Career Recommendation System based on Student Preferences

A.Madhavi[1], Chathala Swapna[2] , Bavikadi Mrudula[3] , Sunkoju Swathi[4]

[1] Assistant Professor – Computer Science and Engineering, VNR VJIET, Medchal, Hyderabad
[2] UG -Computer Science and Engineering, VNR VJIET, Medchal, Hyderabad.
[3] UG -Computer Science and Engineering, VNR VJIET, Medchal, Hyderabad.
[4] UG -Computer Science and Engineering, VNR VJIET, Medchal, Hyderabad.
Corresponding Author Orcid ID : https://orcid.org/0000-0002-8858-5239

Abstract
Choosing right career is a challenging decision for students because career is a significant period of one's life and with opportunities for progress. Due to their lack of understanding, students are unable to make an informed decision, which could cause problems down the road. Students should choose a career that will be the most beneficial to them in order to avoid issues in the future. Students rely on the arbitrary judgements of their family and friends because they have access to so many online resources and no guidance. The interests and abilities of the pupils may not match as a result. To address the aforementioned issues, our project develops a recommender system based on student preferences, which gathers information about the students and suggests a better profession field based on the talents, interests, and academic achievement supplied. In this research, we conducted a comparison of various algorithms, including SVM, Decision Tree, and Random Forest, and we took into consideration the model with the highest precision, recall rate, F1-score, and accuracy.
Keywords: Student preferences, Job domain, Recommendation system, SVM, Decision tree, Random Forest.

Introduction

Career recommendation systems have become increasingly popular in recent years as individuals face increasing choices and information overload when making career decisions. With the advent of artificial intelligence and machine learning algorithms [7], career recommendation systems have become more efficient and effective in providing personalized recommendations to individuals. The main goal of career recommendation systems is to assist individuals in making informed career choices that are aligned with their skills, interests, and goals [3]. These systems typically use a variety of inputs such as academic performance, work experience, skills, and personal preferences to generate career recommendations. Machine learning algorithms are then used to analyse this data and provide personalized recommendations to individuals based on their unique profile [6].

One of the most common types of career recommendation systems is based on clustering algorithms [11]. Clustering algorithms group individuals based on their preferences, skills, and interests, and then recommend careers that are most closely aligned with their cluster [3],[7]. Another common type of career recommendation system is based on decision trees. Decision trees use a set of rules to recommend careers based on an individual's input data. Other types of career recommendation systems include neural networks, support vector machines, and random forests. These systems are typically more complex and require larger amounts of data to generate accurate recommendations

One of the biggest advantages of career recommendation systems is that they provide personalized recommendations that are tailored to the individual's unique profile. This can help individuals make more informed career decisions and avoid costly mistakes. In addition, these systems can help

individuals explore careers they may not have considered otherwise. However, there are also some challenges associated with career recommendation systems. One of the biggest challenges is ensuring that the recommendations generated are accurate and reliable. This requires large amounts of data and complex algorithms that can be expensive to develop and maintain. In addition, some individuals may not be comfortable sharing their personal information with these systems, which can limit their effectiveness. Despite these challenges, career recommendation systems have the potential to revolutionize the way individuals make career decisions. By leveraging the power of artificial intelligence and machine learning algorithms, these systems can provide personalized recommendations that are tailored to an individual's unique profile. As technology continues to evolve, it is likely that career recommendation systems will become even more sophisticated and effective in helping individuals achieve their career goals [5].

Literature Survey
The survey paper reviews the resume and job description using stop words removal and Porter's stemming algorithm, followed by conversion of the text into a matrix using a tf-idf vectorizer [1]. The system then ranks the job descriptions based on their similarity scores, computes the cosine similarity between the job descriptions and resumes, and identifies skills that the user may need to improve upon by comparing their resume with a skills dataset [3].

An overview of a content-based job recommendation system using Cosine or Jaccard similarities [12] is provided in a survey article. Job skills are given higher weights than job domain while computing similarity scores. [5] The system involves web scraping job offers from sa.indeed.com, pre-processing the data through tokenization [2] and keyword extraction, and recommending jobs with high similarity scores for a job seeker's resume.

The survey paper presents utilizing the Systematic Literature Review (SLR) method to explore ITS characteristics and applications [3]. The research questions addressed include the recommendation system approach, AI techniques used [10], necessary features, and evaluation methods employed for career recommendation systems.

The paper provides various methods used for job recommendations include knowledge- based, collaborative, hybrid, and content-based filtering [12]. competitions, validations along with them regarding the reciprocal and time-based nature of Job referrals, Ethical aspects in job recommender system and JRS at scale: notes from LinkedIn [4] are also considered as a part of methodology.

About 40% of students are unsure of their employment prospects, according to a survey by the Council of Scientific [5] and Industrial Research (CSIR). This System would recommend to the student a based on their abilities in many subjects and locations, a career option [3]. The implementation steps include data collection, data pre-processing and using of XG Boost and Decision tree are examples of machine learning algorithms [4] used for prediction.

This paper presents a unique approach, that is they made an interactive platform which allows students to interactively perform [6] the task and get results. In this paper they have used many approaches for supervised classification like Logistic Regression [11], Decision Tree, KNN, Naive Bayes, SVM, Random Forest, Stochastic Gradient Descent, AdaBoost, XgBoost, and several hybrid algorithms involving stacking like SvmAda, RfAda, and KnnSgd.

The project involves creating a job recommendation engine based on two datasets: the jobs dataset and the user's dataset. Exploratory data analysis [7] is conducted in order to divide the data into train and test sets and display the data distribution using bar graphs. Collaborative filtering is then used to match job listings to a candidate's profile, utilizing item-item similarity and user-user similarity [16]. Finally, cosine similarity is applied to compare job descriptions with a person's job profile, specifically comparing the text keywords.

The article describes a Python-based a system for career recommendations that attempts to solve the issues that conventional collaborative [8] filtering systems have with cold start, trust, and privacy [10],[12]. In order to deliver appropriate and reliable career recommendations based on user inputs, the system makes use of a variety of Python modules and machine learning techniques, including

cosine similarity. To help students utilizing the system get better results, the system also has a feedback and remark part that uses NLP to assess whether input is favorable, negative, or neutral [11].

Data collection, data pre-processing, and classification made up the three stages of the research technique. Data from Saudi Arabian IT staff members were gathered using a survey [9]. In this pre-processing phase, data was cleaned, transformed, and feature engineered, with irrelevant features removed. The recommender model was constructed with the K-Nearest Neighbor method. with label encoding used for categorical variables. The model aimed determine for IT graduates the ideal career route based on their skills and academic profession.

This project aims to create a recommendation engine that matches job listings to a candidate's professional profile. Relevant information is extracted from two datasets: jobsand users. Exploratory data analysis is performed to distribute the data into training and test sets and visualize data distribution [1]. Collaborative filtering is used with item-item and user-user similarity, and cosine similarity is applied to compare job descriptions and skills in the job profile [11].

Existing system

The same problem is addressed by a variety of models in the current system, each with unique benefits and drawbacks. These techniques are used by some of the existing systems to collect input.

Prediction using course grades:

Using a student's academic course grades as input may not be the most effective approach because each student's grades may vary depending on their college, test models, paper evaluation, and other factors.

Questionnaires with a yes/no response option:

These are not recommended for use in prediction because they may confuse students who are asked questions from a variety of required categories. If a student just had rudimentary knowledge of the subject matter of a given course, he might not know whether to answer yes or no. Therefore, in this instance, he wouldn't supply input, which would result in an erroneous forecast.

Synthetic dataset used:

Online surveys, in-person interviews, focus groups, panel sampling, telephone, post-call, mail-in, pop-up, and mobile surveys are all common types of research methods.

However, these methods have their own limitations and drawbacks. For example, using course grades as input may not be efficient as different elements, including college, paper grading, and exam formats, can have an impact. Similarly, using YES/NO questionnaire can be confusing for some students, especially when they lack knowledge about a particular course or subject.

Proposed system

Various algorithms and techniques of machine learning are used to develop the proposed system. It is built using the framework development method, which is represented in Figure [1] and consists of four phases. Data extraction and feature selection, data cleansing, establishing a recommended model (career recommendation), and developing an interface are the processes.
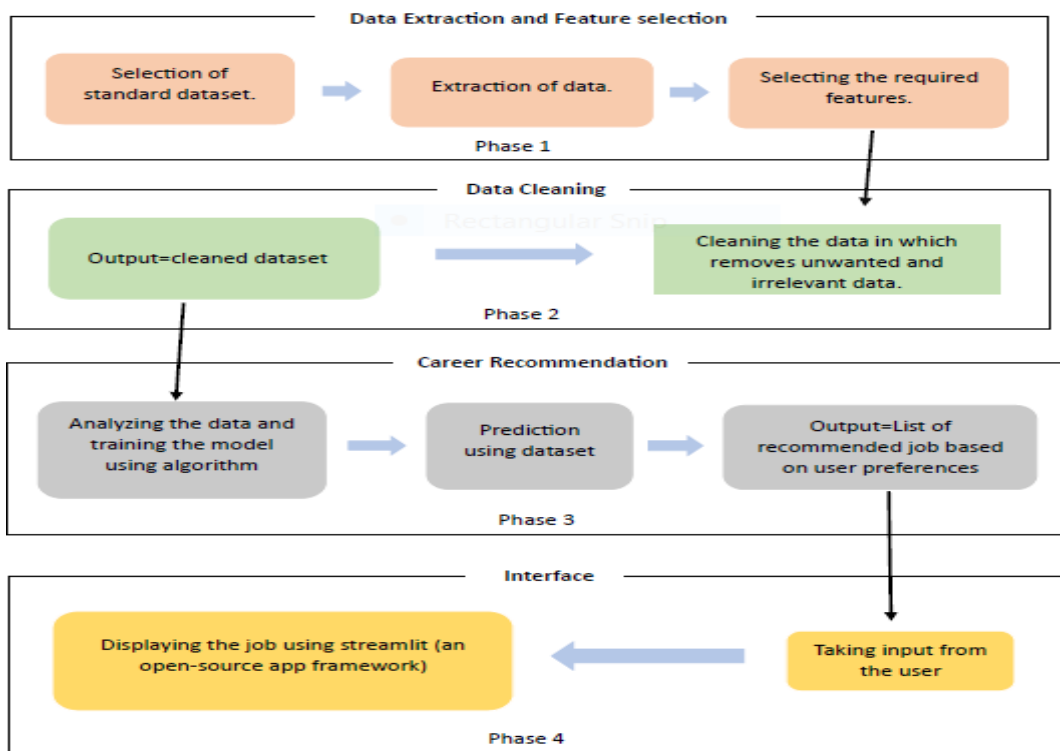
System Architecture

Figure :1 System Architecture

The **first phase** involves data extraction and feature selection. The dataset used for this phase is collected and consists of 6901 rows and 20 columns. The data is then extracted and processed to remove any irrelevant or unwanted information. Feature selection is also performed to choose the most important features that contribute to the prediction of the output. This is accomplished by the use of numerous techniques such as Pearson correlation, Linear Discriminant Analysis, Analysis of Variance (ANOVA), and the Chi Square Test.

In this particular project, the ANOVA test is used for feature selection. The Anova coefficient is calculated by comparing the mean sum of squares between groups with the mean sum squares due to error. The selected features are then used for further processing in the next phases of the system.

The **second phase** is data cleaning phase, we work on identifying and handling issues in the dataset that can potentially affect the accuracy of the model. The first step is to identify missing values and

decide on how to handle them, either by imputing them or removing rows with missing values. Next, we look for duplicates and outliers in the data and handle them accordingly.

We also remove columns that do not contribute significantly to the prediction of the target variable. Another crucial step is to check for data consistency and integrity, ensuring that the data values are within the expected range and format. Overall, the goal of data cleaning is to prepare a clean and reliable dataset for the subsequent modelling phase.

In **phase third** of our project, we implemented a career recommendation system that allows users to give input to the trained model. The system receives user input in the form of various parameters such as certifications, coding skills, workshops attended, and interested subjects. The algorithm offers an appropriate career for the user to pursue based on the input provided by the user.

The recommendations are generated by a machine learning model that was trained on a big collection of career data. The model uses the input parameters provided by the user to match them with the relevant careers from the dataset. The suggested careers are prioritised according to their relevancy to the user's input.

In the **fourth phase** of our project, we utilize the machine learning model that we trained in the previous phase. This phase takes inputs from the user interface built using Streamlit framework. The inputs are then fed to the trained model which makes predictions based on the input parameters.

To anticipate the ideal job role for the user, the trained phase use machine learning methods such as decision trees, random forests, or neural networks. This phase also involves the deployment of the trained model on a web server or a cloud platform, enabling users to access the service through the internet.

The trained phase plays a crucial role in the project, as it provides the end-users with a personalized career recommendation based on their unique interests, skills, and qualifications.

Feature selection is a crucial step in machine learning where the goal is to identify the most important features that contribute to the prediction of the output variable.

In the present work, Anova Test has been used as a filter method for feature selection. Anova Test is a statistical technique that helps in detecting the substantial differences between the means of two or more groups.

$F = MST/MSE$ is the general formula for the Anova Test, where F is the Anova coefficient, MST is the mean sum of squares between groups, and MSE is the mean sum squares attributable to error.

Anova Test compares the variation between groups to the variance within groups. If the variation between groups is much greater than the variance within groups, it indicates that the characteristic is significant and contributes to the prediction of the output variable. Feature selection is a way for selecting the best features that can contribute the most to the prediction of the outcome.

Results

Data Set

The dataset contains 20 columns and 6901 rows. Each row corresponds to an individual and the columns represent various features related to their skills, capabilities, interests, and preferences. The columns include features such as logical quotient rating, hackathons, coding skills rating, public speaking points, self-learning capability, extra-courses done, certifications, workshops, reading and writing skills, memory capability score, interested subjects, interested career area, type of company want to settle in, taken inputs from seniors or elders, interested type of books, management or technical skills, hard/smart worker, worked in teams ever, introvert, and the suggested job role which serves as the output label.

The suggested job role is the target variable that we want to predict based on the given features of an individual. The objective of this dataset is to recommend a suitable job role for an individual based on their skills, interests, and other relevant factors. This dataset used to train machine learning models to predict the suggested job role for a given individual.

- Number of records: 6901
- **Number of attributes:** 20

- **Column names:** 'Logical quotient rating', 'hackathons', 'coding skills rating', 'public speaking points', 'self-learning capability?', 'Extra-courses did', 'certifications', 'workshops', 'reading and writing skills', 'memory capability score', 'Interested subjects', 'interested career area', 'Type of company want to settle in?', 'Taken inputs from seniors or elders', 'Interested Type of Books', 'Management or Technical', 'hard/smart worker', 'worked in teams ever?', 'Introvert', 'Suggested Job Role'

- **Output label:** 'Suggested Job Role

Accuracy Graph

An accuracy graph is a graphical depiction of a machine learning model's performance. It plots the model's accuracy against a certain variable, such as the number of training iterations or the values of hyperparameters. By analyzing the accuracy graph, we can evaluate the performance of the model.

The Figure [2] shows that the Decision Tree and Random Forest models have the highest accuracy score of 87%, while the SVM model has an accuracy score of 83%.

This information is useful in evaluating and selecting the best-performing model for the career recommendation system.
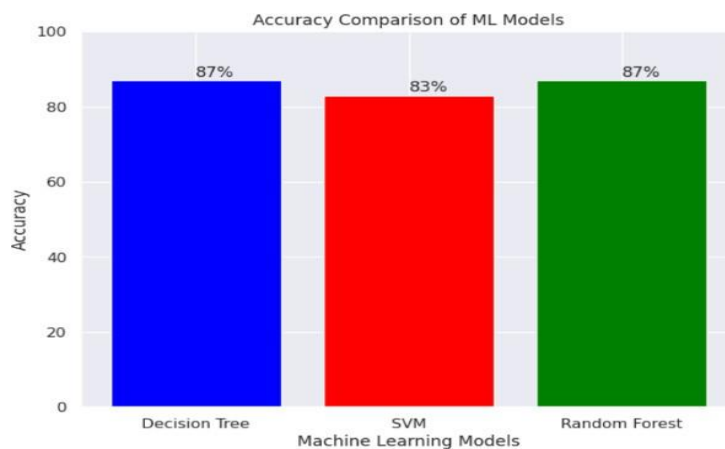

Fig [2] Comparison of algorithms

Evaluating the Performance

A binary classification model's performance is measured using measures such as precision, recall, and F1-score.

Performance / Algorithm	Precision	Recall rate	F1-Score
Decision Tree	0.0875	0.0868	0.0870
SVM	0.0835	0.0868	0.0829
Random Forest	0.0886	0.0868	0.0834

Table 1: Comparison analysis

The three machine learning models' accuracy, recall, and F1-score, i.e., Decision tree, SVM, and Random Forest. It seems [1] that the precision for all models is relatively low, around 0.08-0.09, which indicates that the models may have a high number of false positives.

The recall rate is almost identical for all models (0.0868), which indicates that the models have a similar ability to identify true positives. The F1-score, which is the harmonic mean of accuracy and recall, is also quite low, ranging between 0.08 and 0.09 for all models, indicating that the models may perform poorly overall.

Future-scope and Conclusion

In conclusion, career recommendation based on student preferences is a useful application of machine learning. By analyzing the preferences of students, such as their interests, skills, and personality traits, machine learning algorithms can suggest suitable career options that match their strengths and interests. Machine learning techniques such as Decision Trees, Support Vector Machines, and Random Forests are used to anticipate career suggestions based on student preferences. The models' accuracy was assessed using the confusion matrix, precision, recall, and F1-score, and the findings revealed that Decision Trees and Random Forests had the greatest accuracy.

It is crucial to remember, however, that the accuracy of these models is determined by the quality and quantity of data utilised to train them. Therefore, it is necessary to collect reliable and diverse data to train the models and continually update them to ensure their accuracy. Overall, career recommendation based on student preferences using machine learning is a promising area of research that has the potential to help students make informed career choices and lead fulfilling lives.

In addition, the system can provide insights into the current job market and industry trends, enabling students to make informed decisions about their career path. This might assist them in staying current with industry advancements and make strategic decisions about their education and career.

Ultimately, the future scope is vast and holds great potential for both students and career professionals. By incorporating advanced technology and machine learning algorithms, these systems can provide more personalized and accurate career guidance to students, helping them make informed decisions about their future.

References

[1] TY - JOUR AU - Hulbatte, Shivraj PY - 2019/05/31 SP - 3556 EP - 3563 T1 - Enhanced Job Recommendation System VL - 7 DO - 10.22214/ijraset.2019.5583 JO - International Journal for Research in Applied Science and Engineering Technology ER, 2019.

[2] TY - JOUR AU - Paparrizos, Ioannis AU - Basagiannis, Stylianos AU - Petridou, Sophia PY - 2012/01/05 SP - T1 - Quantitative Analysis for Authentication of Low-cost RFID Tags ER, 2012.

[3]   TY - BOOK AU - Valverde-Rebaza, Jorge AU - Puma, Ricardo AU - Bustios, Paul AU - Silva, Nathalia PY - 2018/03/26 SP - T1 - Job Recommendation based on Job Seeker Skills: An Empirical Study ER, 2018.

[4]   José-García, A., Sneyd, A., Melro, A. *et al.* C3-IoC: A Career Guidance System for Assessing   Student Skills using Machine Learning and Network Visualisation. *Int J Artif Intell Educ* (2022). https://doi.org/10.1007/s40593-022-00317-y, 2022

[5]   Kumar N, Gupta M, Sharma D, Ofori I. Technical Job Recommendation System Using APIs and Web   Crawling. Comput Intell Neurosci. 2022 Jun 21;2022:7797548. doi: 10.1155/2022/7797548. PMID: 35774438; PMCID: PMC9239795.

[6]   TY - JOUR AU - Alsaif, Suleiman PY - 2023/01/23 SP - 1 EP - 10 T1 - Machine Learning-Based Ransomware Classification of Bitcoin Transactions VL - 2023 DO - 10.1155/2023/6274260 JO - Applied Computational Intelligence and Soft Computing ER

[7]   Qing Wan, Lin Ye, "Career Recommendation for College Students Based on Deep Learning and Machine Learning", Scientific Programming, vol. 2022, Article ID 3437139, 10 pages, 2022. https://doi.org/10.1155/2022/3437139

[8]   Y. Zhang, C. Yang and Z. Niu, "A Research of Job Recommendation System Based on Collaborative Filtering," 2014 Seventh International Symposium on Computational Intelligence and Design, Hangzhou, China, 2014, pp. 533-538, doi: 10.1109/ISCID.2014.228.

[9]   P. C. Siswipraptini, H. L. H. S. Warnars, A. Ramadhan and W. Budiharto, "Trends and Characteristics of Career Recommendation Systems for Fresh Graduated Students," 2022 10th International Conference on Information and Education Technology (ICIET), Matsue, Japan, 2022, pp. 355-361, doi: 10.1109/ICIET55102.2022.9779037.

[10]   A. H. A. Rashid, M. Mohamad, S. Masrom and A. Selamat, "Student Career Recommendation System Using Content-Based Filtering Method," 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 2022, pp. 60-65, doi: 10.1109/AiDAS56890.2022.9918766.

[11]   K. Appadoo, M. B. Soonnoo and Z. Mungloo-Dilmohamud, "Job Recommendation System, Machine Learning, Regression, Classification, Natural Language Processing," 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 2020, pp. 1-6, doi: 10.1109/CSDE50874.2020.9411584.

[12]   Yang, Shuo, et al. "Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach." *Knowledge-Based Systems* 136 (2017): 37-45.

[13]   T. V. Yadalam, V. M. Gowda, V. S. Kumar, D. Girish and N. M., "Career Recommendation Systems using Content based Filtering," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 660-665, doi: 10.1109/ICCES48766.2020.9137992.

[14]   Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020) IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1- 7281-5371-1 978-1-7281-5371-1/20/$31.00 ©2020 IEEE 665

[15]   M. Qamhieh, H. Sammaneh, and M. N. Demaidi, "PCRS: PersonalizedCareer-Path Recommender System for Engineering Students," IEEE Access, vol. 8, pp. 214039– 214049, 2020.

[16]   L. D. Kumalasari and A. Susanto, "Recommendation System of Information Technology Jobs using Collaborative Filtering Method Based on LinkedIn Skills Endorsement," Sisforma, vol. 6, no. 2, p. 63, 2020.

[17]. A. Ghosh, B. Woolf, S. Zilberstein, and A. Lan, "Skill-based Career Path Modeling and Recommendation," Proc. - 2020 IEEE Int. Conf. Big Data, Big Data 2020, pp. 1156– 1165, 2020.

[18]   H. Jain and M. Kakkar, "Job Recommendation System based on Machine Learning and Data Mining Techniques using RESTful API and Android IDE," 9th Int. Conf. Cloud Comput. Data Sci. Eng., pp. 416–421, 2019.

[19]   E. van Laar, A. J. A. M. van Deursen, J. A. G. M. van Dijk, and J. de Haan, "The relation

between 21st-century skills and digital skills: A systematic literature review," Comput. Human Behav., vol. 72, pp. 577–588, 2017.

[20]   W. Chen, P.Zhou, S.Dong, S.Gong, "Tree-based contextual learning for online job or candidate recommendation with big data support in professional social networks," IEEE Access, vol. 6, pp. 77725–77739, 2018. [5] M. H. Mohamed, M. H. Khafagy, and M. H. Ibrahim, "Recommender Systems Challenges and Solutions Survey," Proc. 2019 Int. Conf. Innov. Trends Comput. Eng. ITCE 2019, no. February, pp. 149–155, 2019.