

AI-Based Career Path Predictor Using Machine Learning

Submitted by

Shaik Mohammed Junaid

B.Tech CSE.

Contact:

Ph :- +91 8790039170

Email :- junaidbuss18@gmail.com

1. Objective

The goal of this project is to build an AI model that predicts the best career path for a student based on academic scores, skills, and interest area using machine learning techniques.

2. Tools & Technologies Used

- Python
- Pandas, Scikit-learn
- Random Forest Classifier
- Streamlit (Web App)
- CSV File Handling
- VS Code / Jupyter Notebook

3. Dataset Used

- Records: 1600 (Clean and Balanced)
- Features: Interest Area, Skills, 10th %, 12th %, UG %, Career Path
- Dataset Name: career_data_large.csv

4. Model Architecture

- Data Cleaning and Skill Encoding
- Label Encoding for interest area and output
- Combined all features for training

- Used RandomForestClassifier for multi-class prediction

5. Web Application (UI)

A front-end was built using Streamlit where users enter marks, interest area, and skills. The app returns a predicted career path using the trained model.

6. Results & Accuracy

- The trained model achieved 100% accuracy on the cleaned dataset
- Realistic predictions were validated with test inputs
- Output was displayed through a user-friendly web interface

7. Learnings & Takeaways

- Learned end-to-end AI pipeline: data collection, preprocessing, model training, and UI development
- Improved confidence in AI tools and Python libraries
- Understood how machine learning can be applied to real-world problems

8. Conclusion

This project enhanced my understanding of AI development and gave me hands-on experience in solving a real-world problem using ML and Python.

9. Key Python Code (Simplified) # Load

and preprocess data df =

```
pd.read_csv("career_data_large.csv")
```

```
df['Skills'] = df['Skills'].apply(lambda x: [skill.strip() for skill in x.split(',')]) mlb =
```

```
MultiLabelBinarizer() skills_encoded = pd.DataFrame(mlb.fit_transform(df['Skills']),
```

```
columns=mlb.classes_) le_interest = LabelEncoder() df['Interest_Encoded'] =
```

```
le_interest.fit_transform(df['Interest Area'])
```

```
X = pd.concat([df[['10th %', '12th %', 'UG %', 'Interest_Encoded']], skills_encoded], axis=1)
```

```
le_target = LabelEncoder() y = le_target.fit_transform(df['Career Path'])
```

```
# Train model model =
```

```
RandomForestClassifier()
```

```
model.fit(X, y)
```

```
# Predict def
```

```
predict_career(input_data):
```

```
    interest_encoded = le_interest.transform([input_data['Interest Area']])[0]    skill_vector = [1 if skill in
input_data['Skills'] else 0 for skill in mlb.classes_]    features = [input_data['10th %'], input_data['12th
%'],    input_data['UG    %'],    interest_encoded]    +    skill_vector    return
le_target.inverse_transform(model.predict([features]))[0]
```

10. Screenshots

Terminal Screenshot: (Files in the Folder & Run Streamlit)

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\junai> cd C:\Users\junai\Desktop\Edunet
PS C:\Users\junai\Desktop\Edunet> dir

    Directory: C:\Users\junai\Desktop\Edunet


Mode                LastWriteTime         Length Name
----                -
-a-----         13-07-2025   21:08             843 ABSTRACT.txt
-a-----         13-07-2025   21:42            1769 career_app.py
-a-----         13-07-2025   21:04            9357 career_data.csv.xlsx
-a-----         13-07-2025   21:29           88217 career_data_large.csv
-a-----         13-07-2025   21:36            2050 career_predictor.py
-a-----         14-07-2025   19:11              0 Overview.txt

PS C:\Users\junai\Desktop\Edunet> streamlit run career_app.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.0.3:8501
```

Before Entering the Data -User Interface (Career Path Predictor)



Career Path Predictor (AI Project)

Enter your academic info, interest, and skills to get a suggested career path.

10th Percentage

80

50100

12th Percentage

80

50100

UG Percentage

80

50100

Interest Area


Data

Select your skills

Choose an option

Predict Career Path

After Entering the Data -User Interface (Career Path Predictor)



Career Path Predictor (AI Project)

Enter your academic info, interest, and skills to get a suggested career path.

10th Percentage

95

50100

12th Percentage

88

50100

UG Percentage

94

50100

Interest Area

Data

Select your skills

C++ xJava xJavaScript xPython xSQL x

Predict Career Path

✓

Based on your input, the suggested career path is: Data Scientist