Career Guidance Recommendation System using Machine Learning

Md. Abdul Hakim

ID: 20103281

Md Shahriair Alam

ID: 20103204

A Thesis in the Partial Fulfillment of the Requirements

for the Award of Bachelor of Computer Science and Engineering (BCSE)

Department of Computer Science and Engineering

College of Engineering and Technology

IUBAT – International University of Business Agriculture and Technology

Summer 2023

Career Guidance Recommendation System using Machine Learning

Md. Abdul Hakim (20103281) and Md Shahriair Alam (20103204)

A Thesis in the Partial Fulfillment of the Requirements for the Award of Bachelor of Computer Science and Engineering (BCSE)

The thesis has been examined and approved,

_____

Prof Dr. UtpalKanti Das
Chairman & Professor
Dept. of Computer Science and Engineering
IUBAT – International University of Business Agriculture and Technology

_____

Dr. Hasibur Rashid Chayon
Co-supervisor, Coordinator and Associate Professor
Dept. of Computer Science and Engineering
IUBAT – International University of Business Agriculture and Technology

_____

Md. Mahmudul Hasan
Supervisor and Lecturer
Dept. of Computer Science and Engineering
IUBAT – International University of Business Agriculture and Technology

Summer - 2023

Letter of Transmittal

2 September 2023
The Chair
Thesis Defense Committee
Department of Computer Science and Engineering
IUBAT–International University of Business Agriculture and Technology
4 Embankment Drive Road, Sector 10, Uttara Model Town
Dhaka 1230, Bangladesh

Subject: Letter of Transmittal.

Dear Sir,

,

With due respect, we would like to inform you that it is a great pleasure for us to submit this report entitled **"Career Guidance Recommendation System using Machine Learning"** to complete our research.
It was a great opportunity for us to work on this research to make our theoretical knowledge more realistic and we gained a lot of exposure to the field of machine learning. We are now look forward to your kind commentary on this performance report.

We will always be very grateful to you if you kindly go through this report and check our performance.

Sincerely Yours,

Md. Abdul Hakim	Md Shahriair Alam
20103281	20103204

Student's Declaration

I am Md. Abdul Hakim and my partner Md Shahriair Alam, are student of the BCSE-Bachelor of Computer Science and Engineering program, under the College of Engineering and Technology (CEAT) of the International University of Business Agriculture and Technology (IUBAT) announcing this report entitled **"Career Guidance Recommendation System using Machine Learning"** has been prepared for the completion of the thesis course, which is part of the Bachelor of Computer Science and engineering degree.

The report of **"Career Guidance Recommendation System using Machine Learning"** was edited by me and my partner. All modules and procedures for this research are done after proper testing and online information.

It is not designed for other purposes, awards or presentations.

_____	_____
Md. Abdul Hakim	Md Shahriair Alam
20103281	20103204

Supervisor's Certification

This is to ensure that the thesis report on the **"Career Guidance Recommendation System using Machine Learning"** is compiled by Md. Abdul Hakim with ID#20103281 and Md Shahriair Alam, with ID #20103204, of IUBAT– International University of Business Agriculture and Technology, as part of the fulfillment of the required part of an effective thesis course. The report has been prepared under my supervision and is a record of the work accomplished, successfully completed. To the best of my knowledge and as per there declaration, no portions of this report have been posted anywhere by any degree, diploma or certificate.

You are now allowed to submit a report. I wish your every success in the future endeavors.

Thesis Supervisor

_____

Md. Mahmudul Hasan

Lecturer

Department of Computer Science and Engineering

IUBAT–International University of Business Agriculture and Technology

Department's Certification

On behalf of the Department of Computer Science and Engineering, IUBAT-International University of Business Agriculture and Technology, I undersigned, confirm the performance report on '**Career Guidance Recommendation System using Machine Learning**' for Bachelor of Computer Science and Engineering (BCSE) degrees was duly presented by Md. Abdul Hakim (ID No: 20103281) and Md Shahriair Alam (ID No: 20103204) and approved by the department.

_____

Md. Mahmudul Hasan
Lecturer
Department of Computer Science and Engineering
IUBAT- International University of Business Agriculture and Technology

_____

Dr. Hasibur Rashid Chayon
Coordinator and Associate Professor
Department of Computer Science and Engineering
IUBAT- International University of Business Agriculture and Technology

_____

Prof. Dr. Utpal Kanti Das
Chairman & Professor
Department of Computer Science and Engineering
IUBAT- International University of Business Agriculture and Technology

Abstract

The Career Guidance Recommendation System using Machine Learning is a cutting-edge platform designed to assist individuals in making informed career choices. Most students around the world are in constant confusion after completing their upper secondary education. The stage at which they have to decide on an appropriate career path. Precisely, students don't have enough maturity to understand what individuals must follow in order to determine compatible career paths, while working under the stage. With the rapidly changing job market and increasing competition for jobs, individuals are finding it more challenging to make informed decisions about their career paths. The system utilizes machine learning algorithms to analyze user data and provide personalized career recommendations based on the individual's interests, skills, values, and career goals. The system collects data on the user through a series of assessments and surveys, which are then processed by machine learning algorithms to generate a unique profile for each user. This profile is then used to match the user with a list of careers that best align with their profile. The system's machine learning algorithms are trained on large datasets of career information and job market trends, allowing for highly accurate and up-to-date recommendations. Results from the analysis of user interests, skills and values this recommendation system provides a user-friendly interface that allows individuals to explore and compare different career options, and it provides access to detailed information about each career, including job descriptions, salaries, and educational requirements. Finally, our research can be of real help not only for students but also for the educational institution and other enterprises.

Acknowledgments

Firstly, We want to give Almighty Allah to our sincere gratitude for His wonderful grace, which enabled us to successfully finish the thesis.

We really appreciate and express our gratitude to our thesis supervisor Md. Mahmudul Hasan, lecturer department of the Computer Science and Engineering program at International University of Business Agriculture and Technology. He has a strong background in "Machine Learning" and a genuine interest in it. This thesis was made possible by his never-ending patience, academic leadership, valuable advice, constant and energetic supervision, constructive criticism, insightful counsel and fixing them at all levels.

We would like to extend our sincere gratitude to Dr. Hasibur Rashid Chayon, Coordinator and Associate Professor of the CSE Department of International University of Business Agriculture and Technology for their kind assistance in completing our research.

We would like to extend our gratitude to every classmates of International University of Business Agriculture and Technology who participated in this discussion while also attending class.

Finally, we must respectfully appreciate to our parents' unwavering help and endurance.

Table of Contents

List of Figures

List of Tables

Chapter I. Introduction

A career guidance recommendation system using machine learning is a tool that combines artificial intelligence and big data analytics to provide individuals with personalized recommendations for their future careers. The system utilizes algorithms and models trained on large datasets of job market information, job requirements, and individual profiles to generate tailored recommendations for each individual user. To develop a career guidance recommendation system that provides personalized recommendations to individuals based on their skills, interests, values, and aspirations, and helps them make informed decisions about their future careers. The system works by first collecting data on an individual's skills, interests, values, and aspirations through a series of assessments or surveys. This information is then used to create a unique profile for each individual user. The system can also take into consideration other factors such as location, industry, and desired work-life balance to provide even more tailored recommendations. The machine learning algorithms used in the career guidance recommendation system can be supervised or unsupervised, depending on the type of data being analyzed and the goals of the system. For example, decision trees and random forests can be used for supervised learning, where the system is trained on a labeled dataset to predict a target variable. On the other hand, unsupervised learning algorithms such as k-means clustering and principal component analysis can be used to uncover patterns and relationships in the data that may not be immediately obvious. The system can also utilize deep learning techniques such as artificial neural networks to generate more accurate recommendations. It can provide individuals with highly personalized recommendations based on their unique profiles, helping them make informed decisions about their future careers. The recommendation system should

consider an individual's education, work experience, skills, interests, values, and aspirations and use machine learning algorithms to generate a list of suitable career options. It should also provide detailed information about the recommended careers, including job descriptions, salary ranges, and growth prospects, to help individuals make informed decisions about their future careers. Additionally, the system can provide valuable insights into the job market and help individuals make better-informed career decisions, leading to increase job satisfaction and success.

1.1 Context and Background

Career guidance recommendation systems use data analysis and algorithms to provide personalized career recommendations based on an individual's skills, interests, and preferences. They aim to address the challenges individuals face in identifying the best-suited career paths due to the rapidly changing job market and difficulty in self-assessment. Advancements in artificial intelligence and machine learning have made these systems increasingly powerful and have contributed to a growing interest in their development. Overall, career guidance recommendation systems are valuable tools for individuals seeking to navigate the complex world of work.

1.2 Problem Statement

The problem statement for career guidance recommendation systems is that individuals face challenges in identifying the most suitable career paths due to the rapidly changing job market and difficulty in accurately assessing their own strengths and weaknesses. This can lead to career dissatisfaction, underemployment, and a lack of fulfillment. Traditional career guidance methods may not provide personalized recommendations that account for an individual's unique profile. Therefore, there is a need

for more effective career guidance tools that utilize data analysis and algorithms to provide highly personalized career recommendations based on an individual's skills, interests, and preferences.

1.3 Scope

The scope of career guidance recommendation systems is broad and includes students, job seekers, and those considering a career change. These systems provide personalized recommendations for a wide range of careers and industries based on an individual's unique profile, such as educational background, work experience, personality traits, and interests. Educational institutions, career centers, and employers can also use these systems to guide their students, clients, and employees. As these systems evolve, they have the potential to revolutionize career decision-making and help individuals navigate the complex world of work more effectively.

1.4 Research Objectives

The research objectives of career guidance recommendation systems may vary depending on the specific context and goals of the study. However, common objectives may include:

- To investigate the effectiveness of career guidance recommendation systems in helping individuals make informed career decisions.
- To identify the key factors this will devolve personalized career recommendations.
- To evaluate the accuracy and reliability and uses of the algorithms used in these systems.

- To assess the user experience and usability of the recommendation systems.

To understand the impact of career guidance recommendation systems on individuals' career satisfaction and success.

- To explore the potential applications of these systems in educational institutions, career centers, and employers.

Overall, the research objectives of career guidance recommendation systems aim to enhance our understanding of the role these systems can play in helping individuals navigate the complex world of work and make informed career decisions.

1.5 Significance of the Research

Research on career guidance recommendation systems is highly significant due to its potential to revolutionize the way individuals make career decisions. These systems provide personalized recommendations based on an individual's unique profile, such as their educational background, work experience, personality traits, and interests. By doing so, they can help individuals make more informed and confident career decisions, leading to greater career satisfaction and success.

As the job market continues to evolve, there is a growing need for workers with specific skills. Career guidance recommendation systems can help individuals identify the skills needed for success in different industries, thereby helping to address the skills gap. They can also help to streamline the career decision-making process, potentially saving individuals and organizations time and resources.

1.6 Research Questions

Research questions for career guidance recommendation systems may change for the specific context and goals of the study. However, some potential research questions could include:

- How effective are career guidance recommendation systems in helping individuals make informed career decisions?
- What factors should be considered when developing personalized career recommendations for individuals?
- How can algorithms be improved to increase the accuracy, uses and reliability of career guidance recommendation systems?
- What is the user experience of career guidance recommendation systems, and how can they be made more user-friendly?

Chapter II. Literature Review

An Intelligent Career Guidance System using Machine Learning refers to a computer-based platform that uses machine learning algorithms to provide personalized career guidance to individuals. The system analyzes the user's educational background, skills, interests, and other relevant factors to suggest suitable career paths and recommend appropriate courses or training programs. By analyzing a user's historical data, such as education, work experience, and personal interests, machine learning algorithms can generate insights and recommendations for career paths that match their strengths and preferences. One potential benefit of an intelligent career guidance system is that it could help individuals navigate the complexities of the modern job market, which can be overwhelming due to the proliferation of different job titles and descriptions.

2.1 Literature Review

Vignesh et al., (2021) developed an intelligent career guidance system using machine learning algorithms. The system analyzed various factors such as personality traits, work preferences, and learning styles to provide personalized career guidance to individuals. The system used a combination of decision trees and logistic regression algorithms to provide accurate and relevant guidance. The study found that the system was highly effective in providing personalized career guidance to individuals, resulting in better career prospects and increased job satisfaction. An Intelligent Career Guidance System using Machine Learning proposes a system that utilizes machine learning techniques to provide intelligent career guidance to students. The paper has made a significant contribution to the field of career guidance by proposing a novel approach that can provide personalized and accurate career guidance to students. The paper has provided a detailed description of the proposed

system, including the different modules and algorithms used. The methodology used for the development and evaluation of the system is also well-defined, with the authors using various machine learning techniques such as decision trees, K-nearest neighbor, and Naïve Bayes algorithms. The study has several strengths, including its focus on providing personalized career guidance, the use of a comprehensive dataset, and the integration of different machine-learning techniques.

In 2018, Sripath et al. presented a study with the objective of creating a system that utilizes machine learning techniques to anticipate the optimal career path for a student. This system considers several factors, including academic performance, personality traits, interests, and other relevant information. The authors emphasize the importance of collecting high-quality data to ensure the accuracy and efficiency of the algorithms used. The collected data includes various parameters such as academic scores, programming and analytical capabilities, personal details, and interests. The authors explain the process of OneHot Encoding and how it transforms categorical values into a form suitable for machine learning algorithms (SVM, XG Boost and Decision Tree) making the representation of categorical variables more expressive. When inputs are not directly provided but rather student parameters are determined by evaluating students using various assessments and examinations, a more sophisticated web application can be created. The design of technical, analytical, logical, memory-based, psychometric, general awareness, interest-based, and skill-based tests and the collection of parameters through them can be done to ensure that the results are unquestionably accurate and that the system is more dependable to use. The authors collected data from sources including employee records, LinkedIn API, and alumni databases. They used various machine learning techniques such as OneHot Encoding, Random Forest, and Support Vector Machine to process and analyze the data. The expected results showed that the Random Forest model which outperformed other models in

predicting the career paths of students.

Career Guidance using Machine Learning (Metha, Kamble&Vijay,2021)which indicates that the research aims to develop a system that uses machine learning algorithms to provide career guidance to users. The system could be designed for students, job seekers, or individuals looking to switch careers. The development process for a career guidance system using machine learning involves creating a dataset from different sources, splitting the data into training and test sets, and performing pre-processing activities for accurate output. The system uses the methods supervised machine learning, where the data is organized and used to learn the model to expect accurate results for career guidance. Supervised machine learning predicts the target output value from the input features, and the features used in the model are similar to the previous input features. This technique is appropriate for students who require career counseling for HSC Class, however it does not cover other academic topics. This system is still being worked on. The other two options are suitable for higher education and international study; but, in order to access their full service, students must purchase a membership.The paper lacks detailed information on the data collection and pre-processing techniques used, which could be important factors affecting the accuracy of the predictions. The evaluation of the proposed system has been done using accuracy as the performance metric. However, accuracy alone might not be sufficient to evaluate the system's performance, and other metrics such as precision, recall, and F1 score could have been used to provide a better evaluation. Additionally, the authors have not compared their system's performance with existing state-of-the-art systems in the field.

An Intelligent Career Guidance System using Machine Learning is a computer-based platform that provides personalized career guidance to individuals. The "Intelligent Career Guidance System using Machine Learning" (Dahankeet al., 2022) paper discusses the

primary and core module of the system, the prediction module, which is built with various technologies such as machine learning algorithms, APIs, and datasets to train the model. The K-Nearest Neighbors algorithm is used for classification purposes, and K-Means Clustering is used for clustering purposes. The Flask API is used for communication between the front-end and back-end. The dataset used for the machine learning model is developed manually, containing seven different features related to core and sub-skills. The machine learning model uses multi-class classification techniques, and 80% of the dataset is used for training, while 20% is used for validation and testing purposes.

One limitation of the paper is that the dataset used for training the machine learning model was developed manually as there was no appropriate data available related to the core concept of the application. Additionally, the paper lacks discussion on the generalizability of the system to different populations and cultures, which could limit its effectiveness in providing career guidance to individuals from diverse backgrounds. The paper discusses the use of K-Nearest Neighbors for classification and K-Means Clustering for clustering purposes, along with Flask API for communication between the front end and back end. The paper presents a detailed description of the methodology used and the results obtained from the system. However, the paper lacks in-depth discussion of the limitations and future scope of the system, as well as the ethical considerations that arise with the use of machine learning algorithms for career guidance. Overall, while the paper presents an interesting application of machine learning algorithms, further work is needed to address the ethical concerns.

AI-based Career Guidance refers to the use of Artificial Intelligence (AI) technologies to provide personalized and automated career guidance to individuals. This technology aims to assist individuals in making informed decisions about their career paths by analyzing their skills, interests, and personality traits.

AI-based Career Guidance (Firdoshet al., 2020) indicates the potential of artificial intelligence (AI) in providing career guidance to individuals. The authors highlight the importance of career guidance in today's rapidly changing job market and how traditional methods of career guidance are not always effective. The paper presents a system that uses AI to provide personalized career guidance to individuals. The system uses natural language processing (NLP) to analyze the user's responses to a series of questions about their interests, skills, and education. The system then recommends career options that match the user's profile. The authors argue that AI-based career guidance can be more effective than traditional methods because it can provide personalized recommendations based on the user's unique profile. They also discuss the potential limitations of the system, such as the quality of the data used to train the AI algorithms and the need for ongoing updates to ensure the system stays current with changing job market trends.

Chapter III. Research Methodology

3.1 Recap of the research questions

The research question focuses on the development of an intelligent career guidance system that utilizes machine learning algorithms to provide accurate and personalized career advice to users. This research question aims to investigate the various machine learning algorithms that can be used to analyze the user's personal and professional characteristics, identify their skills, strengths, weaknesses, and interests, and suggest suitable career paths based on their profile. This question requires exploring the existing literature on the various machine-learning techniques and determining which algorithms are most effective in predicting and recommending career options. Then another research question focuses on the design and development of a user-friendly interface for the intelligent career guidance system. This question requires investigating the various design principles and guidelines for developing a user interface that is intuitive, easy to use, and aesthetically pleasing. The research will also investigate the usability of the system and gather feedback from users to improve the interface and overall user experience.

Another research question aims to evaluate the effectiveness of the intelligent career guidance system in terms of its accuracy, reliability, and usefulness. This research question requires conducting a series of tests to measure the system's ability to accurately predict and recommend suitable career paths for users. The research will also investigate the reliability of the system by comparing its recommendations to those of human career advisors. The usefulness of the system will be evaluated by gathering feedback from users on the quality and relevance of the career advice provided by the system.The next research question focuses on the ethical implications of using machine learning algorithms in career guidance. This research question requires investigating the potential ethical concerns that

may arise from using machine learning algorithms to predict and recommend career paths. The research will investigate issues such as bias, privacy, and transparency and develop ethical guidelines for the use of machine learning algorithms in career guidance.

Overall, the research questions in this thesis topic aim to investigate the development of an intelligent career guidance system using machine learning algorithms, focusing on the design and development of a user-friendly interface, evaluating the effectiveness of the system, and investigating the ethical implications of using machine learning algorithms in career guidance. By addressing these research questions, the thesis will contribute to the development of a more accurate, personalized, and ethical approach to career guidance, which can help individuals make informed decisions about their careers and achieve their professional goals.

3.2 Description of the method

Selecting the right algorithm for implementing machine learning can be a daunting task. It's essential to preprocess data before using algorithms to avoid bad results. In a study, two algorithms were used to recommend careers for students, and the k-nearest neighbor was found to be the most accurate. Data preparation and preprocessing are crucial steps in the proposed methodology. After preprocessing, algorithms were applied, and the results were compared to determine the best algorithm. Therefore, selecting the right algorithm and preprocessing data can significantly improve the accuracy of machine learning outcomes.

In figure 3.2.1 the proper steps of the proposed methodology are encapsulated below:

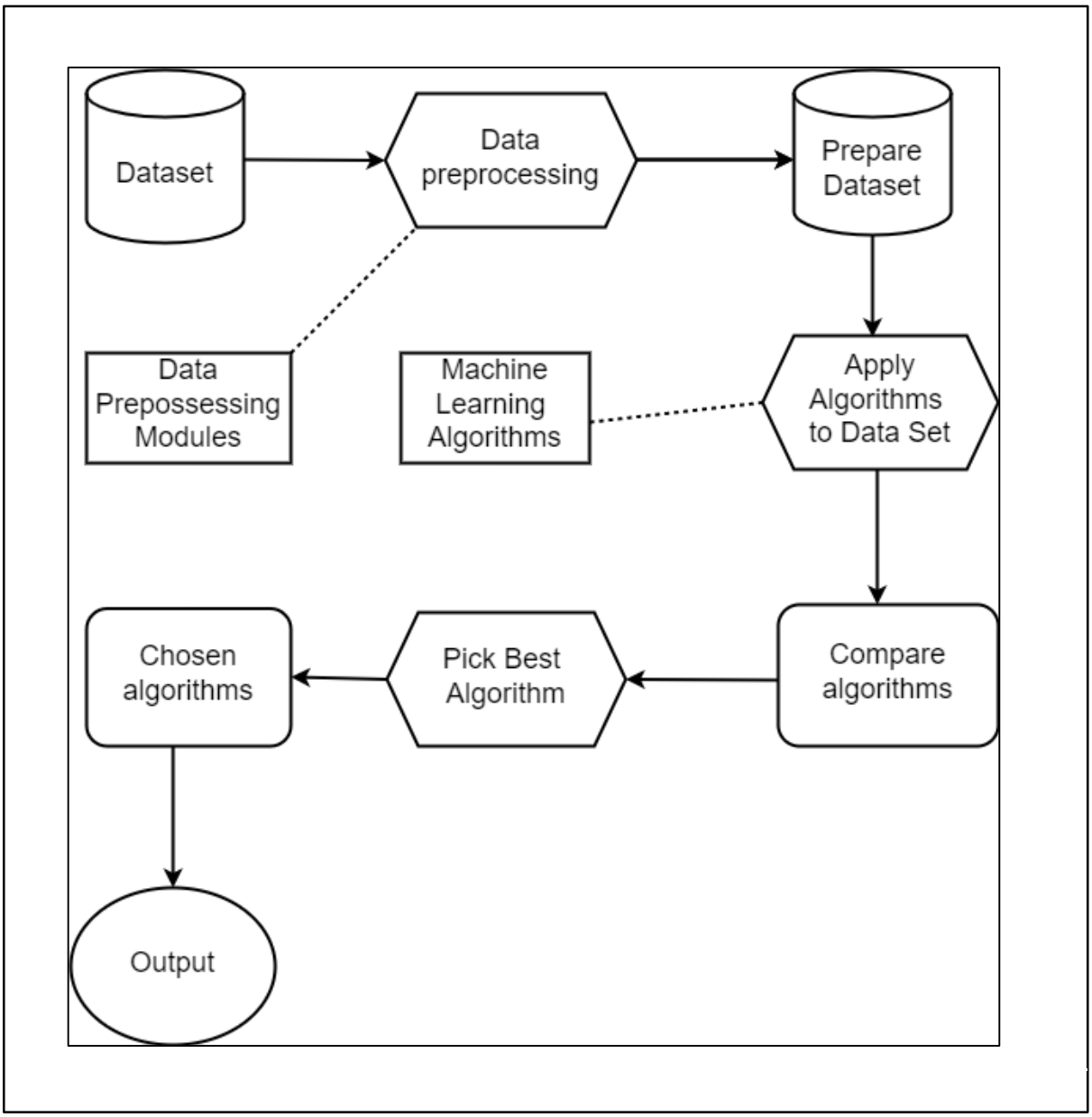


Figure: 3.2.1 Proposed methodology of work

3.3 Dataset Description

Due to difficulties with collecting datasets, we resorted to using dummy datasets that were specifically created from a Bangladeshi career perspective. Although dummy data does not contain any valuable information, it can be used to reserve space for real data and

for both operational and testing purposes. We used dummy datasets for testing purposes by creating questions related to students' interests, academic backgrounds, and environmental situation. By gathering the answers to these questions, we can predict the appropriate field of study for a specific student. Various fields of study were included so that students could choose their own area of study by answering questions to a machine, which could also determine their future job opportunities based on their interests chart of the data reveals that we included ten distinct areas of study in our dummy dataset to obtain the recommended output for students.

In figure 3.3.1 the illustration of our dataset has been given below:

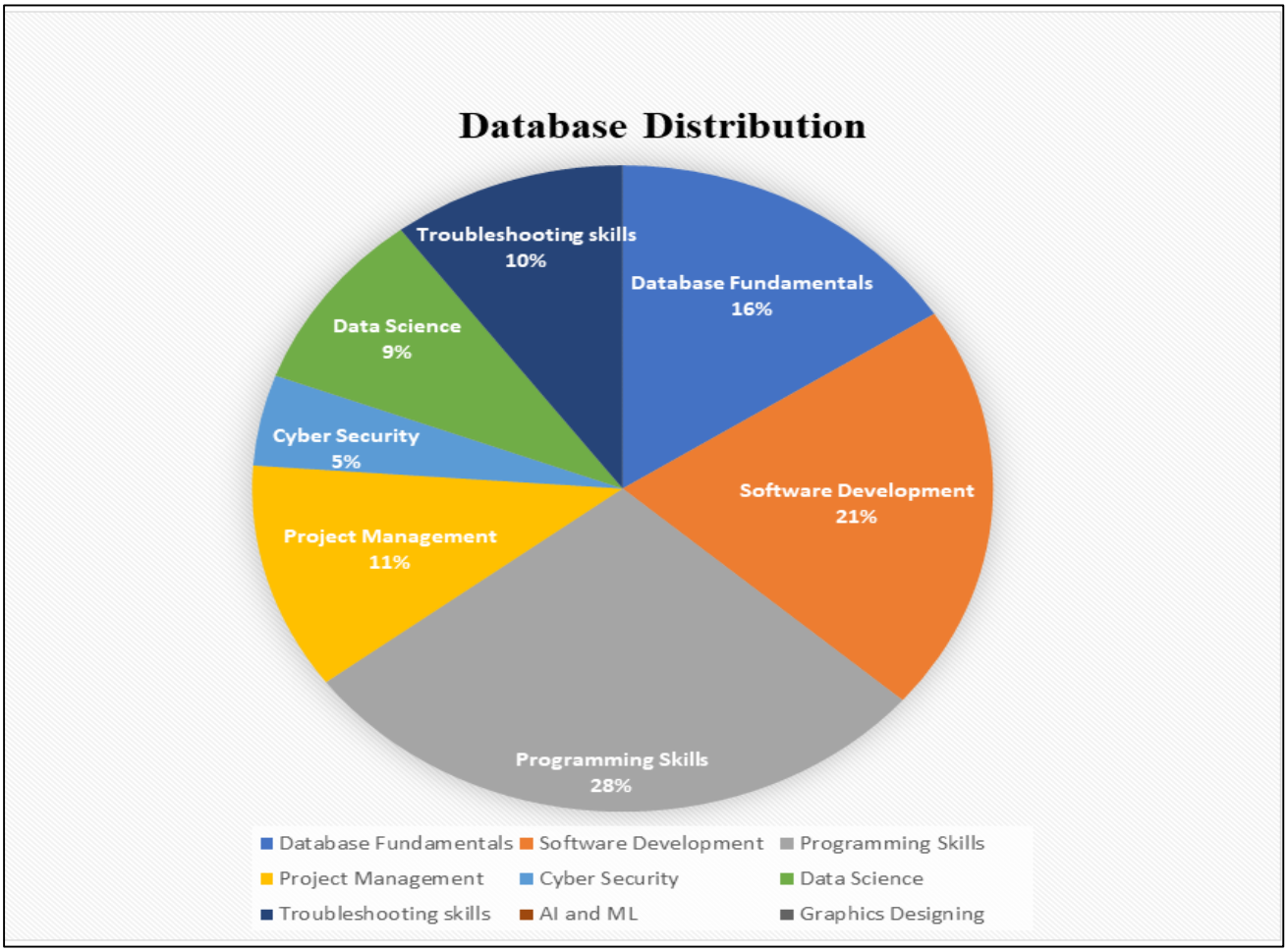


FIGURE: 3.2.2 VARIOUS DISCIPLINES OF DATA

3.3.1 Data Statistics

Statistics of my Dataset –

TABLE 3.3.1: DATA STATISTICS

Names	Numbers
Question of skillsets	10
Categories	10
Students	More than 1000

Figure 3.3.1: Various Disciplines of Data

The table indicates that the questions and categories of skillsets pertain specifically to computer skills and that there are more than 1000 students involved in this scenario. This suggests that the questions and categories of skillsets are likely related to computer programming, IT, or another field of technology and that the students in question are likely studying or training in this area.

3.4 Data preprocessing

Since we utilized dummy data, we needed to preprocess the data to allow our model to perform well. To accomplish this, we followed a few steps. Initially, we performed one hot encoding, which involved encoding categorical features as a one-dimensional array of numbers. This step was necessary for providing categorical data to many scikit-learn estimators and applying datasets to linear models or other machine-learning models. Next, we removed any extraneous data using the panda's library. Once this was completed, we

divided our input and output variables into separate columns to prepare for further training in machine learning algorithms.

3.4.1 OneHot Encoding:

To improve prediction accuracy in machine learning, it is common practice to convert categorical information into a format that can be utilized by machine learning algorithms. One-hot encoding is a popular method for achieving this conversion. This involves transforming categorical variables during pre-processing since machine learning models require numeric input variables.

3.5 Machine Learning Algorithms

The way in which an AI system performs its task, typically making predictions of output values based on input data, is referred to as a machine learning algorithm.

3.5.1 Naïve Bayes

Naïve Bayes is a classification method that relies on Bayes' Theorem, assuming that the predictors are independent. Essentially, the Naive Bayes classifier assumes that the existence of a particular feature in a category is not connected to the existence of any other feature. The Naive Bayes algorithm is adaptable for various applications, including real-time predictions, multi-class predictions, text classification, sentiment analysis, and recommendation systems.

Bayes' theorem offers a method for computing the posterior probability P(A|B) using P(A),

P(B), and P(B|A).

The formula:

$$P(\frac{A}{B}) = \frac{P\left(\frac{B}{A}\right).P(A)}{P(B)}$$

Where -

P(A/B) = Probability of A occurring given evidence B has already occurred

P(B/A) =Probability of B occurring given evidence A has already occurred

P(A)=Probability of A occurring

P(B)=Probability of A occurring

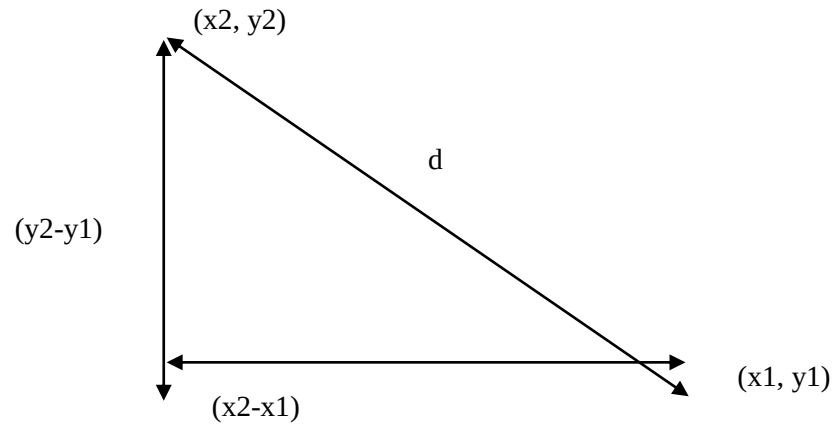
3.5.2 K-NN algorithm

Given that we are developing a recommendation system, the k-NN algorithm is a suitable option for our dataset.The k-nearest neighbor (k-NN) algorithm is a supervised machine learning technique that can be used for both classification and regression tasks, although it is mainly used for classification in the industry. It works by calculating the distance between data points, which is done using the Euclidean Distance formula.

Formula:

The Euclidean Distance formula can be used to calculate the distance between two points in 2-dimensional space. This formula is commonly known as the Euclidean distance formula and is widely used for measuring distances in the plane.

If we consider two points (x1, y1) and (x2, y2), the distance between them can be calculated using the Pythagorean Theorem as follows:



The Euclidean distance between (x1, y1) and (x2, y2) is,d = √ ((x2 - x1) ^2 + (y2 - y1) ^2)

So, in short form,

$$d(x,y) = \sqrt{\sum_{i=1}^{n} \quad (i - yi)\text{^}2}$$

.5.3 Fuzzy Logic

Fuzzy Logic (FL) is a reasoning technique that mimics human reasoning. FL follows the human decision-making approach, which considers all the possible intermediate options between digital values YES and NO. On the other hand, computers use a conventional logic block that accepts precise input and produces an output that is either TRUE or FALSE, similar to humans' YES or NO.

The inventor of fuzzy logic, Lotfi Zadeh, observed that unlike computers, the human decision making includes a range of possibilities between YES and NO,

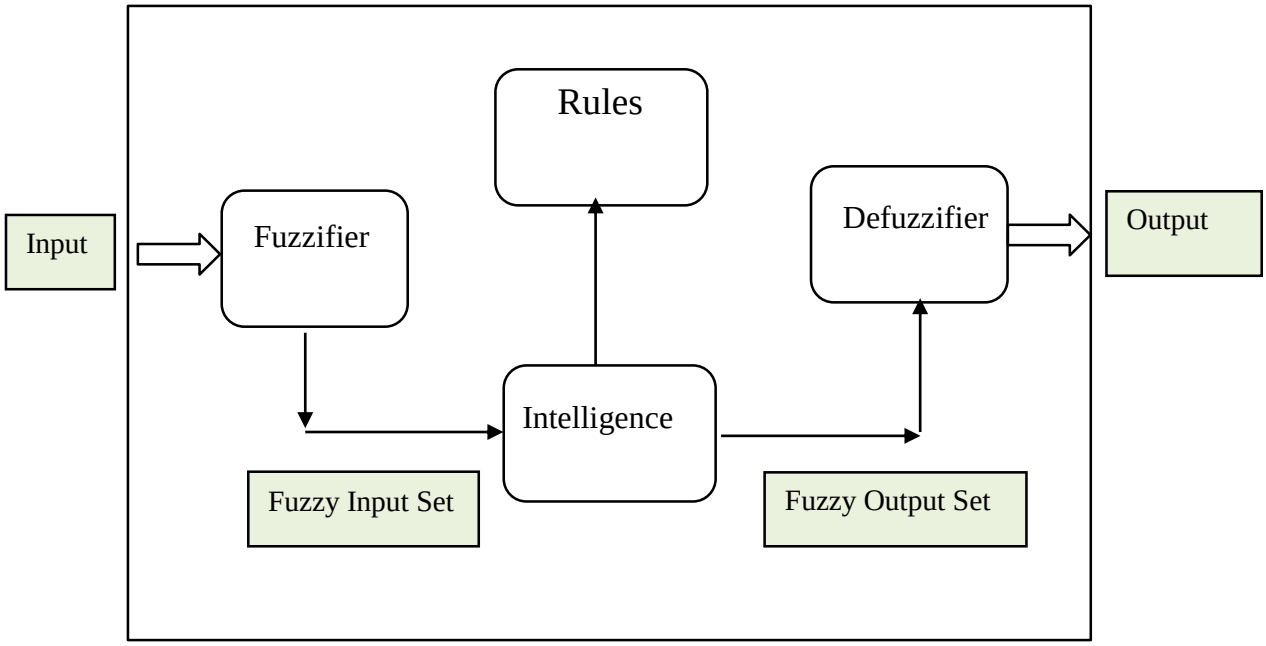


Figure 5.2.1 Fuzzy Logic

3.6 The background and rationale of design choice

Naive Bayes is a classification algorithm based on Bayes' Theorem, named after the 18th-century statistician Thomas Bayes. It is a simple but powerful algorithm that has been used in various applications, such as text classification, spam filtering, and recommendation systems. The Naive Bayes algorithm assumes that the features in the dataset are independent of each other, and it calculates the probability of a particular class given a set of features. It particularly useful when dealing with high-dimensional datasets, and it is computationally efficient.

The k-NN algorithm, short for k-Nearest Neighbors, is another widely used machine learning algorithm. It is a non-parametric algorithm that can be used for both classification and regression problems. The algorithm works by calculating the distance between the data points and identifying the k-nearest neighbors to a given point. The class of the new data point is then determined by majority voting of the k-nearest neighbors. The k-NN algorithm

is particularly useful when dealing with small datasets, and it is easy to interpret and implement.

Fuzzy Logic is a method of reasoning that is based on degrees of membership rather than binary true/false values. It is a way to represent imprecise and uncertain information in a formal way. Fuzzy Logic was introduced by Lotfi Zadeh in the 1960s, and it has been used in various applications, such as control systems, decision-making, and pattern recognition. Fuzzy Logic is particularly useful when dealing with complex systems that are difficult to model using traditional mathematical approaches.

The rationale behind these algorithms is to provide a way to model and make predictions based on data. Naive Bayes and k-NN are both examples of supervised learning algorithms, which means that they require labeled data to learn from. Fuzzy Logic, on the other hand, can be used for both supervised and unsupervised learning. The goal of these algorithms is to learn a model from the data that can be used to make predictions on new data.

In summary, Naive Bayes, k-NN algorithm, and Fuzzy Logic are all powerful machine learning algorithms that can be used for a variety of applications. Naive Bayes is a simple and efficient algorithm that works well with high-dimensional datasets, while k-NN is easy to implement and interpret and is useful for small datasets. Fuzzy Logic is a flexible method of reasoning that can handle uncertainty and imprecise data. Together, these algorithms provide a range of tools for data modeling, classification, and prediction.

3.7 An evaluation of my choice of method

Based on the requirements of my thesis, I have chosen to apply three machine learning algorithms: Naive Bayes, K-NN, and Fuzzy Logic. They are popular classification algorithm that is particularly useful when dealing with large datasets with a high number

of features.

Naive Bayes is a probabilistic algorithm that calculates the probability of a given event based on the occurrence of other events. It is widely used in natural language processing (NLP) and text classification. Naive Bayes is popular because it is simple to implement and computationally efficient. It requires a small amount of training data, making it ideal for applications where data is scarce. Naive Bayes also performs well with high-dimensional data, such as text data, and can handle missing data points easily.

KNN algorithm is a non-parametric algorithm that works by comparing the distance between the nearest neighbors to determine the classification of the new data point. KNN is used for both classification and regression problems and can handle both structured and unstructured data. KNN is popular in industry because it is easy to understand and can produce accurate results with large datasets. KNN can also adapt to changing environments, making it useful in real-time applications such as recommendation systems and fraud detection.

Fuzzy logic is a mathematical approach that allows for uncertainty and imprecision in decision-making. Fuzzy logic is used in applications where traditional logic fails, such as in control systems and pattern recognition. Fuzzy logic is popular because it can handle uncertainty and imprecision in data, making it ideal for real-world applications where data is often ambiguous or incomplete. Fuzzy logic is also useful for solving complex problems that require human-like decision-making, such as in robotics and artificial intelligence.

Naive Bayes, the KNN algorithm, and fuzzy logic in machine learning and artificial intelligence for their unique features and capabilities. Naive Bayes is popular because it is simple, efficient, and performs well with high-dimensional data, such as text data. KNN is popular because it is easy to understand, can handle both structured and unstructured data,

and can adapt to changing environments. Fuzzy logic is popular because it can handle uncertainty and imprecision in data, making it ideal for real-world applications where data is often ambiguous or incomplete.

3.8 Requirements

Once we carefully examined all the essential statistical and theoretical concepts and techniques, we compiled a roster of the required hardware, software, and development tools necessary for forecasting career recommendations.

Hardware/Software Requirements:

- Operating System (Windows 7 or more)
- Ram (4 GB or more)
- Web Browser (chrome or any suitable one)

Developing Tools

- Python 3.10 or latest version
- Anaconda
- Jupiter notebook
- Pandas
- Sklearn

Chapter IV. Result and Discussion

This chapter will cover the details of our research, including the experimental results. We will compare our algorithms to select the most effective one for prediction, and visualize the output using a confusion matrix. To conclude the chapter, we will summarize the results.

4.1 Experimental Results

We utilized more than 1000 simulated datasets from various fields of Computer Science and Engineering and crafted different questions to evaluate the recommended field of study for a student's future career and job. The dataset comprises ten diverse attributes covering several questions related to a student's personal life and interests. The output of this analysis is a recommendation of a student's field of study based on their interests and previous academic pursuits, specific to the Bangladeshi environment. Our output class for the system is a student's chosen discipline or field of study, with ten fields. We employed two machine learning models, namely Naïve Bayes and k-NN and Fuzzy Logic to analyze our datasets and generate recommendations. Although the initial results were unsatisfactory, a real-world dataset is expected to enhance the accuracy of the output.

4.2 Prediction of Results

The machine learning model's performance can be assessed using a confusion matrix, which is a table consisting of four values: true positive, true negative, false positive, and false negative. This technique can be applied to the test dataset, where the true values are already known. True positives (TP) are the values where the examples are accurately identified as positive. False positives (FP) are the values where the examples are negative,

but they are mistakenly classified as positive. True negatives (TN) are the values where the examples are accurately identified as negative. Finally, false negatives (FN) are the values where the examples are positive, but they are incorrectly classified as negative.

Performance measures can be calculated using the following formulas:

1) TP = TP/(TP+FN)
2) FP = FP/(FP+TN)
3) Precision = TP/(TP+FP)
4) Recall = TP/(TP+FN)
5) Accuracy = (TP+TN)/(TP+TN+FP+FN)
6) F-measure=2 * (precision*recall)/(precision+recall)
7) Error- rate = 1-accuracy

TABLE 4.2.1: PERFORMANCE MEASUREMENTS

Classification Model	Accuracy	F-measure	Error- rate
Naïve Bayes	0.931	0.913	0.069
K-NN	0.894	0.90	0.106
Fuzzy Logic	o.871	0.882	0.129

Based on the table presented above, the Naïve Bayes algorithm has been identified as the most effective classifier for providing recommendations regarding suitable departments.

The figure 4.2.1 shows the accuracy of the models:

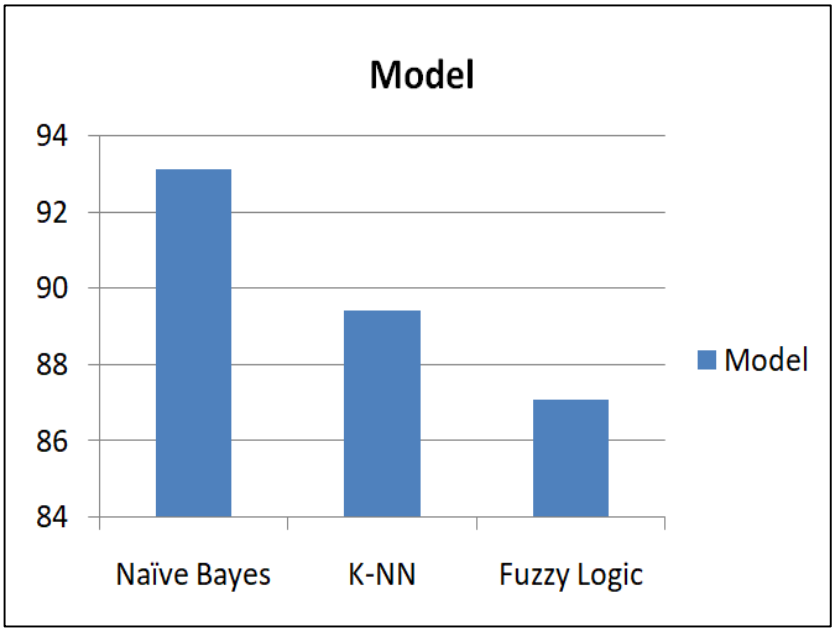


Figure 4.2.1: Accuracy

F=-Measures values are mentioned below:

TABLE 4.2.2: F-MEASURE OF CATEGORY

Categories	F-Measure
Software Development	0.951
Data Science	0.90
Programming Skills	0.963
Database Fundamentalist	0.936
Troubleshooting skills	0.891
Cyber Security	0.873
Project Management	0.945

4.3 Significance of Research

The research on career guidance recommendation system holds significant importance in today's rapidly changing and competitive job market. With the increasing number of fields and specializations, it can be overwhelming for students to make the right career choice. A recommendation system based on machine learning algorithms can provide personalized career guidance to students, which can help them make informed decisions about their future. The significance of this research lies in its potential to benefit students, educational institutions, and industries. By providing students with personalized recommendations based on their interests and academic history, the system can assist them in selecting the most suitable career path. Educational institutions can also use this system to enhance their career guidance services and ensure that their students are well-equipped for the job market. Additionally, industries can benefit from a more skilled and specialized workforce, which can ultimately lead to economic growth. Moreover, this research can contribute to the development of machine learning algorithms and techniques, which can be applied to various other fields. It can also provide insights into the factors that influence career choices and help researchers and policymakers develop strategies to address issues related to unemployment and skill gaps. Overall, the research on career guidance recommendation systems can have a significant impact on the education and job sectors, the economy, and the advancement of machine learning techniques. It can ultimately benefit society as a whole by providing individuals with the guidance they need to succeed in their chosen careers.

4.4 Limitations

We are working with dummy datasets but real world datasets are better for prediction. If we use dummy datasets our prediction can be or results can be different. Our research is only for the fields of Computer Science and Engineering Departments. We have not covered other departments like CE, ME, EEE etc. Also we have not covered Medical and Business field.

4.5 Implementation for further study

There are several implications for further study that can be drawn from our research:

- Data sets can cover the real world data sets and departments can be other engineering departments, Business and Medical fields.
- Since our research was conducted using dummy datasets, one important future direction would be to test our approach on real-world datasets to evaluate its effectiveness in practical settings.
- We can also consider adding more fields of study to our research work to provide a wider range of recommendations to students.
- Another potential avenue for further study is to deploy our research as an Android and website application to make it more accessible and user-friendly for students seeking career guidance.
- Lastly, we can explore the possibility of publishing a publication that details our research work to disseminate our findings to a wider audience and invite further discussion and collaboration in the field.

Chapter V. Conclusion

5.1 Conclusion

The career guidance recommendation system is designed for Bangladeshi students who struggle with choosing a subject for their graduation. By using machine learning models like Naïve Bayes, k-NN, and Fuzzy Logic, the system recommends a field of study based on a student's interests and previous academic pursuits. This research can assist students in selecting a field of study and can benefit industries that employ these students in the future. Furthermore, the system has the potential to be extended to other fields of study and deployed as a mobile application or website. This can greatly benefit students in making informed decisions about their career paths and help them achieve their desired career goals. The future work for this research includes using real-world datasets to further validate the accuracy of the system, adding more fields of study, and improving the system's user interface. Overall, the Career Guidance Recommendation System using Machine Learning can have a significant impact on improving the career guidance services provided to students, and help them make informed decisions about their future career paths.

References:

- Vignesh, S., Priyanka, C.S., Manju, H.S. and Mythili, K., 2021, March. An intelligent career guidance system using machine learning. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 987-990). IEEE.

- Roy, K.S., Roopkanth, K., Teja, V.U., Bhavana, V. and Priyanka, J., 2018. Student career prediction using advanced machine learning techniques. International Journal of Engineering & Technology, 7(2.20), pp.26-29. [3] Metha S., Kamble S., Kumar. (2021) 'Career Guidance using Machine Learning.' 9(VI) pp. 4124-4127.

- Metha S., Kamble S., Kumar. (2021) 'Career Guidance using Machine Learning.' 9(VI) pp. 4124-4127.

- Ajay, D., Nilesh, S., Anirudh, D., Huzaif, S. and Nagare, P.V.(2022) 'AN INTELLIGENT CAREER GUIDANCE SYSTEM USING MACHINE LEARNING.' 9(3) pp. 1684-1691.

- sayyed F.,Sanghani R., Vora A., Lemos. (2020) 'AI based Career Guidance.' 7(5) pp. 2505-2509

- Copeland, L.Y., Dik, B.J., McLaren, M.R., Onder, C., Wolfson, N.E. and Kraiger, K., 2011. Recommendations for using computer-assisted career guidance systems (CACGS) in career counseling practice. *Journal of Psychological Issues in Organizational Culture*, *2*(3), pp.86-94.

- Gati, I., Levin, N. and Landman-Tal, S., 2019. Decision-making models and career guidance. *International handbook of career guidance*, pp.115-145.

- Roy, K.S., Roopkanth, K., Teja, V.U., Bhavana, V. and Priyanka, J., 2018. Student career prediction using advanced machine learning techniques. International

Journal of Engineering & Technology, 7(2.20), pp.26-29. [3] Metha S., Kamble S., Kumar. (2021) 'Career Guidance using Machine Learning.' 9(VI) pp. 4124-4127.

- Nabil Ahmad, S., 2020. Personalized Career Path Recommender System for Engineering Student. *IEEE*, 8 October , 6(14), p. 9.

- Vignesh, A. H. N. H., 2021. Career Path Recomendation System. *IEEE*, 17 June, 13(16), p. 7.