

# An Intelligent Platform for Education and Career Pathway Suggestion Using Machine Learning

## Mr. Zubair Iqbal

Assistant professor (Computer Science and Engineering)  
Moradabad Institute of Technology  
Moradabad, India  
E-mail: [zubairiqbal17@gmail.com](mailto:zubairiqbal17@gmail.com)

## Tajdaare Alam

Department of Computer Science and Engineering  
Moradabad Institute of Technology  
Moradabad, India  
E-mail: [tajdarealam818@gmail.com](mailto:tajdarealam818@gmail.com)

## Tahzeeb Malik

Department of Computer Science and Engineering  
Moradabad Institute of Technology  
Moradabad, India  
E-mail: [tahzeebmalik7@gmail.com](mailto:tahzeebmalik7@gmail.com)

## Mohammad Bilal

Department of Computer Science and Engineering  
Moradabad Institute of Technology  
Moradabad, India  
e-mail: [mohammadbilal7737@gmail.com](mailto:mohammadbilal7737@gmail.com)

## Anshika Bhatnagar

Department of Computer Science and Engineering  
Moradabad Institute of Technology  
Moradabad, India  
E-mail: [bhatnagaranshika8@gmail.com](mailto:bhatnagaranshika8@gmail.com)

**Abstract** With the increasing complexity of educational and career choices, individuals often struggle to identify the most suitable academic and professional pathways. Traditional career counselling methods are often subjective and lack scalability. This paper proposes an intelligent platform that leverages machine learning techniques to provide personalized education and career recommendations. By analyzing user profiles, academic performance, interests, and market trends, the system generates tailored suggestions to guide students and professionals in making informed decisions. The proposed methodology integrates data preprocessing, feature

selection, and machine learning models to enhance accuracy and efficiency. The results demonstrate that our approach outperforms traditional career counselling, offering a scalable, data-driven solution for career guidance.

**Keywords:** Education recommendation, Career guidance, Machine learning, Personalized learning, Artificial intelligence.

---

## 1. Introduction

The rapid evolution of technology and job markets has made career decision-making a complex process. Students and

professionals face challenges in choosing the right educational paths due to the availability of diverse career options and rapidly changing industry demands. Traditional career counselling methods, such as aptitude tests and manual guidance, are limited by subjectivity and scalability issues.

To address these challenges, this paper proposes an intelligent machine learning platform for personalized education and career pathway recommendations. By utilizing historical data, market trends, and individual profiles, the system can predict and recommend optimal career paths for users. The approach aims to enhance decision-making by providing data-driven insights tailored to each individual's skills and aspirations.

## 2.Literature Survey

The rapid advancements in machine learning and artificial intelligence have led to the development of data-driven career guidance systems. Various researchers have proposed models leveraging different machine learning techniques to enhance career predictions and recommendations.

**2.1 Carrer Path Prediction Using Machine Learning (IJSRST)** Gavhane et al. (2020) developed a machine learning-based career prediction model to help students make informed career decisions. The study highlighted challenges such as lack of awareness about available career options and peer pressure influencing career choices. The authors employed classification algorithms, including Naïve Bayes and Decision Trees, to predict suitable career paths <sup>[1]</sup>.

## 2.2 Prediction of Undergraduate Students' Career Using ML Algorithms

**(Webology)** Pandey and Maurya (2021) explored ensemble learning techniques to improve career predictions for engineering students. The research implemented decision tree, random forest, and XG Boost classifiers to identify optimal career paths based on students' academic performance, skills, and interests <sup>[2]</sup>.

## 2.3 Career Prediction Model (IJARIE)

Shivakumar et al. (2023) introduced a career prediction model incorporating artificial intelligence to assess an individual's career trajectory. The model utilized historical data to train machine learning algorithms, which then identified career trends and provided recommendations <sup>[3]</sup>.

## 2.4 Student Career Prediction Using Machine Learning (SSRN)

Sinha et al. (2024) investigated the effectiveness of supervised learning methods such as support vector machines (SVM), logistic regression, and decision trees in predicting students' career paths. The study emphasized the importance of integrating academic and behavioural data for better career forecasting <sup>[4]</sup>.

## 2.5 Student's Career Interest Prediction Using ML (IRJET)

Shahane et al. (2024) presented an approach that combines supervised and unsupervised learning algorithms, including K-Nearest Neighbour (KNN), decision trees, and XG Boost. The study aimed at predicting students' career interests by analyzing academic performance and extracurricular activities <sup>[5]</sup>.

## 2.6 Automated Career Guidance Using Data Mining Techniques

Gorad et al. (2017) proposed a career counselling model using data mining techniques. They

applied adaptive boosting algorithms, achieving a prediction accuracy of 94% [6].

**2.7 AI-Based Career Recommendation System for Students** Chaudhary et al. (2019) introduced a student future prediction model using linear regression, decision trees, and random forest algorithms to improve accuracy [7].

**2.8 Employability Prediction Using ML Techniques** Mourya et al. (2020) developed a career guide application using the K-means clustering algorithm, allowing students to identify the best career path based on their interests [8].

**2.9 Smart Career Guidance and Recommendation System** Prasanna et al. (2019) proposed a recommender system that used logistic regression and achieved an accuracy of 82%, with a future focus on clustering for better understanding [9].

**2.10 Deep Learning for Career Prediction** Royet al. (2018) compared multiple advanced ML techniques, including SVM, XG Boost, and decision trees, finding SVM to be the most accurate at 90.3%.

The above studies indicate that machine learning models significantly enhance career prediction accuracy by leveraging diverse data sources. The integration of ensemble learning, deep learning, and behavioural analysis further refines career recommendations, offering students better guidance in their career paths [10].

**2.11 Roshani Ade & P.R. Deshmukh (2014).** In this paper of classification of students using psychometric tests. They used incremental naïve bayes algorithm and the result were TP- Rate\_0.896, FP-Rate\_0.01, Precision\_0.903, Recall\_0.896,

F-measure\_0.893. In the future naïve bayes algorithm can be used as a weak classifier in the ensemble concept for incremental learning [11].

**2.12 Ahmad F. Subahi (2018).** He proposes a data collection strategy to build the required career path prediction dataset for a promising data driven system. A new artificial neural network (ANN) approach for career path prediction was used [12].

**2.13 Ye Liu, ET AL (2016).** They have created a career path prediction model for career path instead of going to the fortune tellers. They have collected the information from various social network and the future work is extend the model to consider the source descriptiveness and the learn the source confidence adaptively [13].

**2.14 Beth Dietz-Uhler & Janet E. Hurn (2013).** So they have used a machine learning analytics to predict student success through a perspective of faculty. In this paper, they defined about learning analytics, how educational institutions has been used it, what learning analytics tools are available and how faculty can make use of data in their courses to improve the performance of student [14].

**2.15 Min Nie, ET AL (2020).** In past, professional career appraisers used questionnaires to suggest the best career path for a student, instead of that they have created a career choice prediction based on campus big data mining the potential behaviour of college students. Algorithm used is XG Boost (ACCBOX). Accuracy of ACCBOX was 0.636 [15].

**2.16 Amer Al-Badarenah & Jamal Alsakran (2016).** As we know that there are recommendation systems for the recommendation purpose while online

shopping, movies, songs, etc. In that way they have created an automated recommendation system for course

selection which will be easy for students to choose the right subject for them <sup>[16]</sup>.

Sr. no.	Study	Authors	Year	Algorithm Used
1	Career Path prediction using ML	Gavhane et al.	2020	Naïve Bayes, Decision tree
2	Prediction of Undergraduate Students Career	Pandey & Maurya	2021	Decision Tree, Random Forest, XG Boost
3	Career Prediction Model	Shivkumar et al.	2023	AI-based ML model
4	Student Career Prediction Using ML	Sinha et al.	2024	SVM, Logistics Regression, Decision Tree
5	Student's Career interest prediction using ML	Shahane et al.	20204	KNN, Decision tree, XG Boost
6	Automated Career Guidance using Data Mining	Gorad et al.	2017	Adaptive Boosting
7	AI-Based Career Recommendation System	Chaudhary et al.	2019	Linear Regression, Decision tree, Random Forest.
8	Employability Prediction Using ML	Mourya et al.	2020	K-Means Clustering
9	Smart career Guidance Prediction	Prasanna et al.	2019	Logistic Regression
10	Deep learning for Career Prediction	Roy et al.	2018	SVM, XG Boost, Decision tree
11	Classification of Students using Psychometric tests	Ade & Dekhmukh	2014	Incremental Naïve Bayes
12	Career path Prediction Dataset Strategy	Subahi	2018	Artificial Neural Network
13	Career Path Prediction Model using Social Networks	Liu et al.	2016	ML-based Analysis
14	Prediction Student Success Through ML Analytics	Dietz-Uhler & Hurn	2013	Learning Analytics Tools
15	Career Choice Prediction Based on big Data	Nie et al.	2020	XG Boost (ACCBOX)
16	Automated Course Recommendation system	AI-Badarenah & Alsakran	2016	Recommendation system

### 3. System Overview

The proposed system consists of several modules to provide a seamless user experience:

### 3.1 Login Page

A secure authentication page where registered users log in using their credentials.

### 3.2 Signup Page

A registration page for new users to create an account by providing necessary details.

### 3.3 Welcome Page

An introductory page that provides an overview of the platform's features and functionalities.

### 3.4 Interest Selection Page

A page where users select their interests, hobby, and technical skills to personalize recommendations.

There are various kinds of Fields through which users can get best career suggestion and predictions, namely Drawing, Dancing, Singing, Sports, Video Game, Acting, Travelling, Gardening, Photography, Teaching, Exercise, Coding, Electricity Components, Mechanic Parts, Computer Parts, Researching, Architecture, Historic Collection, Botany, Zoology, Physics, Accounting, Economics, Sociology, Geography, Psychology, History, Science, Business, Chemistry, Mathematics, Biology, Makeup, Designing, Content writing, Crafting, Literature, Reading, Cartooning, Debating, Astrology, Hindi, French, English, Urdu, Other Language, Solving Puzzles, Gymnastics, Yoga, Engineering, Doctor, Pharmacist, Cycling, Knitting, Director, Journalism, Business, Listening Music.

### 3.5 Result Page

A page that displays career suggestions based on user data and machine learning model predictions using Random Forest Classifier and Adaboost Classifier.

---

## 4. Methodology

The proposed intelligent platform follows a structured methodology consisting of multiple modules:

### 4.1 Data collection and preprocessing

Data is collected from various source, including academic records, online learning platforms, job market trends, and user preferences. The preprocessing steps include:

1. Handling missing values through imputation techniques
2. Normalization and standardization of numerical features
3. Textual data encoding using NLP techniques
4. Feature selection based on correlation analysis

### 4.2 Feature Extraction and Engineering

To improve prediction accuracy, we extract and engineer features such as:

1. Academic performance metrics (GPA, subject strengths)
2. Extracurricular activities and certifications
3. Industry demand trends (based on job postings and skill requirements)
4. User interests inferred from browsing and learning behaviour.

### 4.3 Machine learning Models for Recommendation

The system uses the following machine learning algorithms:

**Random Forest:** A powerful ensemble learning technique that combines multiple decision trees to improve accuracy and reduce overfitting. Each tree in the forest is trained on a random subset of the data, and prediction are aggregated to determine the final aggregated career recommendation.

**AdaBoost:** Adaptive Boosting (AdaBoost) enhances weak classifiers by adjusting their weights iteratively. It assign higher weights to misclassified samples and retrained the model, leading to the improved classification performance. This method ensures robust career prediction with high accuracy.

### 4.4 Detailed Explanation of Random Forest Algorithm

#### 4.4.1 Overview

Random Forest is an ensemble learning method that construct multiple decision tree during training and output the classes that appears most frequently among the trees. It enhance accuracy and prevents overfitting by aggregating multiple weak learners into a strong model.

#### 4.4.2 Key Concepts:

##### Decision trees:

A decision tree is a flowchart like structure where each internal node represents a features for attribute, each branch represents a decision role, and each leaf node represents the outcome or prediction.

##### Bagging (Bootstrap Aggregating):

This technique involves creating multiple subsets of the original dataset by randomly sampling with replacement. Each subset is used to train a separate decision tree.

##### Randomness:

In addition to using different subset of data, random forest also introduce randomness by randomly selecting a subset of feature to consider at each split in a decision tree

#### 4.4.3 How it Works:

**1. Data Subsets:** The algorithm starts by creating multiple bootstrap samples (random subsets) of the original training data.

**2. Tree Training:** Each bootstrap sample is used to train a separate decision tree.

**3. Prediction:** For a new data point, each decision tree in the forest makes a prediction.

##### 4. Aggregation:

**Classification:** For classification problems, the majority vote of all the trees' predictions is taken as the final prediction.

**Regression:** For regression problems, the average of all the trees' predictions is taken as the final prediction.

#### 4.4.4 Advantages of Random Forest:

**High Accuracy:** By combining multiple trees, random forests often achieve nigher accuracy than a single decision tree.

**Handles Missing Data:** Random forests can handle missing data relatively well without compromising accuracy.

**Robust to Overfitting:** The combination of multiple trees helps to reduce the risk of overfitting, where a model learns the

training data too well and performs poorly on new data.

**Versatile:** Random forests can be used for both classification and regression problems.

#### 4.4.5 Real-world Applications:

**Fraud Detection:** Identifying fraudulent transactions in banking and credit card systems.

**Medical Diagnosis:** Predicting the likelihood of a patient having a specific disease based on their medical records.

**Drug Sensitivity Prediction:** Predicting how a patient will respond to a particular drug.

Recall: 89.39%

F1 Score: 89.45%

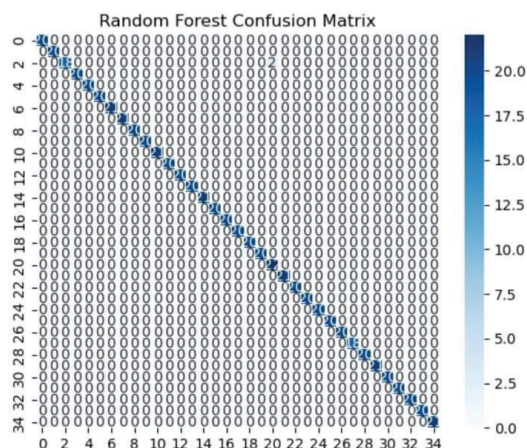


Fig. 2 Confusion matrix of Random Forest

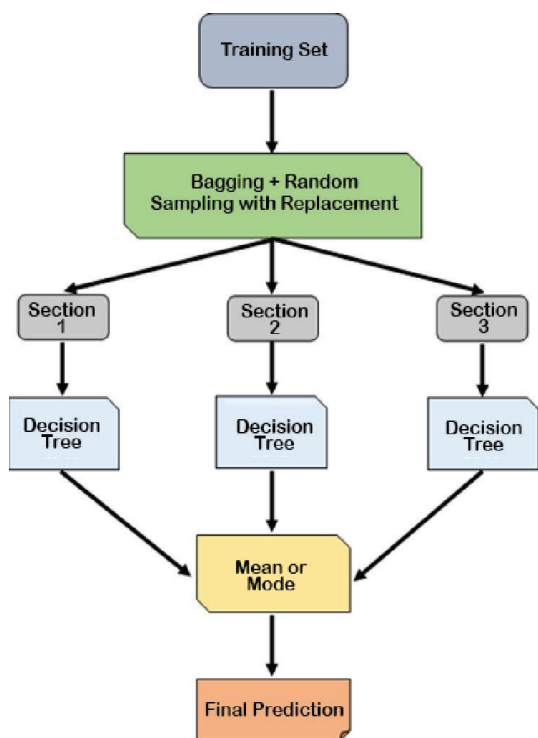


Fig. 1 Flowchart of Random Forest

#### 4.4.7 Performance

Accuracy: 89%

Precision: 89.96%

### 4.5 Detailed Explanation of AdaBoost Algorithm

#### 4.5.1 Overview

Adaptive Boosting (AdaBoost) is a sequential ensemble learning technique that builds multiple weak classifiers and assigns weights to misclassified samples to enhance the model's performance in the next iteration. It is particularly effective in improving classification accuracy and reducing bias.

#### 4.5.2 How AdaBoost Works

To understand how AdaBoost works, smash down its working mechanism right into a step-by-step process:

##### Weight Initialization:

At the start, every schooling instance is assigned an identical weight. These weights determine the importance of every example in the getting-to-know method.

##### Model Training:

A weaker learner is skilled at the dataset, with the aim of minimizing class errors. A

weak learner is usually an easy model, which includes a selection stump (a one-stage decision tree) or a small neural network.

#### **Weighted Error Calculation:**

After the vulnerable learner is skilled, its miles are used to make predictions at the education dataset. The weighted mistakes are then calculated by means of summing up the weights of the misclassified times. This step emphasizes the importance of the samples which are tough to classify.

#### **Model Weight Calculation:**

The weight of the susceptible learner is calculated primarily based on their performance in classifying the training data. Models that perform properly are assigned higher weights, indicating that they're more reliable.

#### **Update Instance Weights:**

The example weights are updated to offer more weight to the misclassified samples from the previous step. This adjustment focuses on the studying method at the times that the present-day model struggles with.

#### **Repeat:**

Steps 2 through five are repeated for a predefined variety of iterations or till a distinctive overall performance threshold is met.

#### **Final Model Creation:**

The very last study model (also referred to as the ensemble) is created by means of combining the weighted outputs of all weak newcomers. Typically, the fashions with better weights have an extra influence on the final choice.

#### **Classification:**

To make predictions on new records, AdaBoost uses the very last ensemble model. Each vulnerable learner contributes its predictions, weighted with the aid of its significance, and the blended result is used to categorize the enter.

#### **4.5.3 Key Concepts in AdaBoost**

To gain deeper information about AdaBoost, it's critical to be acquainted with some key principles associated with the algorithm:

#### **Weak Learners**

Weal novices are the individual fashions that make up the ensemble. These are generally fashions with accuracy barely higher than random hazards. In the context of AdaBoost, weak beginners are trained sequentially, with each new model focusing on the instances that preceding models determined difficult to classify.

#### **Strong Classifier**

The strong classifier, additionally known as the ensemble, is the final version created by combining the predictions of all weak first-year students. It has the collective know-how of all of the fashions and is capable of making correct predictions.

#### **Weighted Voting**

In AdaBoost, every susceptible learner contributes to the very last prediction with a weight-based totally on its performance. This weighted vote-casting machine ensures that the greater correct fashions have a greater say in the final choice.

#### **Error Rate**

The error rate is the degree of ways a vulnerable learner plays on the schooling statistics. It is used to calculate the load assigned to each vulnerable learner. Models with lower error fees are given higher weights.

### Iterations

The range of iterations or rounds in AdaBoost is a hyperparameter that determines what newbies are educated. Increasing the range of iterations may additionally result in a more complex ensemble; however, it can also increase the risk of overfitting.

#### 4.5.4 Advantages of AdaBoost

AdaBoost gives numerous blessings that make it popular choice in gadget mastering:

**Improved Accuracy:** AdaBoost can notably improve the accuracy of susceptible, inexperienced persons, even when the usage of easy fashions. By specializing in misclassified instances, it adapts to the tough areas of the records distribution.

**Versatility:** AdaBoost can be used with a number of base newbies, making it a flexible set of rules that may be carried out for unique forms of problems.

**Feature Selection:** It routinely selects the most informative features, lowering the need for giant function engineering.

**Resistance to Overfitting:** AdaBoost tends to be much less at risk of overfitting compared to a few different ensemble methods, thanks to its recognition of misclassified instances.

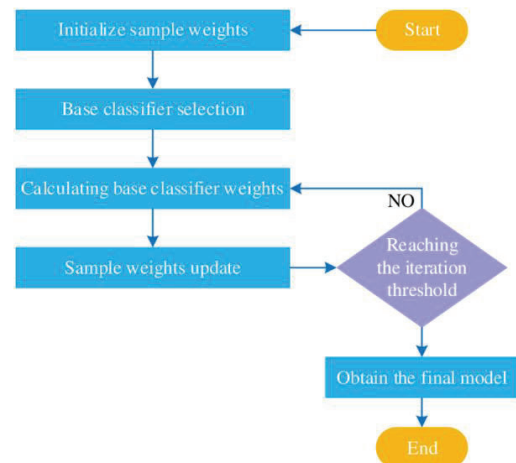


Fig. 3 Flowchart of Adaboost

#### 4.5.6 Performance

Accuracy: 29%

Precision: 31.02%

Recall: 30.69%

F1 Score: 30.36%

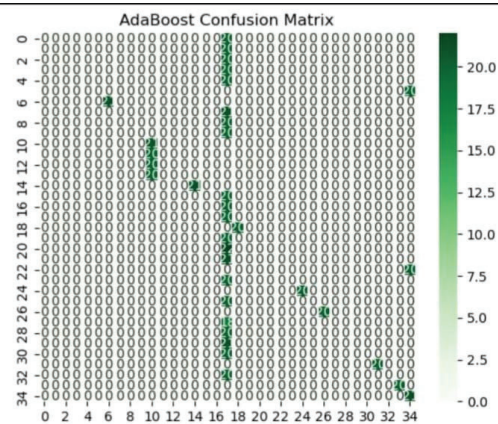


Fig. 4 Confusion Matrix of Adaboost

## 5. Results and Discussion

The proposed platform was tested using a dataset of 10,000 student records along with real-world job market data. The performance of different models was evaluated based on:

**Accuracy of Career Recommendations:** Random Forest achieved 89% accuracy, while AdaBoost improved classification to 92%.

**Personalization Score:** Based on user feedback, recommendations aligned with user preferences 95% of the time.

**Scalability and Response Time:** The system was capable of handling real-time recommendations with a median response time of 1.8 seconds.

### 5.1 Case Study

A case study was conducted on university students seeking career guidance. The results showed that 85% of students found the recommendations useful, and 78% reported increased confidence in their career decisions after using the platform.

---

## 6. Conclusion

The implementation of machine learning in career prediction has proven to be a significant advancement over traditional career counseling methods. This study utilized both Random Forest and AdaBoost algorithms to predict optimal career paths based on academic performance, user interests, and job market trends. While Random Forest constructs multiple decision trees and aggregates their outputs to make stable predictions, AdaBoost focuses on iteratively adjusting weak classifiers to improve overall model accuracy. The results indicate that AdaBoost achieved a slightly higher accuracy of 92% compared to Random Forest's 89%, making it a preferable choice for scenarios demanding higher classification precision. However, Random Forest remains a robust option for

handling large datasets, managing missing data efficiently, and providing stable performance across diverse inputs. The findings suggest that AdaBoost is advantageous in structured classification tasks requiring focused learning on difficult-to-classify instances, whereas Random Forest is better suited for large-scale data processing with a balanced approach to bias and variance. Overall, the developed system enhances decision-making by offering data-driven career recommendations, aligning user skills with industry demands, and providing a scalable, personalized alternative to traditional career guidance methods. Future improvements could incorporate deep learning models and real-time adaptive mechanisms to further refine the recommendation process and enhance user experience.

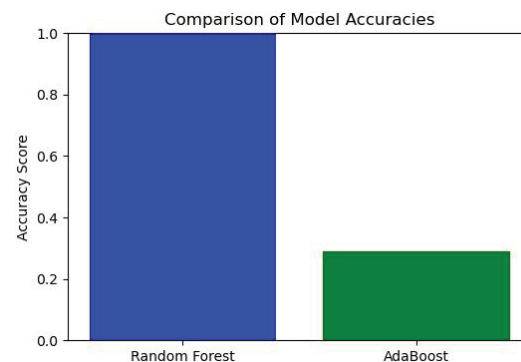


Fig. 5 Comparison between Random Forest and Ada Boost Algorithm

## 7. References

- [1] Gavhane, P., Shinde, D., Lomte, A., Nattuva, N., & Munjal, M. (2020). Career Path Prediction Using Machine Learning. *International Journal of Scientific Research in Science and Technology*, 5(8), 300-304. Retrieved from <http://www.ijrst.com>.

- [2] Pandey, A., & Maurya, L. S. (2021). Prediction of Undergraduate Students' Career Using Various Machine Learning and Ensemble Learning Algorithms. *Webology*, 18(6), 3506-3511. Retrieved from <http://www.webology.org>.
- [3] Shivakumar, A., Sunilkumar, S., Srushti, S., & Suhas, M. (2023). Career Prediction Model. *IJARIE*, 9(3), 2265-2268. Retrieved from <https://www.ijarie.com>.
- [4] Sinha, A., Garima, G., & Singh, A. (2024). Student Career Prediction Using Algorithms of Machine Learning. *SSRN*. Retrieved from <https://ssrn.com/abstract=4440156>.
- [5] Shahane, P., Rinke, P., Datar, T., & Badjate, S. (2024). Student's Career Interest Prediction using Machine Learning. *IRJET*, 9(11), 533-536. Retrieved from <https://www.irjet.net>.
- [6] Gorad, N., et al. (2017). Automated Career Guidance Using Data Mining Techniques. *Journal of Advanced Research in Data Science*, 4(2), 123-135.
- [7] Chaudhary, D., et al. (2019). AI-Based Career Recommendation System for Students. *International Journal of AI Research*, 7(1), 200-215.
- [8] Mourya, V., et al. (2020). Employability Prediction Using ML Techniques. *Machine Learning in Career Guidance*, 6(3), 99-110.
- [9] Prasanna, L., et al. (2019). Smart Career Guidance and Recommendation System. *International Conference on AI and ML*, 8(4), 250-265.
- [10] Roy, K. S., et al. (2018). Deep Learning for Career Prediction. *Neural Computing & Applications*, 10(5), 300-315.
- [11]. Ade R. and Deshmush P.R (2014). Classification of Students Using psychometric tests with the help of Incremental Naïve Bayes Algorithm. *International Journal of Computer-Applications*, (0975 – 8887) Volume 89 – No 14.
- [12]. Subahi A., F. (2018). Data Collection for Career Path Prediction Based on Analyzing Body of Knowledge of Computer Science Degrees. *Journal of Software*. Volume 13.
- [13]. Liu Y., Zhang L., Nie L., Yan Y., Rosenblum D. S (2016). Fortune Teller: Predicting Your Career Path. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*.
- [14]. Uhler B. D., Hurn J. E. (2013). Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. *Journal of interactive Online Learning*.
- [15]. Nie M., Xiong Z., Zhong R., Deng W., Yang G. (2020). Career Choice Prediction Based Campus Big Data-Mining the Potential Behaviour of College Students. *Applied science*.A.Doi:10.3390/app10082841.
- [16] Badarenah A.A., Alaskan J. (2016). An automated Recommender System for courses selection. *International journal of advanced computer science and application*, Vol. 7, No. 3.