DESIGN AND IMPLEMENTATION OF CAREER ASSIST: A COMPREHENSIVE CAREER COUNSELLING AND EDUCATION PLANNING WEBSITE

*Note: Sub-titles are not captured in Xplore and should not be used

1st Mohit Singh
Department of Computer Application
Babu Banarasi Das University
Lucknow, India
line 5: mohitsingh281998@gmail.com

2nd Prachi Singh
Department of Computer Application
Babu Banarasi Das University
Lucknow, India
prachisingh2342000@gmail.com

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 4th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 5th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 6th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

Abstract— The Career Prediction Website is a web-based application worked with Django, a high-level Python web framework. The goal of this site is to offer career guidance to professionals and students alike by anticipating their most suitable profession based on their interests and skills. The platform is outfitted with numerous features such as Sign-up/Login system, Resume Analyzer, Knowledge Network, and Courses for the users to upgrade their skills. The central aspect of the website is the Career Prediction module that implements the K-Nearest Neighbors (KNN) algorithm to predict the user's career by analyzing the data the user inputs. The user inputs their skills on different ratings provided in menu, and the KNN algorithm processes the data to recognize the most fitting career. Besides, the site also provides a drop-down menu with various options such as Services, Contact Us, Home, and more. The interface of the website is user-accommodating and simple to navigate, making it accessible to a broad spectrum of users. The Resume Analyzer feature provides the users with suggestions for working on their resumes. Overall, the Career Prediction Website presents a pragmatic and innovative solution for career guidance and improvement, making it a valuable tool for students and professionals alike.

Keywords— *Prediction module, K-Nearest Neighbors (KNN) algorithm, forecast, user's career, data analysis, drop-down menu, user-accommodating, pragmatic.*

1. Introduction

The Career Prediction System developed using the Django framework is designed to address the need for a reliable career guidance platform that uses advanced technology to predict career options.[1] The website has been developed keeping in mind the user's perspective, offering an interactive and user-friendly interface that allows students and professionals to input their academic details and interests with ease. The system has a range of features, including sign-up and login forms, navigation bar, services, courses, and a drop-down menu, making it a comprehensive platform for career guidance.

The implementation of the Career Prediction System using Machine Learning and Django involves the use of the KNN algorithm, which has been trained on previous data analysis to anticipate the most appropriate career path. The algorithm's machine learning approach ensures the predictions are reliable and accurate, reducing the possibility of selecting an unacceptable career path. The results obtained through the system's implementation demonstrate encouraging accuracy and effectiveness, making it a valuable tool for students and professionals alike.

The website's user-friendly interface is an essential aspect of the Career Prediction System, offering users access to relevant information regarding their career path. The platform provides a clear roadmap for achieving objectives, thereby promoting work satisfaction and mental prosperity. The Resume Analyzer feature is another valuable tool that provides users with suggestions for improving their resumes, increasing their chances of securing a job in their desired career path. The Courses and Knowledge Network sections `offer additional resources to help users enhance their skills and knowledge, further increasing their chances of success in their chosen career.

In conclusion, the Career Prediction System using Machine Learning and Django is an innovative solution for career guidance and improvement, providing students and professionals with a reliable tool to make informed decisions about their future career paths. The system's user-friendly interface, along with its machine learning approach, offers an accurate and effective way to predict suitable career paths, promoting work satisfaction and mental prosperity.

2. Problem Statement

In every person's life, a pivotal decision is selecting a career path that will have lasting implications on their future. Opting for a career that is ill-suited to one's interests and skill sets can cause turbulence, stunting one's career progression and leading to feelings of discontent and tension. Difficulties frequently arise when one is lacking in knowledge when it comes to making career decisions, exacerbating hardships within the workspace.

Due to the present education system's lack of emphasis on unique skills and interests, students may opt for careers that do not coincide with their abilities or personal interests. Additionally, they may lack knowledge about certain career paths and the necessary skills to be successful in them. Indecisiveness and confusion can arise when personalized career recommendations are lacking, resulting in job dissatisfaction and burnout due to a mismatch between one's skills and interests and their current profession. A system that offers guidance and awareness could prevent such issues for students and working professionals alike.

As a result, there is a pressing need for a comprehensive, interactive system that employs machine learning to provide accurate and personalized career guidance to individuals. The system should take into account the interests, talents, and educational background of the individual to create a concise plan for achieving their career goals.

This investigation seeks to create a system that utilizes machine learning and the Django framework to create an interactive and user-friendly website that features multiple features including a registration and login form, a navigation bar, services, courses, and a drop-down menu with options including the Knowledge Network, Career Prediction module, and Resume Analyzer. The proposed system is intended to recommend career paths to individuals that are personalized based on their individual needs, this would allow them to make informed decisions regarding their career paths, which may lead to increased work satisfaction and mental health.

3. Related Work

"Student Future Prediction Using Machine Learning" by Dileep Chaudhary et al.[2] provides an overview of existing systems that aim to predict students' suitable career paths. The authors note that most existing systems only use personality traits or interests as the sole factor to predict career paths, which may lead to inconsistent results. However, the paper by Beth Dietz-Uhler and Janet E. Hurn [3] emphasizes the importance of learning analytics in predicting and improving student performance, highlighting the importance of students' interests, abilities, and strengths in their success.

The paper by Lokesh Katore, Bhakti Ratnaparkhi, and Jayant Umale [4] utilizes 12 attributes of students and various classifiers to predict career paths, with c4.5 producing the highest accuracy at 86%.The paper by Roshani Ade and P.R. Deshmukh suggests an incremental ensemble of classifiers that experiment with hypotheses from multiple classifiers and use a majority voting rule to determine the final result, resulting in an accuracy of 90.8%.[5] Additionally, Mustafa Agaoglu's paper emphasizes the importance of different attributes in evaluating faculty performance and compares the accuracy of various classifiers, with c5.0 proving the most accurate due to its higher attribute usage compared to other classifiers like CART, ANN-Q2H, and SVM.[6]

However, the authors note that the recommendations provided by these systems tend to be generalized and not specific to a particular university or country/state. For example, some systems recommend groups of courses like data analyst, accountant, and law that belong to different fields, potentially causing confusion for students. Overall, the related work section highlights the limitations of existing systems and the need for more comprehensive approaches that consider multiple factors beyond just personality traits or interests to accurately predict suitable career paths for students.[2]

4.1 Literature Review

This section provides an overview of the main categories in computing, namely Computer Science, Computer Information Systems, and Information Technology, which offer various career paths in the computing industry. The current prediction techniques are assessed, with particular attention to the models and algorithms employed. However, the analysis of the prediction models based on Naïve Bayes and Random Forest has revealed a research gap in accurately and efficiently addressing career prediction. Therefore, this study aims to investigate the use of the KNN algorithm to develop a more effective and reliable career prediction model.[7]
.

4.2 Career in Computer Science

With various skillsets and focuses, there is a vast array of career possibilities within computer and information technology.Applying mathematical concepts to achieve specific objectives is the cornerstone of careers in computer science. The field covers several areas, including software development, artificial intelligence, cybersecurity, and database management. For success in this branch, it's critical to possess robust math skills, programming knowledge, and problem-solving abilities.[7]

Installation, improvement, and maintenance of computer systems, networks, and databases are abilities included in information technology (IT) professions. Professionals in this field of work will ensure that technology efficiently supports business activities. Different jobs included in this career path are IT consultant, systems analyst, network administrator, and database administrator. Those who practice IT careers must possess technical know-how, issue-solving skills, and mastery of various software and hardware.[7]

To solve business problems, individuals pursuing Computer Information Systems (CIS) careers apply technology. Their responsibilities include developing, designing, and implementing information systems that align with organizational objectives. Some areas of focus within CIS include project management, data analysis, and system administration. In addition to understanding information systems and technology, individuals employed in CIS careers require communication skills and business savvy.[7]

A crucial aspect of selecting a profession in computing is taking into account individual aptitudes, proficiencies, and passions. An investigation of the various career possibilities and the corresponding qualifications and expertise demanded for each job is indispensable.[7]
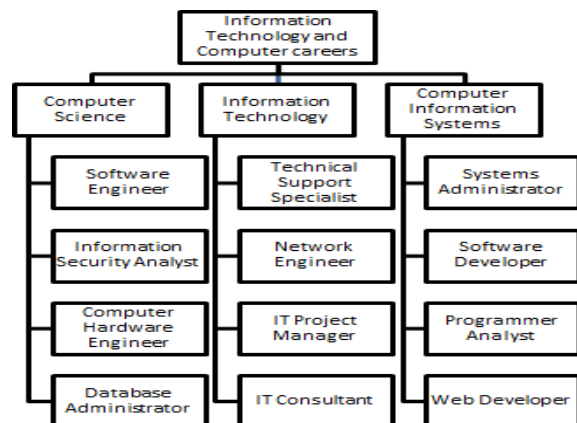

Figure 1 Computing career classification by Wong & Kemp, 2017 [8]

In India, there are numerous career options available in the field of research and technology. Within computing, individuals can pursue various roles or career paths, such as data scientist, UI developer, business process analyst, database administrator, software testing, network analyst, network manager, and more. Each role requires fundamental knowledge in the respective fields.[7]

Due to the rapid evolution of business innovation and technology, there is an increasing demand for qualified IT professionals seeking effective careers in a variety of computer career disciplines. Degrees in computer science (CS), information technology (IT), software engineering (SE), information systems (IS), and management information systems (MIS) are a few examples of computing programs that equip future professionals to enter the workforce and maintain their competence in productivity. Graduates of computing programs possess a wide range of hands-on skills and understanding in various computer and information technology areas. Furthermore, business knowledge sets and non-technical skills are also necessary due to the enormous influence that Information Technology has on commercial processes and setups.[7]

5. Career Prediction

Our career prediction module training code is written in python , which reads in a CSV file 'dataset9000.data' and uses the data to train a K-Nearest Neighbor (KNN) classification model to predict career roles based on a set of 17 skills. First, the code imports the necessary libraries such as pandas, NumPy, pickle, and scikit-learn. The CSV file is read into a Pandas data frame using the read_csv function.

Next, the input features (X) are extracted from the data frame by selecting all the rows and columns 0 to 16, which represent the 17 skills. The output labels (y) are also extracted by selecting all the rows in column 17, which represents the career roles. The code then assigns labels to the column headers for better readability and drops any rows that have all null values. The data is split into training and test sets using the train test split function from scikit-learn. 70% of the data is used for training and 30% for testing, and the random state parameter is set to 524 to ensure reproducibility.

To ensure reproducibility, the Python code loads the dataset into a Pandas data frame using the read_csv function. The code then implements a K-nearest neighbor (KNN) classifier with k=5 neighbors and trains the model using the training data. To evaluate the accuracy of the model, the code predicts career roles for the test data and compares the predicted values with the actual values using the accuracy_score function from the metrics module. The accuracy score is stored in a dictionary with k=5, and the trained model is saved to a pickle file called 'careercounselling.pkl'. However, it should be noted that this code does not include validation or hyperparameter tuning, which are important steps for ensuring that the model is robust and not overfitting the data.Overall, this code is a simple implementation of a KNN classification model for predicting career roles based on a set of 17 skills. However, there is no validation or tuning of hyperparameters to ensure that the model is robust and not overfitting the data.

Then code uses the 'pickle' module to load the pre-trained model saved in the " careercounselling.pkl" file. This model is used to predict the user's career based on their skills. The input data is passed to the model and the predicted career is returned.

The predicted career is then used to index into a dictionary that contains the names of the various careers. The name of the predicted career is displayed to the user.

The code also calculates the probability of the prediction and displays the top careers with a probability greater than 0.05. These careers are displayed in descending order of probability.

Computing Career Prediction

Figure 2 Taking user input

6. Research design

The aim of this study was to develop a predictive model of computing occupations, which was achieved by using the CRISP-DM (Standard Process for Cross-Industry Data Mining) model. Due to its strengths and requirements in the academic research community, this approach has been chosen over other models such as KDD, SEMMA, and KDP. The CRISP-DM model provides a more general and research-oriented explanation of the procedures involved in research. Nai [7] In another study, Abd Alazeez and Thaher [8]

Research designs involve manipulating variables and making predictions based on previous observations or history. The research focuses on identifying and integrating appropriate parameters to create an optimal model that is accurate and reliable in predicting computing occupations.[7]
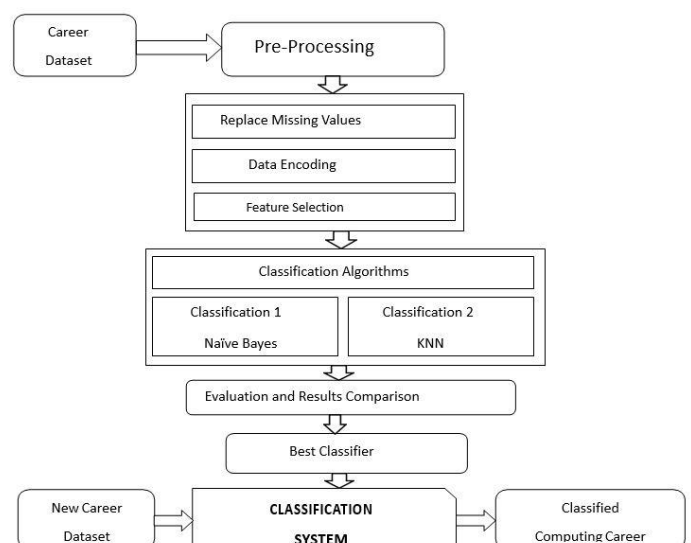

Figure 3 Proposed Computing career classification system overall Architecture

7.1 Resume Analyzer

Corporate entities and recruitment agencies face the daunting task of processing numerous resumes on a daily basis, a task that is nearly impossible for humans to accomplish efficiently. To address this challenge, an automated intelligent system is required to extract vital information from unstructured resumes and transform them into a common structured format that can be ranked for a specific job position. Parsed information includes the candidate's name, email address, social profiles, personal websites, years of work experience, education experiences, publications, certifications, volunteer experiences, keywords, and the cluster of the resume (e.g. computer science, human resources, etc.). The extracted information is then stored in a database (in this case, SQLite) for later use. Although resumes are somewhat structured, information is stored in discrete sets that may vary in type, order, writing style, and format (e.g. .txt, .pdf, .doc, .docx, .odt, .rtf). Parsing such data from different types of resumes effectively and efficiently requires a model that is not reliant on the order or type of data.[10]

7.2 PREPROCESSING

Pre-processing is the primary step in natural language processing that involves transforming raw data into a comprehensible format. Real-world data is often inadequate, conflicting, and contains numerous errors, and pre-processing techniques have been developed to address such issues. Pre-processing involves a series of steps, including:

Data Cleaning: This step involves processes such as filling in missing values, smoothing noisy data, or resolving inconsistencies to clean the data.[10]

Data Integration: Data with different representations are clustered together, and any conflicts between the data are resolved.[10]

Data Transformation: Data is distributed, assembled, and theorized.[10]

Data Reduction: The aim of this step is to present a condensed model in a data warehouse.[10]

Data Discretization: This step involves reducing the number of values of a continuous characteristic by dividing the range of intervals of characteristics.[10]

Tokenization: Tokenization involves dividing a given character sequence or a detailed document unit into smaller units called tokens by removing certain characters such as punctuation. It can be illustrated as follows:[10] in figure 4.

```
In [1]: word="where there is a will, there is a way"
        tokens=[i for i in word.split()]
        print(tokens)

['where', 'there', 'is', 'a', 'will,', 'there', 'is', 'a', 'way']
```

Figure 4 Tokenization

Stemming: Stemming is a crucial step in natural language processing, which involves reducing words to their root or base form. This is done to facilitate analysis and improve search results. The process of stemming involves removing prefixes and suffixes from words, so that variations of the same word are mapped to the same root word. However, it is important to note that the stem may

not always correspond to the actual root of the word. Various algorithms have been developed to perform stemming, dating back to the 1960s. The goal of stemming is to improve the efficiency and accuracy of search engines by expanding the number of results that match a given query. This is accomplished by treating words with the same stem as synonyms, a process known as conflation.[10]

Part-of-speech Tagging: Part-of-speech tagging (POS tagging or POST) is a process in corpus linguistics where a text (corpus) is marked up based on its part of speech, which depends on its definition and context. Tagging is done by identifying the adjoining words in a sentence, phrase, or paragraph. It is also called grammatical tagging or word-category disambiguation. Children in school are often taught a simplified version of this process, where they learn to identify whether a word is a noun, verb, adjective, adverb, etc.[10]

Chunking: Chunking, also known as shallow parsing, is the process of recognizing parts of speech and short phrases. Part-of-speech tagging can determine if the words are nouns, verbs, adjectives, etc., but it cannot provide information about the structure of the sentence or phrase. Sometimes, more information than just the part of speech is needed, but a full parse tree is not necessary. Named-entity recognition is an example of when chunking may be preferred. In NER, the goal is to find named entities, which are usually noun phrases, such as whether "The angry bear" is present in the following sentence: "The angry bear chased the frightened little squirrel." However, it is not necessary to determine if the angry bear is the subject of the sentence. Chunking is also commonly used as a pre-processing step in tasks such as example-based machine natural language understanding, speech generation, and others.[10]

7.3 Implementation

1. Resume Information Extraction: The first key feature of your resume analyzer is the ability to extract important information from the uploaded resume. This includes the name and contact details of the applicant, as well as any skills or technologies mentioned in the resume. This feature is crucial in order to automate the process of screening and shortlisting candidates.

2. Job Field Recommendation: Once the information has been extracted from the resume, the next step is to provide a set of recommendations regarding which job fields the applicant should apply for based on their skills and experience. This feature is particularly useful for candidates who may not be sure about which field they are best suited for or who are looking to explore new opportunities.

3. Course Recommendation: In addition to job field recommendations, your resume analyzer also provides a method to recommend courses that the applicant should enroll in. These courses may be specific to the job field they are interested in or maybe more general courses to help them build their skills and knowledge. This feature is particularly useful for candidates who are looking to upskill or reskill in order to be more competitive in the job market.

4. Key Skills Suggestions: Finally, your resume analyzer also suggests some key skills that the candidate may want to develop in order to enhance their employability. These skills may be related to the job field they are interested in or maybe more general skills that are in high demand in the current job market.

Overall, your resume analyzer provides a comprehensive solution to help job seekers identify their strengths and weaknesses, explore new job opportunities, and develop their skills and knowledge to enhance their employability.

8. Course Module

Our course module is a collection of various computer science courses available on the internet to learn skills. Some of the key features of your course module may include:

1. Course Selection: Our course module will offer a wide range of courses covering different aspects of computer science. This will help students to choose courses that are relevant to their interests and career goals.

2. Skill Mapping: Our course module will provide a detailed description of each course, along with a skill map. This will help students to understand what skills they can learn by taking a particular course.

3. Personalized Learning: Our course module will use machine learning algorithms to personalize the learning experience for each student. This will help students to learn at their own pace and focus on areas that they need to improve.

4. Interactive Learning: Our course module will include interactive learning tools, such as quizzes, assignments, and projects. This will help students to apply the skills they learn in a practical way.

5. Certification: Our course module will offer resources where you get certification upon completion of each course. This will help students to showcase their skills to potential employers.

9. Knowledge Network Module

Our knowledge network module is a collection of various job roles in the IT field with skills required and the role. Some of the key features of your knowledge network module may include:

1. Role Description: Our knowledge network module will provide a detailed description of each job role, along with the skills required. This will help students to understand what skills they need to develop to excel in a particular job.

2. Skill Mapping: Our knowledge network module will provide a skill map for each job role. This will help students to identify the skills they need to develop to qualify for a particular job.

3. Career Guidance: Our knowledge network module will provide career guidance to students, helping them to understand the different career paths available in the IT industry.

4. Industry Insights: Our knowledge network module will provide insights into the IT industry, such as emerging trends and skills that are in demand. This will help students to make informed decisions about their career.

10. Conclusion and Future Scope

In conclusion, the Career Prediction Website is an innovative and practical solution for individuals seeking accurate career guidance and improvement. The platform's user-friendly interface and various features, such as the Sign-up/Login system, Resume Analyzer, Knowledge Network, and Courses, make it accessible to a broad range of users. The Career Prediction module, which utilizes the K-Nearest Neighbors (KNN) algorithm, is the central feature of the website and provides accurate career recommendations based on the user's skills and interests. However, the potential for career counselling apps on the web is limitless, and there are several areas for future development. Expanding the scope of career guidance web applications to cover all fields can enable students from different backgrounds to receive personalized career advice. The addition of machine learning and NLP techniques can significantly enhance the accuracy of the Resume Analyzer, and the incorporation of a recommendation engine can suggest jobs that match a student's skills, interests, and education level. Furthermore, integrating psychometric tests can help identify the student's personality traits, strengths, weaknesses, and interests, guiding them to choose the right career path according to their personal characteristics. Overall, the Career Prediction Website is a valuable tool for professionals and students alike, providing accurate career guidance and recommendations to enable users to make informed decisions about their future. By incorporating new technologies and expanding the platform's features, individuals can take control of their career and work towards achieving their professional goals. The future scope of career counselling apps on the web is vast, and the incorporation of new features such as expanding the scope of fields, enhancing the accuracy of the Resume Analyzer, incorporating a recommendation engine, and integrating psychometric tests can provide more accurate and personalized career advice to students from different backgrounds.

References

[1] Django.[Online].Available: https://www.djangoproject.com/. [Accessed: May. 8, 2023].

[2] D. Chaudhary, H. Prajapati, R. Rathod, P. Patel, and R. K. Gurjwar, "Student future prediction using machine learning," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 5, no. 2, pp. 1104-1108, Mar.-Apr. 2019, doi: 10.32628/CSEIT1952300.

[3] U.D. Beth and H.E. Janet, "Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective," Journal of Interactive Online Learning, vol. 12, pp. 17-26, 2013.

[4] K.S. Lokesh, R.S. Bhakti, et al., "Novel Professional career prediction and recommendation method for individual through analytics on personal traits using C4.5 algorithm," in IEEE Communication Technology (GCCT), Dec. 2015.

[5] A. Roshani and P.R. Deshmukh, "An incremental ensemble of classifiers as a technique for prediction of student's career choice," in IEEE Networks & Soft Computing (ICNSC), Sep. 2015.

[6] A. Mustafer, "Predicting Instructor performance using data mining technique in higher education," IEEE, vol. 4, pp. 2379-2387, 2016.

[7] S. Nai, "Career Prediction Model for Computing College Students in Kenya," Scholarly Project, University of Nairobi Digital,Repository,2022.Available: http://erepository.uonbi.ac.ke/handle/11295/161620.

[8] Abd Alazeez, Y., & Thaher, A. (2021). Data Mining Between Classical and Modern Applications: A Review. AL-Rafidain Journal of Computer Sciences and Mathematics, 171-191.

[9] Wong, Billy & Kemp, Peter. (2018). Technical boys and creative girls: the career aspirations of digitally skilled youths. Cambridge Journal of Education.

[10] S. Sanyal, S. Hazra, S. Adhikary and N. Ghosh, "Resume parser with natural language processing," International Journal of Engineering Science, vol. 4484, 2017.