

# Student Career Prediction Using Advanced Machine Learning Techniques

**Mrs.Amrita Kindalkar, Kiran Naik**

Asst Professor, Department of Computer Science Engineering,

Srinivas University Institute Of Engineering and Technology Karnataka, India-574146

Student, Department of Computer Science Engineering,

Srinivas University Institute Of Engineering and Technology Karnataka, India-574146

## Abstract

Students and professionals need to make informed career choices based on their skills, interests, and industry demands. Career Quest: AI-Powered Pathway Finder with Real-Time Job Insights and Skill Development provides personalized career recommendations by analyzing academic backgrounds, skill sets, and preferences while incorporating real-time job market trends. The system also offers skill development resources aligned with industry requirements, ensuring users remain competitive. By integrating AI-driven analytics with up-to-date job insights, the platform enhances career decision-making for students and professionals while assisting recruiters in identifying the best-fit roles for candidates based on their competencies.

**Keywords:** Student Career Prediction, Decision Tree, Machine Learning, Job Market Analysis, Skill Development

## 1.Introduction

Career planning is essential for shaping a student's future, requiring a careful assessment of their skills, interests, and market demands. Career Quest: AI-Powered Pathway Finder with Real-Time Job Insights and Skill Development is designed to help students and professionals navigate their career paths efficiently. By analyzing various factors such as academic background, skills, industry trends, and real-time job insights, the system provides personalized career recommendations. This enables individuals to align their capabilities with suitable job roles and develop the necessary skills for career growth.

In addition to technical skills, Career Quest considers a broader range of factors, including extracurricular activities, interests, soft skills, hobbies, competitions, and overall knowledge. A total of 38 parameters are used as input for career prediction, allowing the system to classify candidates into distinct job roles based on these inputs. Due to the complexity of career pathways and the large number of input parameters, traditional programming and basic algorithms fail to provide accurate classification and prediction. To address this challenge, advanced machine learning algorithms such as Support Vector Machines (SVM), Random Forest, Decision Tree, OneHot Encoding, and XGBoost are employed to enhance prediction accuracy and ensure reliable career recommendations.

Recruiters also face challenges in evaluating candidates and assigning them to the most suitable roles based on their abilities. Various job roles—such as Data Scientist, Software Developer, Network Engineer, and Business Analyst—require specific skill sets. AI-driven career guidance platforms simplify the recruitment process by analyzing candidate competencies and recommending the best-fit roles. While existing career recommendation systems primarily focus on technical skills and psychometric evaluations, Career Quest integrates multiple parameters to provide comprehensive career insights. By leveraging machine learning techniques, it empowers students and professionals to make informed career decisions with greater confidence.

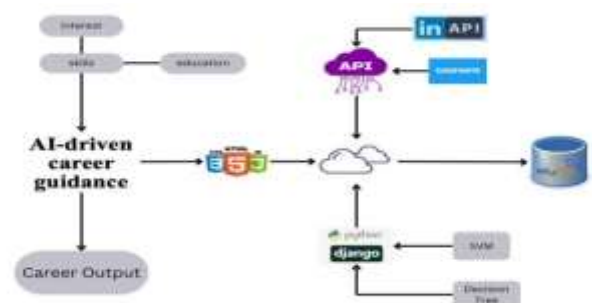


**Figure 1:** Overview of various Advanced Machine Learning Algorithms

Machine Learning is a technique where machines are trained to respond to specific inputs or scenarios based on previously learned data. Simply put, it enables computers to learn using statistical techniques without being explicitly programmed. Machine learning reduces human intervention in machine-dependent tasks, making it easier to solve complex problems efficiently with minimal human effort. Various applications of machine learning include Natural Language Processing (NLP), classification, prediction, image recognition, medical diagnosis, algorithm building, self-driving cars, and more. In this paper, classification and prediction techniques are utilized to enhance career guidance.

Most machine learning problems can be solved using supervised or unsupervised learning. In supervised learning, the final class labels are known beforehand, and new data points are assigned to one of these predefined labels based on their attributes. In contrast, unsupervised learning involves grouping data points based on similarities without predefined class labels. Classification falls under supervised learning, where input parameters are analyzed and assigned to a predefined category. Other approaches, such as clustering and regression, may also be used depending on the nature of the problem. The choice of the model depends on the specific problem being addressed to ensure optimal results.

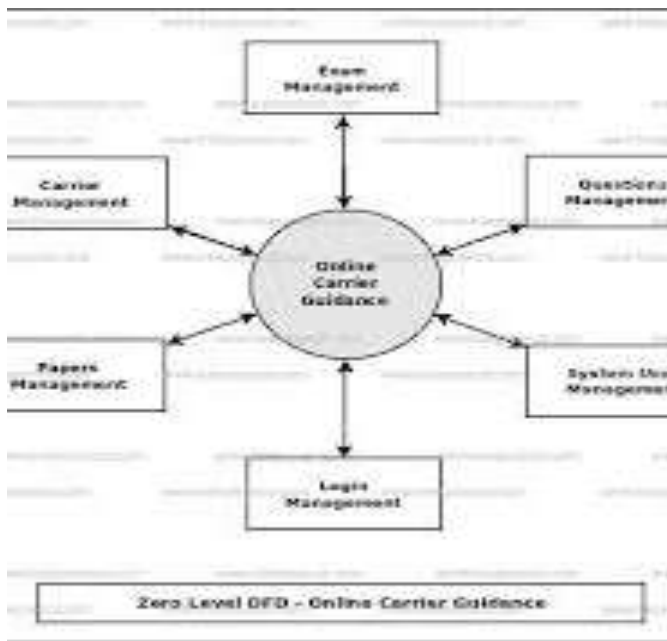


Figure 2: Process Flow Diagram of proposed system

However, algorithms like SVM, OneHot Encoding, Decision Tree, and XGBoost are used. After training and testing the data with these algorithms, the most accurate one is selected for further processing. The initial task involves predicting the output using all the proposed algorithms, followed by analyzing the results to determine the most effective model. Finally, this paper focuses on various advanced machine learning algorithms that involve classification and prediction, aiming to improve accuracy, reliability, and overall performance analysis.

## 2.Implementation

### 2.1 Data Collection and Processing:

The first step in implementation involves gathering relevant data, which includes:

- Academic records (courses, grades, certifications)
- Technical and soft skills assessment
- Industry demand and real-time job trends
- User interests, extracurricular activities, and hobbies

Once collected, the data undergoes preprocessing:

- **Cleaning** – Removing duplicate or inconsistent entries.
- **Encoding** – Converting categorical data into numerical form using techniques like OneHot Encoding.
- **Normalization** – Scaling the features to ensure uniformity.
- **Splitting** – Dividing data into training and testing sets.

### 2.2 Feature Selection and Engineering:

Feature selection helps improve prediction accuracy by focusing on the most impactful attributes.

**Correlation analysis** is performed to identify dependencies between different features.

**Principal Component Analysis (PCA)** is applied to reduce dimensionality while retaining important information.

**New feature creation:** Additional features are derived, such as skill proficiency levels based on project experience or weighted scores for extracurricular activities.

### 2.3 Web Application Development:

A user-friendly **React-based web application** is developed to provide career insights and recommendations.

#### 2.3.1 Frontend Development (React.js)

- Users can input their academic records, skills, and job preferences through an interactive UI.
- The dashboard displays **career recommendations, real-time job insights, and skill development resources**.

#### 2.3.2 Backend Development (Python Flask or Django)

- Handles user authentication, profile management, and communication with the machine learning model.
- Connects to a **MongoDB Atlas or SQL database** for storing user profiles and career history.
- Implements API integration for fetching real-time job market trends.

### 2.3.3 Skill Development Module

To enhance employability, the platform suggests skill development resources: API integration with online learning platforms (Coursera, Udemy, LinkedIn Learning) Personalized recommendations for courses based on career path and skill gaps Tracking of user progress in skill development activities

## 3. Machine Learning Algorithms

### 3.1 SVM:

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression tasks, with its primary application in classification problems. It efficiently distinguishes between different categories by mapping data points into an n-dimensional space, where n represents the number of features. Each feature's value corresponds to a coordinate in this space, and the algorithm finds the optimal hyperplane that best separates the data into distinct classes.

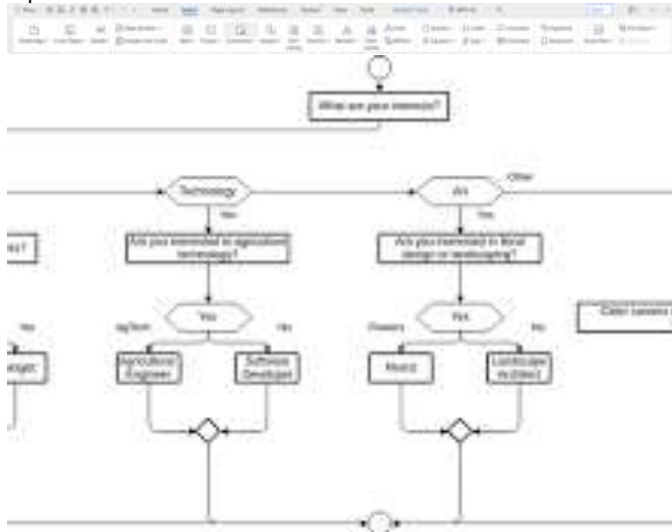


Figure 3: Support Vector Machines Example

SVMs are implemented using *kernels*, which transform the input data into higher-dimensional spaces to make classification easier. There are three types of SVMs, and in *linear SVM*, the hyperplane is determined using linear algebra transformations. The fundamental concept behind SVM is that it can be reformulated using the inner product of two observations. The inner product, which is the sum of the multiplication of each pair of inputs, is represented as:

$$f(x)=B0+\sum(ai \cdot (x,xi))$$

Instead of using a simple dot product, a polynomial kernel can be applied for improved classification performance:

$$K(x,xi)=1+\sum(x \cdot xi)^d$$

For more complex decision boundaries, a radial basis function (RBF) kernel can be used, which is defined as:

$$K(x,xi)=exp(-\gamma \sum (x-xi)^2)$$

### 3.2 XG Boost:

XGBoost denotes eXtreme Gradient Boosting. XGBoost is an implementation of gradient boosting algorithms and is available in multiple forms, such as tools and libraries. It primarily focuses on improving model performance while significantly reducing computational time. This implementation incorporates features from scikit-learn and R, along with additional enhancements like regularization. Regularized gradient boosting includes both L1 and

L2 regularization techniques, improving the overall stability and accuracy of the model.

Key features of XGBoost include automatic handling of missing values through a sparse-aware implementation, block structure to promote parallel tree construction, and continued training, allowing further boosting of an already fitted model with new data. Gradient boosting is a technique where new models are trained to predict the errors (or residuals) of previous models and are then combined to improve the final prediction. It uses gradient descent algorithms to minimize loss during the addition of new models. XGBoost supports both classification and regression tasks.

### 3.3 Decision Tree:

Decision Trees are one of the simplest yet widely used machine learning algorithms for classification problems. They serve as the foundation for many advanced algorithms such as bagging, gradient boosting, and random forest. An advanced version, XGBoost, enhances predictive performance through optimization techniques. Common decision tree algorithms include CART, C4.5, C5, and ID3.

In a decision tree, a node represents an input variable (X) and a split on that variable, particularly for numerical data. The terminal nodes, or leaves, contain the output variable (y), which is crucial for prediction. The decision tree follows a structured approach: first, it selects a root node and calculates the information gain or entropy for each node before splitting. The node with the highest information gain or lowest entropy is chosen for the next split. This process continues iteratively until further splitting is not possible or entropy reaches its minimum value.

Entropy measures the uncertainty or randomness in data, while information gain quantifies the reduction in entropy before and after a split. These principles enable decision trees to efficiently classify data and serve as a foundation for more advanced machine learning models.

## 4. Training and Testing:

After processing the data and training the model, the next crucial step is testing. This phase assesses the algorithm's performance, data quality, and output accuracy. Typically, 80% of the collected dataset is used for training, while the remaining 20% is reserved for testing.

Training enables the model to learn patterns from the given data, allowing it to make future predictions. In contrast, testing involves using a predefined dataset with labeled outputs to verify the model's accuracy. If the model achieves a high percentage of correct predictions, it is considered reliable for further use. Otherwise, modifications to the model or algorithm may be required to enhance accuracy.

Furthermore, new input sets and their corresponding predictions are continuously added to the dataset, improving the model's learning capability over time. This iterative process ensures enhanced accuracy and a more robust prediction system.

## 5.Result:

The data was trained and tested using three different algorithms, with Support Vector Machine (SVM) achieving the highest accuracy of 90.3%, followed by XGBoost with 88.33% accuracy. Since SVM

demonstrated the best performance, it was selected as the primary algorithm for all further predictions.

A web application was developed to accept student input parameters and generate career predictions based on the trained model. The background algorithm used in the application is SVM, and new predictions are continuously integrated to enhance accuracy.



**Figure 4:** Final Output Graphs

## 6.Future Scope:

A more advanced web application can be developed where inputs are not provided directly by students. Instead, student parameters are gathered through comprehensive evaluations and assessments. These assessments can include technical, analytical, logical, memory-based, psychometric, general awareness, interest-based, and skill-based tests. By collecting parameters through these structured evaluations, the system can generate more accurate results, making it more reliable and effective for career predictions.

## References

1. R. K. Sharma, P. Jain, "AI-Based Career Prediction Models Using Deep Learning," Proceedings of the IEEE International Conference on Artificial Intelligence and Machine Learning, 2020.
2. S. K. Singh, "A Study on Predictive Modeling for Career Guidance Using Machine Learning," International Journal of Computer Science and Artificial Intelligence, 2021.
3. Sudheep Elayidom, Dr. Sumam Mary Idikkula, "Applying Data Mining using Statistical Techniques for Career Selection," International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009.
4. Ms. Roshani Ade, Dr. P. R. Deshmukh, "An Incremental Ensemble of Classifiers as a Technique for Prediction of Student's Career Choice," 2014 First International Conference on Networks & Soft Computing.
5. Nikita Gorad, Ishani Zalte, "Career Counselling Using Data Mining," International Journal of Innovative Research in Computer and Communication Engineering.
6. Marium-E-Jannat, Sayma Sultana, Munira Akther, "A Probabilistic Machine Learning Approach for Eligible Candidate Selection," International Journal of Computer Applications (0975 – 8887) Volume 144 – No.10, June 2016.

7. Ali Daud, Naif Radi Aljohani, "Predicting Student Performance using Advanced Learning Analytics," 2017 International World Wide Web Conference Committee (IW3C2).
8. Bo Guo, Rui Zhang, "Predicting Students' Performance in Educational Data Mining," 2015 International Symposium on Educational Technology.
9. P. KaviPriya, "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques," International Journal of Advanced Research in Computer Science and Software Engineering.
10. Dr. Mahendra Tiwari, Manmohan Mishra, "Accuracy Estimation of Classification Algorithms with Demp Model," International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.
11. Rutvija Pandya, Jayati Pandya, "C5.0 Algorithm to Improve Decision Tree with Feature Selection and Reduced Error Pruning," International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 16, May 2015.
12. Shiju Sathyadevan, Remya R. Nair, "Comparative Analysis of Decision Tree Algorithms: ID3, C4.5, and Random Forest."
13. Yu Lou, Ran Ren, "A Machine Learning Approach for Future Career Planning."
14. Anuj Karpatne, Gowtham Atluri, "Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data," IEEE Transactions on Knowledge and Data Engineering, Vol. 29, No. 10, October 2017.
15. Gareth James, Daniela Witten, Trevor Hastie, "An Introduction to Statistical Learning with Applications in R."