Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.8, May : 2024

CAREER GUIDANCE SYSTEM BY MACHINE LEARNING

Ayush Kumar Tiwari Student, Department of Computer Science & Engineering, ABES Engineering College, Ghaziabad
Prakhar Tandon Student, Department of Computer Science & Engineering, ABES Engineering College, Ghaziabad
Pratham Singhal Student, Department of Computer Science & Engineering, ABES Engineering College, Ghaziabad
Utkarsh Dixit, Assistant Professor, Department of Computer Science & Engineering, ABES Engineering College, Ghaziabad

Abstract
The machine learning (ML)-based career guidance system is a state-of-the-art tool that has the potential to completely transform the way people choose and make decisions about their career paths. This system uses ML algorithms to provide accurate and personalised career suggestions in response to the increasing complexity of profession options and the wide range of factors influencing career success.
Through the use of an extensive amount of historical data that includes abilities, academic achievements, career paths, and personal preferences, the system builds dynamic prediction models. These models ensure relevance and flexibility by taking into account both individual profiles and the dynamic structure of the labour market. The user-centric method creates recommendations that are in line with each user's particular abilities and goals by carefully examining academic performance, talents, and preferences.
This Career Guidance System stands out due in part to its focus on predictive analytics. ML algorithms are used to find relationships between different influencing elements and real-world career outcomes, providing consumers with useful information about the chances of success in particular sectors. People are empowered to make strategic decisions that are in line with their goals and strengths because to this predictive aptitude.
Adaptation and ongoing learning are essential components of the system. The system improves its models through an iterative feedback loop by using user interactions and actual results. This flexibility guarantees that, despite changing employment environments, the advice given is relevant and up to date.
Users are actively engaged by the system, which gives them the opportunity to explore several career pathways and offer comments. This interactive feature improves user satisfaction and provides a useful source of data for improving the machine learning models. The system becomes an interactive tool for career exploration and decision-making by encouraging a collaborative and user-driven approach.
In summary, the machine learning-based career guidance system is at the vanguard of innovation when it comes to helping people navigate the complexity of choosing a job. Its incorporation of user participation, continuous learning, and predictive analytics builds a strong foundation for providing personalised and pertinent advice, ultimately enabling people to start fulfilling careers.

I. Introduction

The complexity of today's work environment demands that people make well-informed judgements in order to pursue successful careers. One prospective solution to this problem is the incorporation of machine learning into career advising systems. The goal of this project is to create a sophisticated Career Guidance System (CGS) that can provide customised and flexible career recommendations by applying machine learning techniques.
The convergence of individualised recommendations and data-driven insights has the potential to completely rethink current methods of career counselling by providing people with a more flexible and customised framework. In order to provide insightful information about succeeding in particular

sectors, the CGS analyses past career data, adjusts to individual user profiles, and forecasts possible career success.
The project's core principles of continuous learning and adaptation guarantee that the CGS changes in response to user interactions and practical results. Through active user engagement, the system encourages a collaborative and user-driven approach to career discovery by enabling users to explore a variety of career options and provide comments.
Moreover, the CGS's capabilities are improved by its connection with educational institutions. The system guarantees alignment with the most recent educational opportunities and industry demands by merging academic performance data, course offerings, and business relationships. This enhances the accuracy of the system's suggestions.
Essentially, the goal of this initiative is to completely transform how people choose their career paths. Through the integration of machine learning, data-driven insights, and user involvement, the Career Guidance System aims to become a flexible and dynamic instrument that enables people to make well-informed choices and pursue rewarding careers.

II. Literature Review

Guidance is to support individuals in forging their own life paths by assisting them in making greater use of their own skills and abilities (Peavy, 2000). The Council of the European Union (2008) states that advice includes a range of individual and group actions such as providing information, counselling, assessing competence, helping, and imparting skills in career management and decision-making.
Activities that pertain to both individual and group counselling fall under this description. Information sharing, assisting clients with self-assessments, and offering professional guiding staff for therapy are all included in the offerings. The importance of reskilling and upskilling within further education has taken center stage in career advice in recent years (Toni & Vuorinen 2020).
People may access a wider selection of services and new methods of getting assistance whenever and wherever they need it thanks to digital tools. Increased accessibility, expanded networks and information availability, assessments, decreased overall costs, and enhanced cost-effectiveness are some possible advantages of employing technology in career counselling (Sampson et al., 2020). The ongoing pandemic has made digital and remote advisory services more necessary.
In the past, technology has been used by guidance counsellors in three ways: 1) to support career information and learning, 2) to enable automated involvement through career examinations, simulations, or games, and 3) to offer communication options (Hooley et al., 2015).
To provide integrated or blended guidance, or guidance via digital means, guidance experts and service designers need to determine which technologies to use and how to use them (Bakke et al., 2018).
According to this study, intelligent entities constitute artificial intelligence that sense their environment and take action to change it, as defined by Russell and Norvig (2016). These virtual creatures may mimic mental functions like understanding, picking things up, deductive reasoning, and solving problems.
New advancements in AI are expected to open up new avenues for teaching and learning, and they will also have a big impact on competency requirements and future job markets (Tuomi, 2018). According to research and review studies, artificial intelligence (AI) can be used in education to support a variety of functions, including student self-regulation, motivation and well-being, personalised learning support and feedback, learning process support, assessment and evaluation, profiling and prediction, usability and accessibility, resourcing, and competence management.

III. Proposed Methodology

3.1 Problem description

After graduation, engineering students often face a significant degree of uncertainty and struggle while deciding on a career path. There are several disadvantages to the present methods of career counselling, which rely on mentors, parents, teachers, or professional job counsellors. These include potential biases, unequal accessibility, and a lack of subject-matter competence. Students may not be able to reach their full potential in their selected fields of study as a result of this ambiguity, which could lead to less than ideal career choices.

3.2 Data Gathering

The methodical process of obtaining and evaluating information from a range of sources is known as data collection.

A variety of sources to obtain an accurate and comprehensive picture of a subject. By gathering data, an individual or organization can assess results, anticipate future trends and probabilities, and provide pertinent answers to queries.

3.3 Preparing Data

Any type of processing applied to unprocessed data in order to prepare it for further processing is referred to as data preparation. Data preprocessing, which is frequently used as an initial step in data mining, converts the data into a format that can be processed by the user more quickly and efficiently.

3.4 Investigative Data Analysis

3.4.1 Histograms and Scatterplots

To illustrate the pairwise interactions and correlations between the various variables, we will first create a scatterplot matrix. A brief summary of the distribution of the data and the presence or absence of outliers is also very helpful.

3.4.2 Matrix of Correlations

In order to measure and summarize the correlations between the variables, we have produced a correlation matrix.

This correlation matrix and the covariance matrix are closely connected; in fact, the correlation matrix is simply the covariance matrix rescaled using standardized features. There are exactly the same amount of columns and rows in this square matrix, which holds the Person's r correlation coefficient.

3.4.3 Algorithms for Regression

Modelling a target value based on independent factors is called regression. The main applications of this method are in forecasting and figuring out cause-and-effect relationships between different variables. The number of independent variables and the nature of the relationship between the independent and dependent variables are the main factors that determine how different regression approaches work.

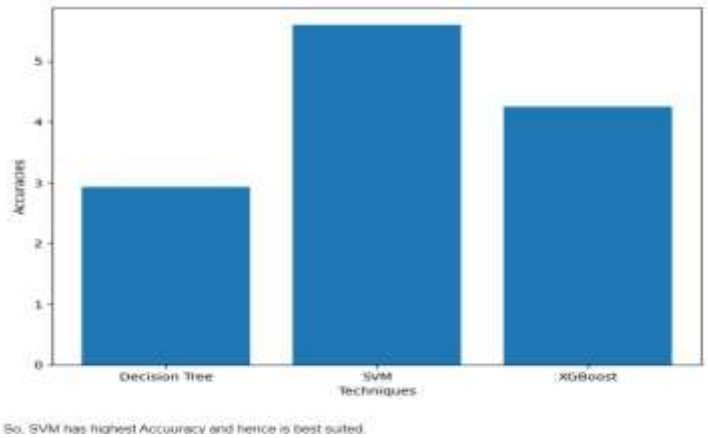
3.4.4 The Linear Regression

Simple linear regression is a type of regression analysis that happens when there is just one independent variable (x) and a linear connection between the dependent variable (y). In the graph above, the red line is the most appropriate straight line. Our goal is to create a line that most accurately depicts the given data points. The line can be represented by the linear equation below.

y = a_0 + a_1 * x Linear Equation

The objective of the linear regression process is to determine the optimal values for a_0 and a_1. Before we go into the technique, let's take a closer look at two fundamental concepts you need understand about linear regression.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.8, May : 2024



So, SVM has highest Accuuracy and hence is best suited.

3.4.5 Regression using Decision_Trees

The decision_tree is a fundamental machine_learning_model for starting regression_tasks. Every internal (non-leaf) node in a decision tree represents a test on an attribute; every branch in a decision tree reflects the test's result; and every leaf (or terminal) node has a class label. The structure is similar to a flowchart. In a tree, the root_node is the highest_node.

Different model types have different advantages.

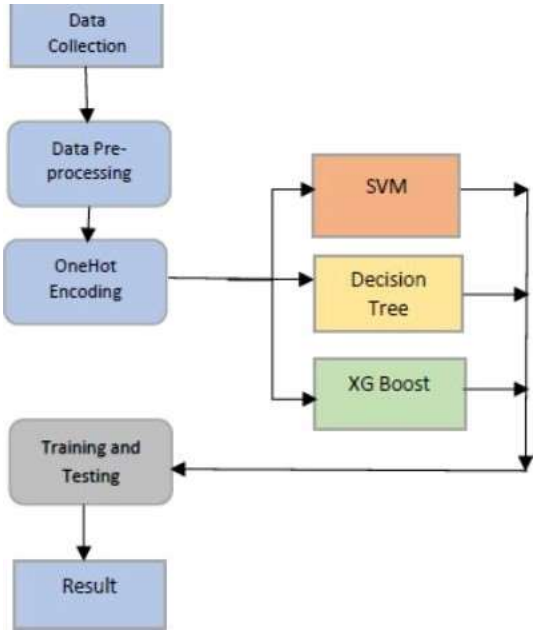
The decision tree model performs exceptionally well when handling numerical characteristics in tabular data or categorical features with fewer_than hundreds of_categories. Decision trees, in contrast to linear models, are able to capture interactions that are not linear between features and the target.

3.4.6 The Algorithm of Random Forest

One of the best machine learning models for predictive analytics is Random Forest, which is a workhorse of the industry. The random forest model, which creates projections by combining selections from a number of base models, is an illustration of an additive model. This class of models can be described formally as follows: where basic base models are the product of the final model. Each base classifier in this case is a simple decision tree. The technique of employing many models to increase prediction performance is known as "model_ensembling." In random_forests, each base model is constructed_independently using a different subset of the data.

IV. Work Flow Diagram

Figure 3.3. I An overview of several sophi sticated algorithms for machine learning



V. Practical Implementation

5.1 Data Collection:

Data gathering is one of the primary and most important tasks of every machine learning project. Given that data is provided to the algorithms as input. Therefore, the accuracy and efficiency of the algorithm depend on the precision and quality of the data that is collected. The outcome will therefore match the facts. To forecast a student's career, many factors are required, such as their academic standing in a range of subjects, specializations, programming and analytical skills, memory, personal information about relationships, sports, competitions, hackathons, workshops, certifications, books they're interested in, and many more. They are all considered because they play a crucial role in predicting a student's future professional success. There are numerous ways to collect data. A portion of the data is created randomly, while other portions are acquired via the LinkedIn API, a database of recent graduates, and staff members of various businesses. About 20,000 records total, consisting of 36 data columns, are gathered.

5.2 Data Pre-processing:

All of these circumstances must be verified and replaced with new values in order for the data to have any real worth and to be pertinent for further processing. Maintaining structured data is necessary. As crucial as collecting data is the challenge of giving the data significance. A variety of methods can yield disorganized, erroneous, and unwanted data, as well as a large number of null values. The main processes in data preparation are cleaning up all of these data, replacing them with appropriate or approximation data, and deleting null and missing data and replacing them with some specified replacement values. Even collected data may contain values that are completely meaningless. In terms of format and style, it could not be exactly as anticipated.

5.3 One Hot Encoding:

A method is called One Hot Encoding that transforms classification of values found in the gathered information into an ordinal or numerical format so that machine learning algorithms can use them to make predictions that are more accurate. Just to recap Categorical values are transformed into the most appropriate form using One Hot encoding before being fed into several ML algorithms. Nearly all ML methods are in line with this algorithm. There aren't many algorithms that handle categorical values as well as random forest. One Hot encoding is not necessary in these circumstances. While the One Hot encoding process may appear challenging, the majority of contemporary machine learning algorithms handle it. Here is a basic rundown of the steps involved: As an example, an integer encoder can assign values like I and O to values in a data set that contain values like yes and no. As long as we

maintain the preset values for yes as I and no as O, this process can be continued. Integer encoding is used while we still allocate or assign these precise labels to these set numbers. However, in this case, consistency is crucial since, should we subsequently reverse the encoding, especially in the prediction situation, we need to be able to derive the appropriate labels from those integer values. Making a vector for every integer value is the next stage. Assuming that this vector is binary, its length would be two for each of the two possible integer values. Next, the zeroth index of a vector [l , l] is assigned the value I will represent the 'yes' label encoded as 1.1n a same manner, a "no" label with a "O" encoded will be rendered as [0,0], indicating that the value of the first index is O.

VI. Expected Outcome

After training and testing the data using all three algorithms, SVM yielded the highest accuracy (90.3%), followed by XG Boost (88.33%). All subsequent data predictions are chosen to be followed using S VM since it provided the best accuracy. Ultimately, the final forecast is generated and presented by a web application that is designed to receive the parameter centered by the students. The fundamental algorithm is S VM, and in order to improve accuracy, new predictions are constantly added to the dataset.

Overall, by utilising machine learning (S VM) to deliver data-driven and coustomised career suggestions for engineering students, our study has effectively addressed the shortcomings of conventional career counselling techniques. For students to get these recommendations, the online application provides an easily navigable interface. Additionally, by adding fresh predictions to the dataset, the system is continuously improved and able to adjust to the evolving demands of students and trends in the workforce.

```
In [1]: import numpy as np
        import pandas as pd
        url = "career_pred.csv"
        df1 = pd.read_csv(url)
        # Dataset is now stored in a Pandas Dataframe

In [2]: df1.head()
```

Out[2]:

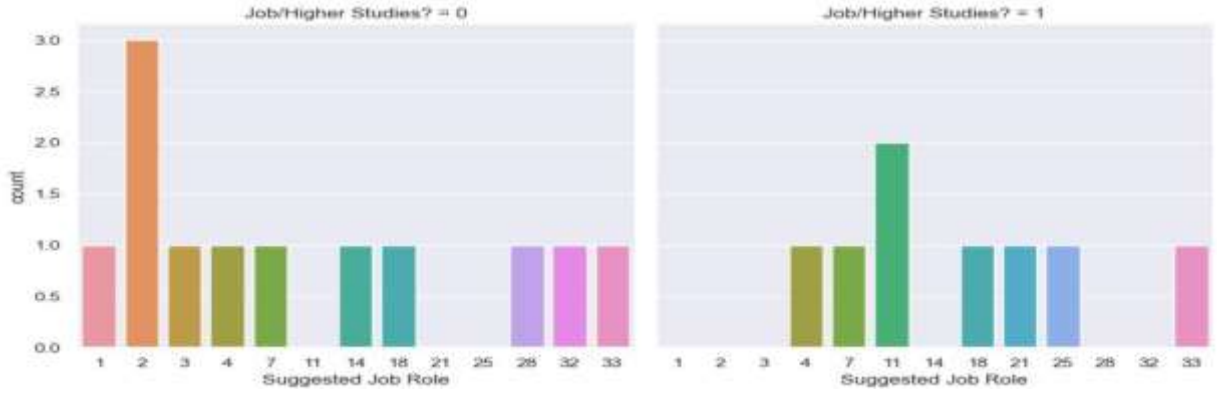
	Acedamic percentage in Operating Systems	percentage in Algorithms	Percentage in Programming Concepts	Percentage in Software Engineering	Percentage in Computer Networks	Percentage in Electronics Subjects	Percentage in Computer Architecture	Percentage in Mathematics	Percentage in Communication skills	Hours working per day	...	Interested Type of Books	Salary Range Expected
0	69	63	75	67	94	94	87	84	61	9	...	Prayer Books	salary
1	78	62	73	60	71	70	73	84	91	12	...	Childrens	salary
2	71	86	91	87	61	61	72	72	94	11	...	Travel	Work
3	76	87	60	84	89	73	62	86	69	7	...	Romance	Work
4	92	62	90	67	71	89	73	71	73	4	...	Cookbooks	salary

VII. FINAL SUMMARY

Students will be able to predict the appropriate course with the aid of this approach. This system aids in reducing the failure ratio and encouraging career-appropriate behavior.

Students can benefit from this approach since it will help them make the right decisions while selecting a career path. Additionally, this method will assist the college in analyzing the admissions branch and determining the appropriate course of action based on the outcome.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.8, May : 2024



VIII. Conclusion

Students will be able to predict the results of this system appropriate training. This method aids in reducing the failure percentage and to act in a way that is suitable for one's job. Students can benefit from this method since it will provide guidance. people to make the right choice when selecting the stream as their profession. Additionally, this technique will assist the university to examine the admissions division and take the necessary steps based on the outcome.

IX. References

[1]. PR Deshmukh, A Roshani, A method for predicting a student's profession choice using an incremental ensemble of classifiers. September 25, 2015, IEEE Networks & Soft Computing (ICNSC).
[2]. R Dymitr, C. Ling, and others. Big Data: Analytics Potential for Big Data. September 10, 2015, IEEE Digital Signal Processing (DSP).