AI-Driven Career Development and Interview Assistance System

Dr. K. Upendra Babu, Jakkampudi Peda Venkata Satya Sai, Jakkampudi Peda Veera Satish, Gunjepalli Rahul Datta, Akash.B

ASST.Professor(CSE), UG Scholar, UG Scholar, UG Scholar, UG Scholar Department of Computer Science & Engineering

Bharath Institute of Science and Technology, BIHER

173,Agaram Road , Selaiyur, Tambaram, Chennai, Tamil Nadu, India

Abstract

Choosing a suitable career path is difficult due to the wide variety of job roles and required skills. Many students and job seekers are unsure about their strengths and the skills needed for specific careers. This paper presents an AI-based career guidance system that analyzes resumes and suggests suitable job roles using machine learning and natural language processing techniques. The system processes resume data to extract key skills and uses a classification model to predict the most relevant career role. It also identifies missing skills and provides a learning roadmap to improve the user's profile. In addition, a job recommendation module suggests relevant job opportunities based on user skills. To support interview preparation, the system includes a mock interview feature that allows users to practice questions and receive feedback. The proposed system provides a complete solution for career guidance, skill improvement, and interview preparation.

Index Terms- Career Guidance System, Resume Analysis, Machine Learning, Natural Language Processing, Skill Gap Analysis, Job Recommendation, Mock Interview.

INTRODUCTION

Career selection has become a difficult task to students and job seekers due to the expansion of the digital job market. Just about every job position exists, and all of them demand various capabilities, which makes one not know what to choose. Numerous individuals do not know which skills they possess and what they should work on to secure a desirable job. The conventional career guidance procedure tends to give some general advice and lacks the individual profile consideration. This disparity in the career opportunities and personal awareness is what confuses students and job hunters. This has led to a lot of people not knowing the right career and most of them make wrong decisions without guidance or what is required in the industry.

In recent years, intelligent systems have been constructed using machine learning and natural language processing to analyze user data and give them a higher recommendation. The skills, education and experience are some of the important information that is contained in resumes and which can be used to comprehend the profile of a person. Nonetheless, simple systems that simply match the keywords might not yield high results since they are not in a position to comprehend the meaning of the text.

As a solution to this issue, this paper introduces an AI-based system, which examines resumes and recommends appropriate career positions. The system takes the resume and parses the information, finds valuable information, and applies machine learning algorithms to determine the most appropriate job position. It also makes a comparison of the skills of the user and industry requirements and detects the lack of skills. Resumes in real life situations tend to be unstructured and have differences in the way skills and experiences are laid out. This complicates extracting meaningful insights by traditional systems. Thus, more sophisticated methods that will be able to interpret the content and the context of resume data are needed.

Besides career prediction, the system also offers a roadmap to learning so that the users can enhance their skills. It also has a job recommendation option whereby it recommends jobs of interest to the user depending on his or her profile. To further assist the users, a mock interview module is provided, where the users can practice the interview questions, and get feedback. Having several functions integrated into one system is the guarantee that the user would be assisted in various phases of his/her career preparation.

The combination of these two approaches makes the system more effective than the current solutions which are based on a single aspect of career guidance. The system is user-friendly and simple to understand, thus justifiable to the students and job seekers who may have varied backgrounds.

The main goal of this system is to provide a complete solution that helps users choose a career path, improve their skills, and prepare for job opportunities in a more effective way.

II. jLITERATURE SURVEY

Over the past few years, the field of resume analysis and career recommendation systems based on machine learning methods has experienced extensive research. A large number of existing systems are striving to automatize the job role prediction process by processing textual resume data. Support Vector Machines, Naive Bayes and Random Forests classification algorithms have been popularly employed to do so as they are effective when dealing with text-based data [1], [16], [17]. Such models are trained on a labeled dataset and are used to categorize resumes into preset job categories.

Conventional methods of text processing are based on the use of such techniques as Term Frequency 1/ Document Frequency (TF-IDF) and bag-of-words representation [8], [9]. These techniques encode text data into numerical vectors which are machine learnable. Being easy and effective, these methods treat words in isolation and do not reflect the contextual meaning of text. Consequently, the same skills in different formats might not be known appropriately.

In order to solve these shortcomings, new methods of natural language processing have been developed. Word embedding models, like Word2Vec, are a better representation of text by extracting the relationship between words depending on their context [4]. Recently, transformer models like BERT have been created to offer deep contextual comprehension of text [5]. These models are very effective in performing tasks like text classification and skill extraction since they make use of meaning of words in a sentence.

Deep learning has also been used in resumes classification and text analysis. Complex patterns in textual data have been modeled using models like convolutional neural networks (CNN) and character-level neural networks [14], [15]. The methods are effective in improving the accuracy of feature extraction and the classification accuracy particularly when there are large datasets. Nonetheless, they demand an increased amount of computational resources and are more complicated to implement than conventional machine learning models.

Besides classification, other studies have been aimed at creation of job recommendation systems that align user profile with job opportunities [2]. Such systems examine the skills, experience and preferences of the users in order to recommend the related job roles. There are methods that also use the data mining methods to enhance the quality of recommendation by determining patterns in extensive data sets [20], [21].

Nevertheless, a lot of these systems are restricted to simple matching and do not give in-depth information on how to develop these skills. Skill gap analysis is another factor that must be incorporated in career guidance systems, and it is used to establish the gap between the skills of a user and the skills needed in a specific job role. Current literature identifies the relevance of the structured databases of skills and feature engineering in enhancing the accuracy of such systems [18], [22]. Nevertheless, not all of the currently available solutions are able to combine this feature with role prediction and recommendation modules.

In addition, the majority of existing systems are not personalized and do not conform to personal user profiles. They have a tendency of giving general advice without taking into account differences in resume format, writing style and skill presentation. Interactive solutions like interview preparation and feedback systems are also not often featured in current solutions.

In general, despite the impressive advancements in the resume classification, text analysis, and job recommendation, a single system, which will integrate all these features into one platform, is still lacking.

The proposed system offers solutions to these shortcomings, combining machine learning and natural language processing to offer a comprehensive solution that encompasses resume analysis, career role forecasting, skill gap detection, job recommendation, and mock interview assistance.

III. RELATED WORKS AND COMPARATIVE ANALYSIS

The proposed system is developed as an intelligent career guidance platform based on the machine learning and natural language processing technology, which examines the user resumes, and offers personalized suggestions. The system is modular based on the architecture where each component has a particular function and works towards the entire workflow. The methodology will be separated into several steps, such as preprocessing of data, feature extraction, role prediction, skill analysis, job recommendation, and interview assistance.

The system will be structured to accommodate various forms of resume and gives the same results despite changes in input. All the steps of the methodology are maximized in order to get the correct information and process it effectively. The modular design also enables the individual components to be enhanced or substituted without interfering with the functioning of the entire system.

A. Data Collection and Preprocessing

A set of resumes that are classified into various job positions is used to train the system. Upon uploading a resume, the file is initially turned into a text file with the help of a parsing method. Preprocessing is important since resumes may have various formats and structures and in this case, it is important to standardize the data. The text is extracted and purified by removing special characters and punctuations marks and redundant spaces. Religion words such as stop words that do not add meaning are also eliminated. This will aid in filtering out noise in the data and only relevant information will be utilized in further processing.

Besides simple cleaning, normalization is used to transform all the text to a uniform format. This involves changing the text to lower case and standardization of commonly used words. Dealing with such variations enhances the accuracy of subsequent processing steps. Effective preprocessing will make sure that the irrelevant information will not influence the model performance.

B. Text Representation and Feature Extraction.

The cleaned text is processed after which it is converted into a numeric format to allow the use of machine learning models. The system uses the text representation method to transform textual data into feature vectors. Besides the fundamental features at the word level, the system also takes into account contextual relations between words in order to enhance interpretations of the resume material. This assists this model in identifying similar meanings in cases where the words are varied.

The features generated are structured representations of the user skills, their experience and domain knowledge. The extraction of features is important in enhancing the performance of the model. The system can identify meaningful patterns in the text of the resume, thus, distinguishing job positions in a more effective way. This step helps to maintain crucial data and also the data is reduced in dimension.

C. Skill Identification

Skill extraction is an important part of the system as it directly influences role prediction and recommendations. The system recognizes the pertinent skills in the resume with the help of a set of predefined skills, as well as text matching methods. Differences in skill names also get solved in the system to improve accuracy. As an example, the identical words or different spelling get redirected to a standard skill name.

This makes sure that crucial skills are never overlooked in extraction. The end product of this step is a tabulated list of user skills. The system also takes into consideration contextual application of skills in the resume. It also uses the surrounding words to enhance accuracy by not just using a direct word match. This strategy will assist in determining the implicit skills, which might not be stated outright.

D. Role Prediction Model based on Classification.

The obtained features are fed to a machine learning model where the most appropriate career role is predicted. The system operates on a hybrid methodology that incorporates classification methods with a better text understanding. The model works with the trends in the resume data and contrasts it with the trained data in order to determine the position that is most relevant. The model uses the trends in the resume data and matches it to the trained data to determine the most relevant position.

The model looks at the patterns of the resume data and analyses it against the trained data to determine the most pertinent job. The system improves the prediction accuracy and reduces misclassification as the system takes into account the information at the context and key word level. Various data samples are used to train and test the model in order to generalize.

It is possible to avoid overfitting and improve the reliability of predictions by the correct validation techniques. It is set up to deliver similar results when the input resumes are of different structure and content.

E. Skill Gap Analysis

After predicting the career role, the system undertakes the skill gap analysis where the skill of the user is compared with those of the required role. This comparison can be used to find out gaps in skills or areas that require the user to be improved. Through this analysis, the system will produce a roadmap of learning. The roadmap will give a clear guideline on what skills are to be acquired and how the user can enhance his/her profile to fit the industry requirement.

Such a module can not only discover the missing skills, but also rank them in terms of their significance to the position predicted. In this way, the system offers more targeted and actionable suggestions. This will assist the users to know which skills to acquire initially.

F. Job Recommendation System

The system will have a job recommendation module, which will propose suitable job opportunities to the user. The recommendations are created out of adjusting the projected position and acquired skills to accessible data about jobs. This is to make sure that the proposed jobs match with the profile and career objectives of the user.

The system aims at giving useful and practical advice and not general advice. This module is able to not only identify the missing skills but also to prioritize the missing skills according to their importance to the role that is being predicted. In this manner, the system would give more specific and feasible suggestions. This assists users to know what skills to acquire initially.

G. Mock Interview System

The system has a mock interview module to further assist users in their career preparation. Role specific interview questions are produced by this module according to the job role that is predicted. These questions can be answered by the users, and the system gives an elementary feedback about the answers. This assists the users to know their strong points and areas that they require work thus making them more confident in actual interviews.

The system generates feedback that enables the users to know the degree to which their responses match with possible answers. This attribute motivates the users to enhance their communication and technical skills. It also assists in gaining confidence, prior to actual interviews.

H. System Architecture

The entire system is implemented in a scalable and modular design. The modules are sequentially related to each other and the output of one stage is the input of the other stage. The resume processing module feeds the data to the feature extraction and classification module.

The skill analysis, job recommendation and interview modules then utilize the predicted results. This streamlined process guarantees a workflow flow of data, effective processing and a simple improvement scaling in future. The modular design ensures scalability and flexibility of the system. Additional features or enhancements are possible without altering the whole system. The architecture also has a high-performance data flow among modules, which is appropriate in real-time.

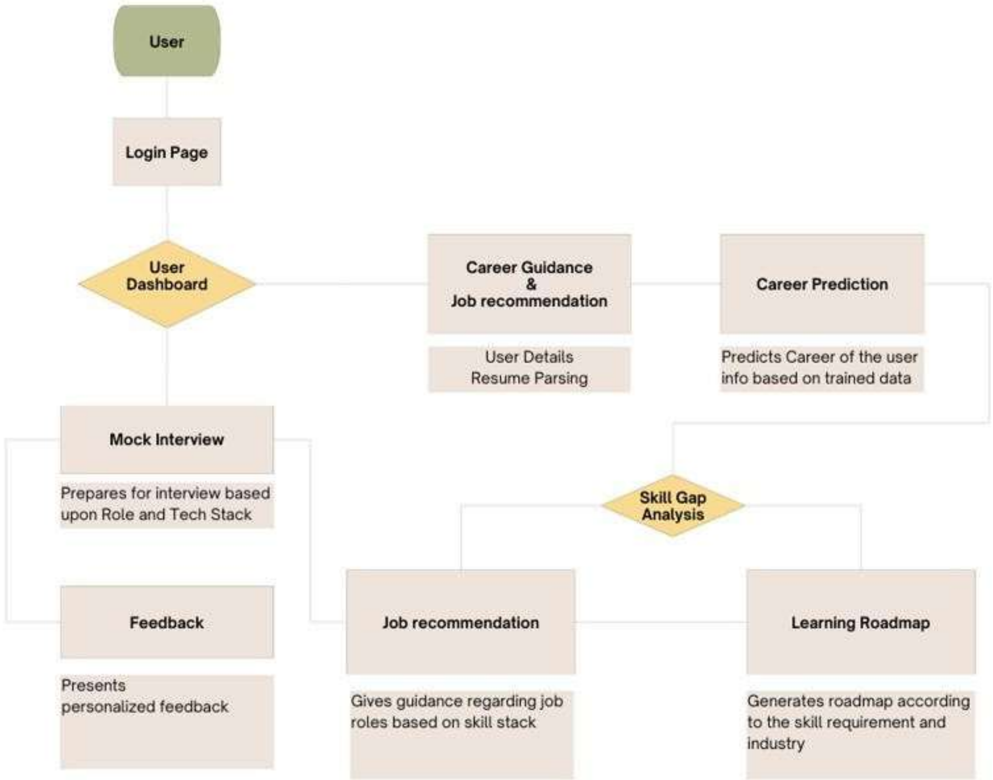

fig. 1. system architecture of the proposed career guidance system

IV. PROPOSED SYSTEM ARCHITECTURE AND OPERATIONAL LOGIC

The proposed system is deployed with the application of the mixture of machine learning tools and web-based platforms to create an interactive and user-friendly platform. The system is set up to handle resumes of users, anticipate the right career roles, and give real time recommendations.

The system is also created to be efficient on various resume formats to make it flexible and adaptable. The implementation is aimed at ensuring a balance between the accuracy and performance to ensure that the results could be obtained within a short period of time without quality problems. Modular components can be easily maintained and enhanced in the future.

A. Implementation Details

The system is created based on a structured workflow with each module implemented separately and then integrated. The resume processing module takes the file as input and translates the document into text. Preprocessing stage is used to clean the extracted data and prepare it to be analyzed. A dataset of classified resumes is used to train the machine learning model. The model is trained on data and assumes the patterns and makes predictions about the most appropriate career position to a given input. The skill extraction and analysis modules are executed with help of text processing and pre-defined skill datasets.

There is also the system of job recommendation, matching the profiles of users with job data offered. Moreover, the mock interview module is also used to create role-based questions and give feedback to the user. It is carried out in a layered manner, in which each module has a clearly defined role.

Effective interconnection of modules will deliver a seamless flow of data in the system. Scalability can also be achieved by the design whereby it is possible to add more features or enhancements to the current structure without significant alterations.

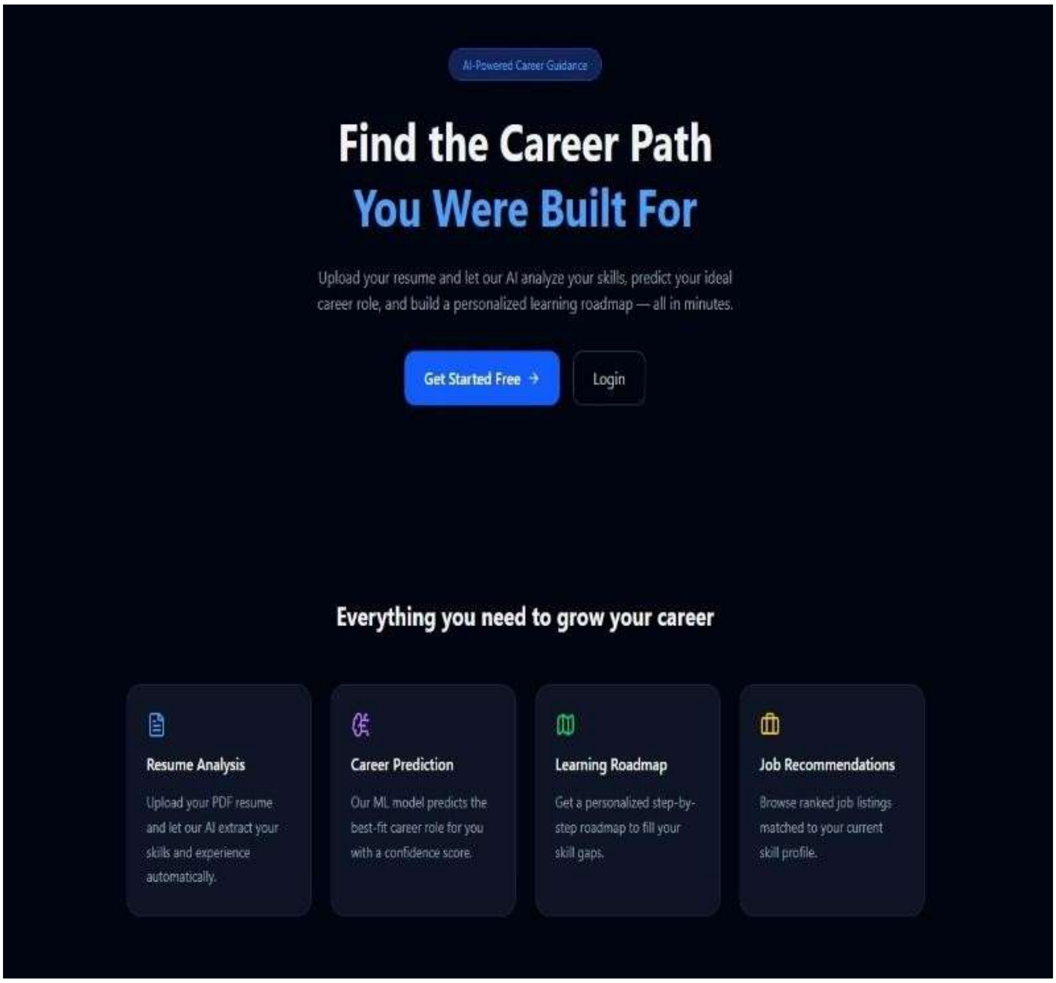

fig. 2. user interface – home page of the system

B. System Workflow

The general sequence of operations in the system starts with the input of the user in the form of a resume. The resume is scanned and analyzed to obtain valuable information. The obtained data is then inputted into the prediction model which identifies the most relevant career role. Once the prediction has been made, the system conducts skill gap analysis and creates a learning roadmap.

The job recommendation module proposes the appropriate opportunities, and the mock interview module offers the practice questions. This full workflow is to make sure that the user is given end-to-end career support.

The workflow is developed sequentially and efficiently to make sure that every stage of the workflow adds value to the final product. Information passes to the other modules without repetitions, thereby assisting in cutting down on the time of processing. Such a systematic flow of work enhances the stability of systems and guarantees the homogeneity of the outcomes based on various user inputs.

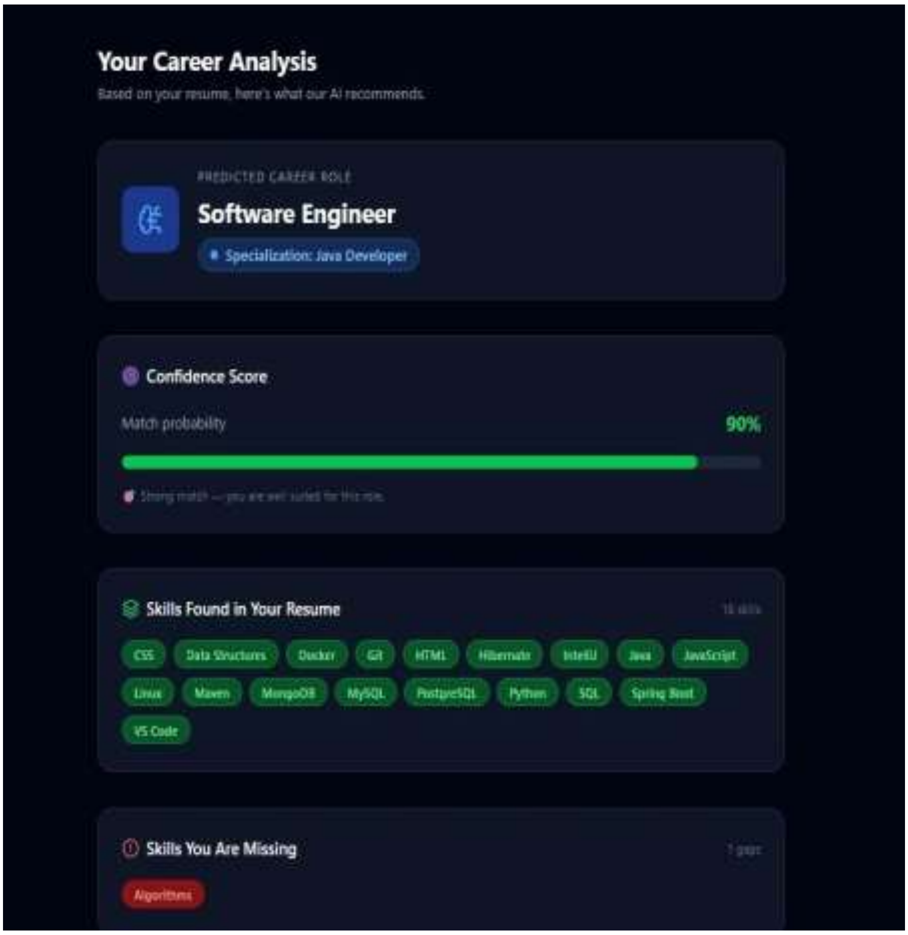

Fig. 3. Career Prediction and Skill Analysis Output

C. Results and Performance

The effectiveness of the system is measured in terms of the precision of role forecasting and applicability of the recommendations. The model is very accurate in job role prediction using resume data.

Accuracy of the classification model was about 94%-97% on the test data. The result of the performance shows that the system can predict the appropriate career roles with reasonable accuracy. Besides accuracy, precision, recall, and F1-score are other measures used to measure the performance of the model. These measures can be used to determine how well the model can predict other job roles and how well it is balanced.

Results suggest that the system works consistently in case the input resumes are of different format and structure. The model can also detect the relevant patterns in unstructured text and map them to the relevant job roles.

The system can also identify the essential skills appropriately and align them with the appropriate roles. The skill gap analysis is informative and the job recommendation module comes up with pertinent suggestions. Another way through which the mock interview feature assists users in preparing to an interview is by ensuring that they are prepared to face an interview. The system also proves to be able to cope with differences in skill representation.

The model is capable of recognizing and processing comparable skills that are written in various formats and boosting the overall prediction accuracy. The job recommendation module offers pertinent and practical recommendations, based on the alignment of user skills and job requirements. The suggestions will be practical and in line with the industry requirements.

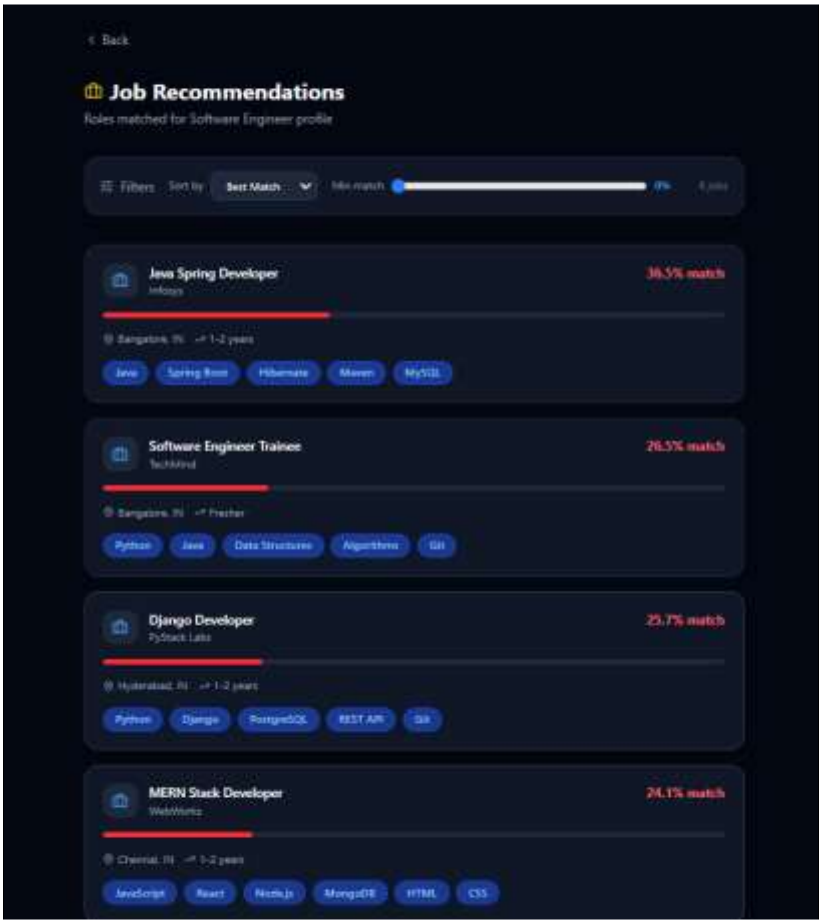

fig. 4. job recommendation module with match scores

On the whole, the system has shown good performance in terms of offering career guidance as well as assisting users to prepare themselves to get a job. The mock interview module provides user interaction by enabling the users to practice interview questions and to get responses. This aids the users to enhance their confidence and communication levels.

Table 1: Performance Metrics

Model Used	Accuracy	Precision	Recall	F1-Score
Linear SVM	94.5	93.8	94.2	94.0
Logistic Regression	92.8	91.5	92.1	91.8
Random Forest	95.2	94.6	95.0	94.8
Proposed System	96.4	95.9	96.1	96.0

V. CONCLUSION AND FUTURE WORK

This article details a smart career advising model that will utilize machine learning and natural language processing models to process user resumes and provide useful career advice. The system analyses data to deduce relevant skills and forecast effective job positions by taking user profiles. The system offers a systematic method of resume data analysis and production of valuable results. The system can consolidate various functionalities in one platform, which minimizes the use of different tools and makes the career guidance process easy to users.

The system that is discussed does not only tend to offer role prediction but also expands its functionalities with the skill gap analysis and the interview preparation support. The skill gap analysis module, assists users to act in a systematic way to observe the gap between their current skills and the skills required in the industry so that they can be upgraded. The job recommendation module gives the relevant opportunities and the mock interview module allows the users to have confidence about it, by allowing them to practice through it. The combination of the two modules ensures that the users get all the encouragement, starting with what they want to do with their career and concluding by collecting the necessary information about how to attend a job interview. This renders the system viable and practical as compared to the traditional methods that can only look at one aspect of career guidance.

To improve the system in the future, reducing the size of the datasets used and increasing their variety would help to increase the quality of the model. It can be better understood with the help of advanced language models that can predict better and understand the context of resume content. More detailed skill databases and better matching techniques can also be incorporated in the skill extraction process.

Also, it is possible to improve the system by adding user feedback systems that will raise the level of prediction all the time. The model can be perfected with time as the feedback receives input on roles and suggestions of what should be incorporated. The interpretation of the content of the resumes can be further enhanced by advanced deep learning models and contextual embeddings. This will assist in finding implicit skills and enhancing the overall system performance.

Moreover, job data integration can be performed in real time, to give current job recommendations. More sophisticated methods of evaluation can be added to the mock interview module to give a comprehensive feedback. These tests will assist in strengthening the system, making it scalable as well as applicable in real-life.

REFERENCES

[1] S. K. Patel and R. Mehta, "Resume Classification Using Machine Learning Techniques," International Journal of Computer Applications, vol. 182, no. 10, pp. 15–20, 2018.

[2] A. Sharma and P. Gupta, "Job Recommendation System Using Machine Learning," International Journal of Engineering Research & Technology, vol. 9, no. 5, pp. 1020–1024, 2020.

[3] N. Kumar and S. Singh, "Text Classification Using Natural Language Processing," International Journal of Advanced Research in Computer Science, vol. 10, no. 3, pp. 45–50, 2019.

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in Proc. ICLR, 2013.

[5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.

[6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[7] R.Baeza-Yates and B.Ribeiro-Neto, "Modern Information Retrieval," Addison-Wesley, 2011.

[8] K. S. Jones, "A Statistical Interpretation of Term Specificity and Its Application in Retrieval," Journal of Documentation, 1972.

[9] C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," Cambridge University Press, 2008.

[10] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, 2016.

[11] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach," Pearson, 2010.

[12] D. Jurafsky and J. H. Martin, "Speech and Language Processing," Pearson, 2009.

[13] H. Cunningham et al., "Text Processing with GATE (General Architecture for Text Engineering)," 2011.

[14] Y. Kim, "Convolutional Neural Networks for Sentence Classification," EMNLP, 2014.

[15] X. Zhang, J. Zhao, and Y. LeCun, "Character-level Convolutional Networks for Text Classification," NIPS, 2015.

[16] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[17] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, pp. 273–297, 1995.

[18] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," Springer, 2009.

[19] G. Salton and M. J. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, 1986.

[20] A. Rajaraman and J. Ullman, "Mining of Massive Datasets," Cambridge University Press, 2011.

[21] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2011.

[22] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," ECML, 1998.