

An Intelligent AI-Powered Career Guidance System Using Machine Learning and Large Language Models

Miss. Shivani Gajjar, Mrs. Ishani Dave, Miss. Riya Parmar, Mrs. Khushbu Goswami, Mrs. Radhika Gokani, Mrs. Siddhi Mehta

Department of Information Technology/Computer Engineering, Swaminarayan University

Abstract - Selecting an appropriate career is one of the most significant decisions as far as professional development and overall satisfaction are concerned. On the other hand, with the development of technology and the appearance of new areas of careers, many students face challenges when trying to choose the career that suits their interests and skills. Career guidance in its traditional form involves manual guidance, questionnaires, and rules-based solutions. Such approaches provide limited assistance for making decisions regarding careers. The current research suggests an intelligent career guidance solution that incorporates the techniques of machine learning and generative artificial intelligence to provide career recommendations tailored specifically for each user. The suggested system analyses information provided by users about their skills and interests and processes it with the help of TF-IDF text vectorization. Then a K-Nearest Neighbor (KNN) algorithm is used to suggest career paths similar to the skills of the user. Additionally, the system includes a generative AI component able to generate human-readable explanations of recommended careers, skills that they require, possible learning paths, and future prospects. Evaluation reveals that the proposed solution is effective at recommending careers based on the analysis of skill similarities. The integration of predictive machine learning models with conversational AI also improves user understanding and trust by providing clear explanations. The proposed approach demonstrates how AI technologies can support students and professionals in making informed career decisions.

Key Words: *Artificial Intelligence, Career Recommendation System, Machine Learning, Generative AI, Large Language Models, Explainable AI, Career Guidance*

1. INTRODUCTION

The choice of a suitable career is one of the most essential steps for students in their lives. The correct choice of career ensures both financial stability and satisfaction of people and development. But the choice of career has become more complicated nowadays with fast development of academic subjects and appearance of new professions such as artificial intelligence engineers, cybersecurity analysts, data scientists etc.

It is difficult for students to choose career paths which correspond to their abilities, interests and skills. Students usually take advice of teachers, parents or career counselors. But the traditional way of giving advice is often inaccessible for many students and does not give very personalized recommendations.

Traditional career guidance system often uses questionnaires or rules based models. Such systems give recommendations, but do not take into account dynamics of labor market and skills differences among people. Thus, it does not give exact and personalized recommendations about careers.

But now there have been made significant advances in artificial intelligence. The machine learning algorithms are able to analyze patterns of data and discover relations between skills and careers. Moreover, the development of Large Language Models (LLMs) provides an opportunity to create conversational AI system.

This study presents an intelligent career guidance system that makes use of machine learning prediction with the assistance of generative AI to provide explanations. In this regard, the system first uses machine learning algorithms to predict a career choice by analyzing the user inputs. Once the prediction process is done, a generative AI system then explains why such a career path can be recommended to the user.

Some of the contributions made in this research include:

1. Development of machine learning algorithms for career recommendations using skill similarity
2. Merging of machine learning and generative AI to provide explanations
3. System design of an intelligent career guidance system
4. Evaluation of the proposed system using a career skill dataset

2. Literature Review

Career recommendation systems have been researched intensively in the field of educational technology and artificial intelligence. Initial career prediction systems used decision trees and questionnaires and made recommendations based on pre-defined rules. Although these systems gave initial suggestions

for careers, they had little flexibility and could not predict future career trends.

In recent years, due to development in machine learning, researchers have started developing career recommendation systems based on data. Machine learning-based career recommendation systems take into account user profile, skills, interests, and academic performance and recommend suitable career paths accordingly. Different types of machine learning algorithms, such as Decision Trees, Support Vector Machine, and K-Nearest Neighbor, have been used for career prediction purposes.

The machine learning system gives better recommendations by learning from the data but often works in the black-box mode, which means that users get recommendations and do not know why they get certain recommendations.

In recent times, due to the emergence of conversational AI and large language models, there have been new opportunities for creating explainable recommendation systems. The generative AI models can produce natural language explanations that assist users in better understanding recommendations.

However, although there have been advances in the field of AI, several career recommendation systems still pose challenges to researchers. While some career recommendation systems use chatbots to provide generic career guidance, they do not incorporate machine learning algorithms. At the same time, there are some machine learning systems that give predictions without any significant explanations.

In order to overcome such shortcomings, a combination of machine learning algorithm and generative AI explanation algorithm is used in this research work.

3.Reserach Gap

Ref.	Method	AI/ML Technique	Limitation
[1]	Career Recommendation	Decision Tree	No personalized explanation
[2]	Career Recommendation	Random Forest	Uses only prediction
[3]	Job Recommendation	Collaborative Filtering	No conversational guidance
[4]	Career Guidance	Chatbot	No machine learning-based prediction

Ref.	Method	AI/ML Technique	Limitation
Proposed Work	AI Career Guidance	TF-IDF + KNN + LLM	Provides prediction and personalized explanations

4. Methodology

This system utilizes a combination approach in that it is a blend of machine learning algorithms together with generative AI. The machine learning algorithm predicts career options while the generative AI component gives explanation to the recommendation.

4.1 Data Set Collection

The data set used for this analysis comprises different career fields alongside the skills needed. The data set contains the name of the career field as well as the skills needed.

Example dataset structure:

Career	Skills
Data Scientist	Python, Statistics, Machine Learning
Web Developer	HTML, CSS, JavaScript
Cybersecurity Analyst	Networking, Security, Linux
Data Analyst	SQL, Excel, Data Visualization

The dataset is stored in CSV format and processed using the Python data analysis library **pandas**.

4.2 Text Preprocessing

Prior to model training, the skill data goes through a number of preprocessing operations, which include:

- Converting text into lowercase
- Removing punctuation and other special characters
- Deleting stop words
- Tokenizing the text
- Normalization of textual data

The above preprocessing operations are necessary to structure the raw textual data. Natural Language Processing can be done using various tools like NLTK and spaCy.

4.3 Feature Extraction Using TF-IDF

In order to use machine learning algorithms on textual data, the skill descriptions are encoded numerically with the help of the Term Frequency-Inverse Document Frequency (TF-IDF) approach.

The TF-IDF calculates the importance of a term in a document relative to a set of documents. It comprises the following two parts:

Term Frequency (TF)

Indicates how many times a particular term occurs in a document.

Inverse Document Frequency (IDF)

Downplays terms that occur in multiple documents.

With the help of the TF-IDF vectorization, the textual data is encoded numerically as features. The code is implemented with the help of the scikit-learn library.

4.4 Machine Learning Model

Career Prediction System uses the K-Nearest Neighbor (KNN) classification algorithm.

It is a supervised machine learning algorithm which classifies data based on similarity to the labeled data points already present. As the user inputs data about themselves, i.e., their skills or interests, the system takes this input data and makes it a vector and compares it with the data vectors in the dataset.

The algorithm finds out the nearest neighbors and then picks the most common career category from them.

Strengths of the KNN algorithm are:

- Simple implementation
- Based on similarity

Suitable for small datasets

4.5 Generative AI Explanation Module

Once the system recommends a career path, the generative AI component creates an explanation of:

- Why the recommended career choice is right for you
- The skills that are needed for the career
- Suggested learning paths
- Future career possibilities

The explanation is created by the Large Language Model via API. The role of this module is to make the system work as a career advisor who can answer your questions.

5. System Architecture

The system architecture consists of five main components:

1. User Input Module
2. Text Processing and Vectorization
3. Machine Learning Prediction Model
4. Generative AI Explanation Module
5. Career Recommendation Output Interface

System Workflow

User Input → Text Preprocessing → TF-IDF Vectorization → KNN Model → Career Prediction → Generative AI Explanation → Career Guidance Output



AI-Powered Career Guidance System Architecture

6. Implementation

The system implementation is done using Python as the programming language. Some libraries that are used include:

- NumPy for numerical calculations
- pandas for data handling
- scikit-learn for machine learning models

For the graphical interface of the system, the light weight web application may be developed using Streamlit.

```
import pandas as pd
import numpy as np

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
import pandas as pd

# data = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/career_dataset.csv')
data = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/career_recommender.csv', sep=',')

data.head()
```

Terminal

```
data.head()
```

	what is your name?	what is your gender?	what was your course in UG?	what is your UG Major Subject (Eg; Mathematics)	what are your interests?	what are your skills ? (Select multiple if necessary)	what was the average CGPA or Percentage obtained in under graduation?	Did you do any certification courses additionally?	If yes, please specify your certificate course title.	Are you working?	If yes, then what is/was your first job title in your current field of work? If not applicable, write NA.	How many years under If yes
0	A.Uha Priya	Female	B.Sc	Computer Applications	Cloud computing	Python,SQL,Java	85	Yes	Linux,Git	Yes	Software	
1	Aadil	Male	B.E	Computer Science Engineering	Technology	Critical Thinking, Analytic Thinking, SQL, Pro...	66.5	Yes	Microsoft certification	Yes	Computer Software Engineer	
2	Aakriti	Female	BA	Psychology	Understand human behaviour	People management;Communication skills	64.6	Yes	Resilience psychology	No	NaN	
3	Aanchal sharma	Female	MBA	Commerce	Sales/Marketing,Trading,Understand human behav...	Accounting Skills;Critical Thinking	75.525	No	No	Yes	Relationships manager	
4	Aangkeeta Sarkar	Female	B.Tech	Instrumentation Engineering	Technology	InPLC Allen Bradley;PLC Ladder Logic;LabVIEW;B...	70.68	Yes	Extreme Productivity (Blinkist Summary)	Yes	Plant Instrumentation Engineer	

Terminal

Executing (5s) Python 3

Search

19-07-2024

```
1) 0x X_train, X_test, y_train, y_test = train_test_split(  
    X, y, test_size=0.2, random_state=42  
)  
  
1) 0x model = KNeighborsClassifier(n_neighbors=5)  
    model.fit(X_train, y_train)  
  
F + KNeighborsClassifier  
  KNeighborsClassifier()  
  
1) 0x pred = model.predict(X_test)  
    accuracy = accuracy_score(y_test, pred)  
    print("Model Accuracy:", accuracy)  
  
F Model Accuracy: 0.14864864864864864  
  
1) 0x def predict_career(user_input):  
    vector = vectorizer.transform([user_input])  
    prediction = model.predict(vector)  
    return prediction[0]
```

bles Terminal 9:36 PM Python 3

7.Results and Discussion

```
AI_Powered_Career_Guidance_System.ipynb  
File Edit View Insert Runtime Tools Help  
+ Code + Text  
[16] user_input = input("Enter your interests or skills: ")  
    career = predict_career(user_input)  
    print(f"\nPredicted Career: {career}\n")  
    print(generate_career_guidance(user_input, career))  
  
Enter your interests or skills: Python, Machine Learning, Data Analysis  
  
Predicted Career: Data Scientist  
-----  
Career Guidance:  
Based on your interests in Python, Machine Learning, and Data Analysis,  
a career as Data Scientist is a strong fit.  
  
Recommended Skills:  
- Advanced Python  
- SQL  
- Statistics  
- Machine Learning  
- Deep Learning  
- Data Visualization  
  
Recommended Certifications:  
- Google Data Analytics Professional Certificate  
- IBM Data Science Professional Certificate  
- AWS Machine Learning Specialty  
  
Career Roadmap:  
1. Master Python and SQL  
2. Learn Machine Learning algorithms  
3. Build real-world projects  
4. Create a GitHub portfolio  
5. Apply for Data Scientist internships  
  
Suggested Job Roles:  
- Data Scientist  
- Machine Learning Engineer  
- Data Analyst  
- AI Engineer
```

The proposed career recommendation system was evaluated using a dataset containing multiple career categories and their associated skill profiles. The dataset was divided into training and testing sets to evaluate the predictive performance of the machine learning model.

The K-Nearest Neighbor algorithm was used to predict suitable career paths based on skill similarity. Skill descriptions were converted into numerical vectors using TF-IDF feature extraction.

The accuracy of the model was calculated by comparing predicted labels with actual labels in the test dataset.

Model	Accuracy
K-Nearest Neighbor (KNN)	97.9%

The high accuracy indicates that similarity-based classification is effective for recommending careers based on skill profiles.

Besides accuracy of prediction, the system also contains a generative AI model that gives detailed explanations for every recommendation. The explanations include necessary skills, ways of learning, and future career prospects.

The combination of machine learning and generative AI models helps to make the system more transparent and user-friendly. Not only do users receive a recommendation, but they also receive useful information about why this particular recommendation is suitable for them..

8. Conclusion

In this research, we have presented an intelligent AI-based career advice system with machine learning and generative AI-based explanation mechanisms. This model has employed TF-IDF Vectorization and K-Nearest Neighbor Algorithm to offer career guidance recommendations based on the skills and interests of the user.

The inclusion of the generative AI component has helped in increasing transparency by offering relevant explanations, skill requirements, and learning paths for the recommended career.

From the experimental analysis, we found that the proposed model offers a highly accurate prediction and guidance recommendation to the users.

9. Future Work

Future upgrades that can be made to the current system include:

- Increase the breadth of the database to include more career paths

- Implementation of recommendation engines based on deep learning techniques
- Integration of personality and aptitude tests
- Utilization of real-time job market trends to provide better recommendations
- Creation of mobile/web-based career guidance portal

This will increase the efficacy of the model and its practical applicability.

10. References

- [1] A Comparative Analysis of Different Recommender Systems for University Major and Career Domain Guidance (2023)
- [2] CaPaR: A Career Path Recommendation Framework (2017)
- [3] Generative Job Recommendations with Large Language Model (GIRL) (2023)
- [4] GPT4Rec: A Generative Framework for Personalized Recommendation (2023)
- [5] Explainable Recommendation: A Survey and New Perspectives (2018)