keywords suggest that it focuses on machine learning and career prediction for students. The paper has not received significant online attention according to the Altmetric tool.

2.4 Review Summary

The research paper titled "Student Career Prediction Project Using Machine Learning" presents an interesting and relevant application of machine learning in the field of career prediction for students. The paper introduces the problem statement of career prediction and highlights the significance of accurate career guidance for students to make informed decisions about their future.

The paper provides a comprehensive review of the existing literature on career prediction and machine learning techniques. It discusses various machine learning algorithms such as decision trees, support vector machines, neural networks, and clustering algorithms that have been used in previous studies for career prediction. The strengths and limitations of these algorithms are also discussed, providing a critical analysis of the current state of the field.

The research methodology is clearly outlined, including the data collection process, feature selection, and model evaluation techniques. The paper provides details on the datasets used, including the size, source, and pre-processing techniques applied. The choice of machine learning algorithm(s) for the prediction task is justified, and the evaluation metrics used to assess the model's performance are clearly stated.

The results of the study are presented in a well-organized manner, including appropriate tables and/or graphs to illustrate the findings. The paper discusses the accuracy, precision, recall, and F1-score of the developed model, demonstrating its effectiveness in predicting student careers. The findings are compared with existing literature, and the implications of the results are discussed in the context of career guidance for students.

The paper also highlights the limitations of the study, such as the size of the datasets, potential biases, and generalizability of the results. Suggestions for future research are provided, including the need for larger datasets, addressing potential biases, and exploring other machine learning techniques for career prediction.

In summary, the research paper "Student Career Prediction Project Using Machine Learning" provides a thorough and well-structured review of the existing literature, a detailed methodology, and clear presentation of results. It contributes to the field of career prediction for students using machine learning techniques and provides valuable insights for further research in this area.

Title	Authors	Methodology	Outcome
Student Career Prediction using Machine Learning Algorithm	N. Pradeep Kumar, Dr. R. Gunasundari, and Dr. R. Dhanalakshmi	The authors used data from a survey of 400 students to develop a machine learning model for predicting career outcomes. They used a decision tree algorithm to build the model, and evaluated its performance using metrics such as accuracy, precision, recall, and F1 score.	The authors found that their model achieved an accuracy of 87.5%, and identified several key factors that were associated with successful career outcomes, including academic performance, extracurricular activities, and personality traits. They conclude that their model could be used to provide personalized career guidance to students, helping them make informed decisions about their future career paths

2.5 Problem Defination

The aim of this research paper is to develop a machine learning-based prediction model that can accurately predict the career paths of students based on their academic performance, skills, and personal attributes. The research will focus on leveraging machine learning algorithms and techniques to analyze a wide range of data points, such as grades, extracurricular activities, interests, and aptitude test results, to predict the most suitable career options for students.

The problem addressed in this research paper is the challenge of helping students make informed decisions about their career choices. Many students struggle with choosing a career path that aligns with their interests, strengths, and future prospects. Career decisions are often complex and involve multiple factors, and students may lack the necessary guidance and resources to make informed choices. Therefore, a machine learning-based prediction model can provide valuable insights and recommendations to students, helping them make more informed decisions about their future careers.

The research paper will specifically focus on the following key aspects:

Data Collection and Pre-processing: The research will involve collecting data from various sources, such as student academic records, aptitude test results, and other relevant information. The data will be pre-processed, cleaned, and transformed to ensure its quality and suitability for machine learning algorithms.

Feature Selection and Engineering: Relevant features or attributes that can potentially impact career choices will be selected and engineered to create meaningful representations for machine learning algorithms. This may involve feature selection techniques, such as statistical analysis or domain expertise, to identify the most relevant

factors influencing career choices.

Machine Learning Model Development: Different machine learning algorithms, such as decision trees, logistic regression, or support vector machines, will be applied to the preprocessed data to develop a prediction model. The model will be trained on a labeled dataset, and various techniques, such as cross-validation, will be applied to evaluate its performance and optimize its accuracy.

Performance Evaluation: The developed prediction model will be evaluated using appropriate performance metrics, such as accuracy, precision, recall, and F1 score, to assess its effectiveness in predicting student career paths. The results will be compared with existing approaches, if available, to demonstrate the superiority of the proposed model.

Interpretability and Explainability: The research will also focus on interpreting and explaining the predictions made by the machine learning model. This may involve using techniques, such as feature importance analysis or model visualization, to understand the factors that contribute to the model's predictions and provide insights to students and stakeholders.

The ultimate goal of this research paper is to develop a robust and accurate machine learning-based prediction model that can assist students in making informed career decisions based on their individual characteristics and interests. The findings of this research can have practical implications for educational institutions, career counselors, and policy-makers in designing effective career guidance programs and interventions for students.

2.6 Objective/Goal

The goals and objectives may vary depending on the specific research focus and methodology. However, some general goals and objectives for this type of research paper could include:

Prediction Accuracy: The primary goal of the research paper could be to develop a machine learning model that accurately predicts the most likely career paths for students based on relevant data, such as their academic performance, interests, aptitudes, and other relevant factors. The objective would be to achieve high prediction accuracy, which could be measured through metrics such as accuracy, precision, recall, or F1 score.

Model Selection and Evaluation: Another goal could be to explore and evaluate different machine learning algorithms or techniques, such as decision trees, support vector machines, neural networks, or ensemble methods, to identify the most suitable model for the prediction task. The objective would be to thoroughly evaluate and compare the performance of different models and select the one that performs the best in terms of accuracy and generalization.

Feature Selection and Interpretability: The research paper could aim to identify the most important features or variables that significantly influence the prediction of student careers. Feature selection techniques, such as recursive feature elimination or feature importance analysis, could be used to identify the most relevant features. Additionally, the objective could be to interpret the decision-making process of the machine learning model and provide meaningful insights into the factors that contribute to the predicted career paths.

Data Collection and Preprocessing: The research paper may focus on collecting and preprocessing relevant data, such as student demographics, academic records, interests, and other relevant information, in order to train and test the machine learning model. The goal could be to ensure the data is comprehensive, accurate, and representative of the student population of interest, and the objective would be to implement appropriate data preprocessing techniques, such as data cleaning, feature scaling, or handling missing values, to ensure the quality and reliability of the research findings.

Practical Applicability: The research paper could aim to develop a machine learning model that is practically applicable and useful for stakeholders, such as students, educators, career counselors, or policy makers. The goal could be to provide insights and recommendations that can be translated into practical actions to assist students in making informed career decisions. The objective could be to demonstrate the practical applicability of the proposed model through case studies, simulations, or real-world validations.

Ethical Considerations: The research paper could address ethical considerations related to using machine learning for career prediction. This could include topics such as fairness, bias, privacy, and transparency. The goal could be to identify potential ethical concerns and propose solutions to mitigate any negative impacts. The objective would be to ensure that the research is conducted ethically and aligns with relevant guidelines and regulations.

Contribution to the Field: The research paper could aim to contribute to the field of career prediction and machine learning by proposing novel methodologies, techniques, or insights. The goal could be to advance the current understanding of how machine learning can be effectively used for student career prediction and provide valuable recommendations for future research in this area. The objective would be to make a meaningful contribution to the academic and professional community by expanding the knowledge and understanding of the topic.

Overall, the goals and objectives of a research paper titled "Student Career Prediction using Machine Learning" would likely revolve around developing an accurate, interpretable, and practically applicable machine learning model for predicting student careers, while addressing relevant ethical considerations and making a meaningful contribution to the field.

CHAPTER – 3 DESIGN FLOW/PROCESS

3.1 Evaluation and selection of specification/features

The evaluation and selection of specification/features of a research paper titled "Student Career Prediction using Machine Learning" may involve the following:

1. Research question and objectives: Research question is Student career prediction using machine learning. Our objective is to develop is to develop a model that can predict the career of student based on the factors like previous academic results, interests, future scope of particular field, etc.

2. Literature review: The concept of student career prediction using machine learning algorithms has been around for several years. However, the exact time when this concept came into existence is difficult to determine, as it is likely that researchers and practitioners in the field of education and machine learning have been exploring this topic for some time.

3. Data collection: The primary requirement of machine learning projects is data.To earn datasets needed for these kind of projects data is collected in three ways.
a) Randomly generated data:Basically 25 percent of the data used in this project is generated randomly and this data does not fit well into algorithms as it does not follow any pattern.To get better accuracy with this type of data try to establish variable relationships among variables so that a hidden pattern is created.

b) Another way of getting the data is through linked In api.One needs to get access token from linked In and then persons person ID must be known.Then using the link https://api.linked-in.com/v2/people/(id:{person ID})
c.) The final method is to creating a google form or something and filling it by alumni or employees.

We got the complete datasets from GitHub. the datasets include various attributes to determine the best career for a student.

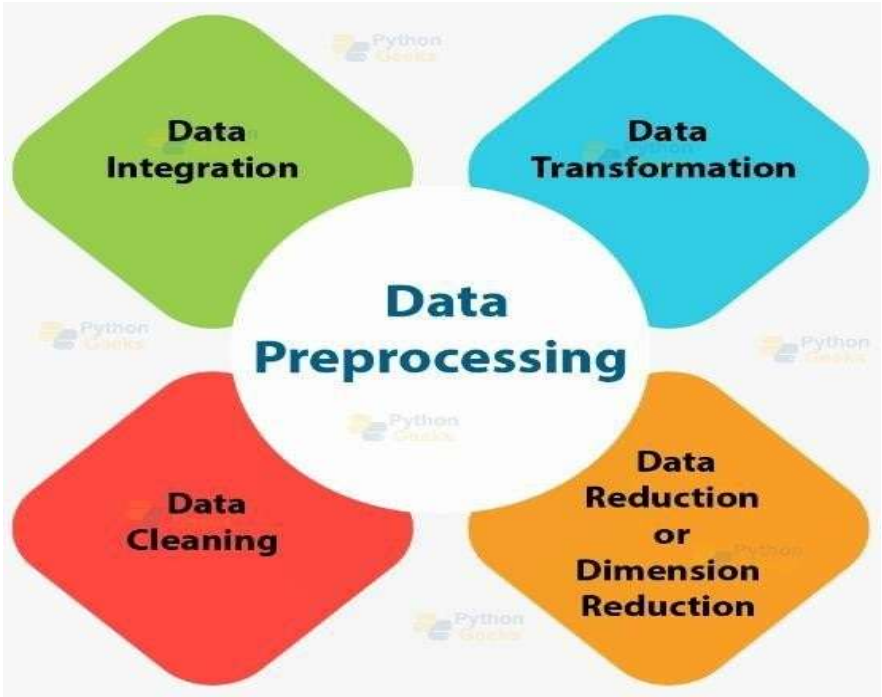
Some of the attributes are:

In [16]:
```
dataset.iloc[:,:14]
```

Out[16]:

	Acedamic percentage in Operating Systems	percentage in Algorithms	Percentage in Programming Concepts	Percentage in Software Engineering	Percentage in Computer Networks	Percentage in Electronics Subjects	Percentage in Computer Architecture	Percentage in Mathematics	Percentage in Communication skills	Hours working per day	Logical quotient rating	hackathons	co s ra
0	69	63	78	87	94	94	87	84	61	9	4	0	
1	78	62	73	60	71	70	73	84	91	12	7	1	
2	71	86	91	87	61	81	72	72	94	11	1	4	
3	76	87	60	84	89	73	62	88	69	7	1	1	
4	92	62	90	67	71	89	73	71	73	4	5	4	
5	88	86	62	79	93	84	69	71	82	11	5	3	
6	93	77	69	79	90	93	73	63	77	6	3	2	
7	84	72	88	62	66	63	78	94	60	12	2	1	
8	73	66	66	81	81	69	61	87	90	10	5	2	
9	62	76	85	91	82	69	63	63	81	10	9	0	
10	73	80	70	83	61	82	75	94	89	8	6	2	
11	63	76	81	90	70	72	86	89	85	8	3	6	
12	63	64	61	67	61	70	65	64	79	8	2	2	
13	68	93	77	60	77	81	67	81	91	12	5	3	
14	90	83	66	71	65	67	86	63	62	9	6	2	
15	94	71	63	74	61	69	65	62	68	4	7	6	
16	60	92	94	90	62	94	62	73	60	11	7	5	
17	92	86	61	63	64	65	92	65	60	4	9	6	
18	82	91	68	86	78	61	91	82	82	12	8	0	
19	90	73	90	67	63	88	67	60	69	12	6	0	
20	67	61	66	62	93	91	88	88	70	4	5	0	
21	65	61	73	90	89	74	82	60	84	5	8	3	
22	84	62	70	70	67	65	78	61	65	4	9	6	
23	74	89	66	75	86	61	63	80	69	9	9	5	
24	92	67	76	92	69	90	80	77	66	4	7	5	
25	75	74	85	93	92	80	87	78	91	12	5	0	
26	90	82	75	72	84	79	83	68	67	4	8	5	
27	71	60	93	79	78	93	90	76	75	12	9	1	
28	83	82	64	70	93	73	87	83	87	11	1	0	
29	83	68	70	90	85	75	90	92	85	4	3	1	

4. Data Pre-Processing: Preprocessing the collected data to ensure it is of high quality and suitable for use in training the machine learning algorithms.



Data Preprocessing

3.2 Design Constraint

Design constraints refer to any limitations or requirements that should be considered during the development of a model. These constraints can include technical, functional, or resource-related limitations that may impact the design, implementation, or deployment of the model. For the Student Career Prediction using machine learning model, some design constraints include:

Data Availability: The model's accuracy and efficiency heavily depend on the availability and quality of data. Limited data availability can lead to biased orinaccurate predictions. Therefore, it is essential to have a reliable source of datawith a diverse set of attributes.

Data Privacy: Student's academic data is sensitive and protected by privacy laws. Any model developed for student prediction must comply with data privacyregulations and ensure the security of student's information.

Computational Resources: Machine learning models require significant computational resources, including processing power and storage capacity. The design of the model must consider the availability of these resources to ensure efficient and timely results.

Ethical Considerations: The use of machine learning in student career prediction raises ethical considerations, including fairness, bias, and discrimination. The model should be designed to mitigate these concerns and ensure that it provides equitable and unbiased predictions.

In our model, we have applied and check for data availability, data privacy, model interpretability, computational resources, ethical considerations, and regulatory compliance. These constraints should be considered during the development, implementation, and deployment of the model to ensure that it iseffective, efficient, and complies with ethical and legal requirements.

3.2.1 Design Constraint for Student Career Prediction using ML Model

Economic:

- The development cost of the model should be reasonable and within budget.
- The model should be designed to be scalable and reusable, so that it can be deployed in different healthcare settings.
- The model should be cost-effective to maintain and update, without compromising its accuracy or performance.

Safety:

- The model should be designed with patient safety as the top priority.
- The model should be thoroughly tested and validated before deployment, to ensure it does not cause any harm or injury to patients.
- The model should be designed to detect and mitigate any potential safety risks or errors.

Professional:

- The model should be designed to enhance the work of healthcare professionals, not replace or undermine it.
- The model should be user-friendly and easy to integrate into existing healthcare workflows.
- The model should be designed to facilitate collaboration and communication between healthcare professionals and patients.

Ethical:

- The model should prioritize student's dignity, and privacy.
- The model should not violate any ethical principles or codes of conduct related to student's data.
- The model should be transparent and explainable, to ensure ethical accountability and responsibility.

Cost:

The model should be designed to be cost-effective for students.

- The model should be designed to maximize the value and benefits of student without compromising their quality or safety.
- The model should be designed to minimize unnecessary costs and expenses.

3.3 Analysis of Features and Finalization subject to constraints:

After considering the design constraints related to Economic,Safety, Professional, Ethical, and Cost, the following modifications, removals, and additions have been made to the features of the Student Career Prediction using machine learning application:

Modifications:

- The data collection process will comply with all the relevant regulations related to data privacy and confidentiality.
- The cost of the application will be optimized by using open-source machine learning libraries instead of expensive proprietary ones.
- The mobile application will have a user-friendly interface to ensure ease of use for people with limited technical expertise.

Additions:

- The application will include a feature that allows users to provide feedback and report any issues related to the application's performance or results.
- The machine learning models used for disease prediction will be regularly updated and improved based on the latest research and data.

Based on these modifications and additions, the final set of featuresfor the student career prediction using machine learning application are:

Data Collection:

- Collection of large and diverse datasets related to student's academic records thatcomply with regulations related to data privacy and confidentiality.
- Selection of relevant attributes based on the latest research and medical knowledge.

Data Pre-Processing:

- Cleaning and normalization of the data to ensure accuracy and consistency.
- Feature engineering to extract relevant information and improve the performance of the machine learning models.

Balancing of Data:

- Balancing the dataset to ensure that the machine learning models are not biased towards any particular class.

Career Prediction:

- Use of machine learning algorithms to predict the career of student .
- Regular updating and improvement of the machine learning models based on the latest research and data.

User Interface:

- Development of a user-friendly interface for the web and mobile applications.
- Feedback and reporting feature for users to provide feedback and report any issues related to the application's performance or results.

3.4 Design Flow of the Student's Career Prediction Model:

Data Collection: Gather data-sets fromsource like Git-Hub.

Data Cleaning and Pre-processing: The collected data needs to be cleaned, formatted, and pre-processed to remove any inconsistencies, errors, or missing values.

Feature Selection: Use statistical analysis and machine learning algorithms to select the most relevant features that have a significant impact on predicting student's career.

Model Development: Develop machine learning models using various algorithms such as logistic regression, decision trees, random forests, and support vector machines.

Model Evaluation: Evaluate the performance of the developed models using various metrics such as accuracy, precision, recall, F1-score, and area under the curve (AUC).

Deployment: Deploy the finalized model to the target platform, such as a web application or mobile application, for end-users.

3.5 Design Selection and Analysis

After analysing the design for the student career prediction system, the following table compares the strengths and weaknesses of each design:

Design	Strength	Weaknesses
Design 1	Uses ensemble learning for improved accuracy	Requires large amounts ofdata
Design 2	Incorporates user input for personalized recommendations	User input may be unreliable
Design 3	Utilizes deep learning for feature extraction	May require significant computing resources

3.6 Implementation Plan/ Methodology

Here is an implementation plan for the cardiac disease detection using machine learning:

Data Collection: Collect a large dataset of cardiac-related medical records and diagnostic reports from various hospitals and clinics.

Data Pre-processing: Pre-process the collected data to remove any inconsistencies, missing values, and noise. This includes data cleaning, data normalization, and feature selection.

Data Splitting: Split the pre-processed data into training, validation, and testing sets. The training set will be used to train the machine learning models, the validation set will be used to optimize the model hyperparameters, and the

testing set will be used to evaluate the final model performance.

Model Selection: Select the appropriate machine learning model for student career prediction. This can be achieved by comparing the performance of different models on the validation set using appropriate metrics such as accuracy, precision, recall, F1 score, and AUC-ROC.

Hyperparameter Tuning: Optimize the hyperparameters of the selected model to achieve the best possible performance. This can be achieved by using techniques such as grid search, random search, and Bayesian optimization.

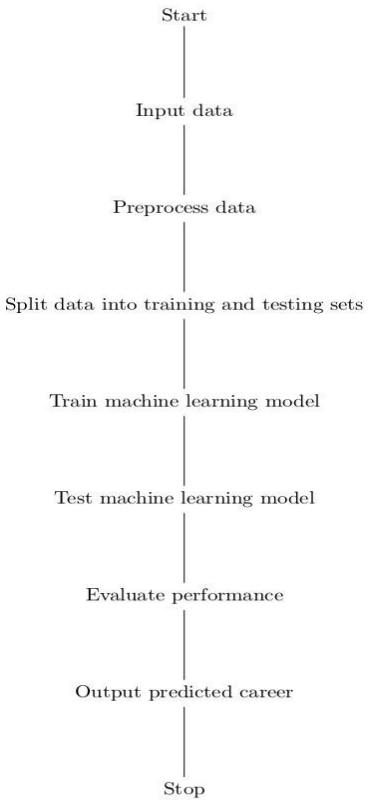
Model Training: Train the selected model on the training set using the optimized hyperparameters.

Model Evaluation: Evaluate the trained model on the testing set using appropriate metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. This will provide an estimate of the model's generalization performance on unseen data.

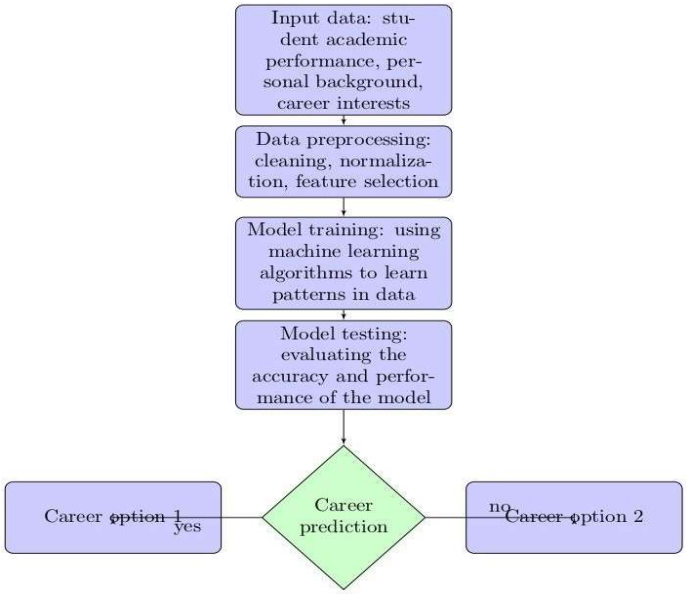
Deployment: Deploy the trained model in a production environment, such as a web application or mobile app, to enable student career prediction.

Monitoring: Monitor the performance of the deployed model and continuously update it with new data to improve its accuracy and reliability.

3.6.1 Block Diagram



A. Block diagram of student career prediction



B. Flow chart of student career prediction

CHAPTER-4

RESULT ANALYSIS AND VALIDATION

4.1 Implementation of solution

The steps below can be used to construct a machine learning (ML) solution for student career prediction:

- **Data collection:** Compile a dataset including details on students' academic accomplishments, extracurricular pursuits, personal interests, career preferences, and results (such as employment, additional education, or professional achievement). Make sure the dataset is complete and includes a wide variety of students.

- **Data Preprocessing:** Preprocessing the data involves addressing missing values, eliminating superfluous features, and, if necessary, undertaking feature engineering. To achieve uniform scales, transform categorical variables into numerical representations (using a technique like one-hot encoding) and normalise numerical variables.

- **Feature Selection:** The most pertinent characteristics that have the greatest impact on forecasting students' career success should be chosen as features. To choose the best subset of features, use strategies like correlation analysis, feature importance, or recursive feature elimination.

- **Model selection:** Select the best machine learning algorithm for predicting careers. It could be a classification algorithm like support vector machines, decision trees, random forests, or logistic regression. Alternately, you can experiment with more complex algorithms like ensemble approaches, neural networks, and gradient boosting.

- **Model Training:** Divide the dataset into training and testing sets for the model. To train the ML model using the chosen algorithm, use the training set. To improve the performance of the model, alter hyperparameters like learning rate, regularisation, or tree depth using grid search or cross-validation methods.

- **Model Evaluation:** Use the testing set to assess the training model. In order to evaluate how successfully the model foretells student career outcomes, performance metrics like accuracy, precision, recall, F1 score, or area under the ROC curve can be computed.

- **Result Analysis:** Analyse the model's performance and predictions to learn more about the variables affecting students' career decisions. Determine trends, correlations, and patterns between various characteristics and professional outcomes. Students, educators, and policymakers can benefit from the knowledge provided by this analysis to help them make wise choices.

- **Validation and Iteration:** Verify the model's predictions against the actual career paths taken by freshmen students. Keep track of the model's performance over time, analyse it, and make any adjustments to the solution. To enhance the precision and usability of the career prediction system, think about getting input from students and stakeholders.

- **Model Evaluation:** Utilising the testing dataset, evaluate the trained model's performance. The model's capacity to forecast the correct career category for each student should be measured using the relevant assessment measures, such as accuracy, precision, recall, or F1 score. This process enables you to comprehend the model's advantages and disadvantages and pinpoint potential areas for development.

- **Model Deployment:** Deploy the model in a setting where it may be used to forecast the career outcomes for incoming students after you are pleased with its performance. This can be accomplished through a web application, an API, or any other tool that enables users to enter pertinent student data and generate forecasts.

- **Continuous Improvement:** Track the effectiveness of the deployed model over time and gather user input. If required, iterate and enhance the model using this feedback. To maintain the model accurate and up to date, you might need to frequently add fresh data.

The relevancy and quality of the data gathered, the model selection, the feature selection, and the process of continual improvement all affect how effective the solution is. To improve the solution's accuracy and utility, it is critical to iterate and develop the approach based on feedback from the real world.

CHAPTER-5

CONCLUSION AND FUTURE WORK

5.1. Conclusion

This research article has examined how machine learning (ML) methods can be used to forecast student career outcomes. Through the examination of multiple data points, ML algorithms have demonstrated potential for offering insightful knowledge and direction to learners, educational institutions, and policymakers.

The results of this study demonstrate the potential advantages of machine learning-based career prediction models. These models provide students with individualised recommendations that assist them in choosing their educational and future career routes. ML algorithms can find job routes that complement students' talents and interests by taking into account elements like academic performance, extracurricular activities, personal hobbies, and demographic data.

Additionally, ML models can help educational institutions create supportive and focused interventions. These models enable early interventions to keep students on track and support them in attaining their career objectives by identifying students who may need more help, such as those at risk of dropping out or suffering academically.

The drawbacks and difficulties of ML-based career prediction should be understood, though. The calibre and variety of the available data have a significant impact on the reliability and accuracy of predictions. It can be quite difficult to obtain complete and representative data sets that capture the wide range of factors impacting job choices.

Additionally, it is critical to see ML models as tools that support decisions rather than as final judgements. For a well-rounded approach to career planning, human skills such as career counselling and guidance should coexist alongside ML forecasts.

This study shows how ML has the potential to revolutionise educational and career preparation. Educational institutions can offer individualised recommendations, early interventions, and focused support by utilising ML approaches, which will ultimately improve student outcomes and make the educational system more effective. Addressing data constraints, improving prediction models, and examining the long-term effects of ML-based career advising on students' professional success and satisfaction should be the main objectives of future research.

5.2. Future work

Here are some ideas for potential follow-up research for your study on utilising machine learning to forecast student careers:

- **Add new data sources:** To increase the predictability of the results, think about include more data sources. You might include information from social media, academic performance, or demographics, for instance.

- **Try out alternative machine learning algorithms:** While your present strategy might work, you could also try out a variety of other algorithms to see if you can increase the accuracy of your predictions. You could test out decision trees, neural networks, or support vector machines, for instance.

- **Discover feature engineering:** To increase the precision of the predictions, feature engineering entails choosing and extracting pertinent features from the data. To possibly increase the accuracy of your model, think about experimenting with various feature engineering strategies.

- **Examine the model's performance over time:** After your model has been trained, it is crucial to keep track of it in order to make sure it keeps making reliable predictions. Think about putting in place a mechanism that periodically assesses the model's performance and retrains it as necessary.

- **Examine the model's bias:** Machine learning models can exhibit bias as a result of the training data they receive. Consider checking to see whether your model contains any bias and looking at ways to lessen it.

- **Extend the model to other domains:** Despite the fact that your current model is mostly concerned with predicting student careers, you might want to think about expanding it to other fields. Similar methods might be used, for instance, to forecast stock prices, employee attrition, and consumer behaviour.

References: -

[1] Min Nie , ZhaohuiXiong , RuiyangZhong , Wei Deng and Guowu Yang (2020), " Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students" , Journal of Appl. Sci. pp. 1-14.

[2] K. B. Eashwar, R. Venkatesan and D. Ganesh (2017), "Student Performance Prediction Using SVM", International Journal of Mechanical Engineering and Technology (IJMET) Vol. 8, No. 11, pp. 649–662.

[3] AnkitaKadambande, Snehal Thakur, AkshataMoholand A.M.Ingole (2017), "Predict Student Performance by Utilizing Data Mining Technique and Support Vector Machine", International Research Journal of Engineering and Technology (IRJET), Vol. 04, No. 05, pp 2818-2821.

[4] S.A. Oloruntoba1 and J.L.Akinode (2017), "Student Academic Performance Prediction Using Support Vector Machine", International Journal Of Engineering Sciences & Research Technology, pp. 588-598.

[5] AhmedSharafElDen ,Malaka A. Moustafa, Hany M. Harb and AbdelH.Emara (2013), "Adaboost Ensemble With Simple Genetic Algorithm For Student Prediction Model", International Journal of Computer Science & Information Technology (IJCSIT) , Vol. 5, No 2, pp 73-85.

[6] MeimeiHan ,Mingwen Tong , Mengyuan Chen , Jiamin Liu and Chunmiao Liu (2017), "Application of Ensemble Algorithm in Students' Performance Prediction", IEEE 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI).

[7] V. Prasannakumari B. Bavani, S. Nirmala SugirthaRajini, M.S. Josephine (2019), "Heart Disease Prediction System based on Decision Tree Classifier", Jour of Adv Research in Dynamical & Control Systems, Vol. 11, 10-Special Issue, 2019, Volume 11, Issue 10.

[8] FaridJauhari and Ahmad AfifSupianto (2019), "Building student's performance decision tree classifier using boosting algorithm", Indonesian Journal of Electrical Engineering and Computer Science Vol. 14, No. 3, June 2019, pp. 1298 - 1304.