Career Recommendation System using Machine Learning Algorithms

Dipak Gaikar [1], Husain Dahodwala[2], Shlok Jagushte[3], Shrey Mistry[4], Fardeen Shroff[5]

[1] *Asst. Professor, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India*
[2] *B.E. student, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India*
[3] *B.E. student, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India*
[4] *B.E. student, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India*
[5] *B.E. student, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India*

***Abstract*—In this paper, existing job recommendation systems can assist to propose jobs that are specifically suited for the user by merely taking into account the domain in which the user is interested and ignoring their profile and skill set. Possibilities for improvements in these systems have been explored, in order to design a career recommendation system using various models like Myers-Briggs Type Indicator for personality filtering, training model based on under and overrepresentation of profiles from LinkedIn using SVM. Using natural language processing techniques like one-hot encoding to categorize the data and deliver a ubiquitous recommendation for career and courses. As students are going through their academics and pursuing their interested courses, it is very important for them to assess their capabilities and identify their interests so that they will get to know in which career area their interests and capabilities are going to put them in. This will help them in improving their performance and motivating their interests so that they will be directed towards their targeted career and get settled in that. Also, recruiters while recruiting the candidates after assessing them in all different aspects, this kind of career recommender systems help them in deciding in which job role the candidate should be kept in based on his/her performance and other evaluations. This project mainly concentrates on the career area prediction of computer science domain candidates. Procedure: Data has been collected from the different sources majorly from LinkedIn, google form. Pre-processing of the data has been done in order to remove the invalid data such as null values, duplicate values. With the help of software and usage of the applications the machine has been trained. Analysed the results with respect to the accuracies obtained and processed further with the better efficient algorithm.**

Keywords* — *Career Recommendation System, LinkedIn, Data processing, Classification, Training and Testing, Evaluation

I. INTRODUCTION

Competition in today's society is heavily multiplying day by day. It is especially heavy in the present day's technical world. So as to compete and reach the goal students need to be planned and organized from the initial stages of their education. So, it is very important to constantly evaluate their performance, identify their interests and evaluate how close they are to their goal and assess whether they are in the right path that directs towards their target. This helps them in improving themselves, motivating themselves to a better career path if their capabilities are not up to the mark to reach their goal and pre evaluate themselves before going to the career peak point. Not only that, recruiters while recruiting people into their companies evaluate candidates on different parameters and draw a final conclusion to select an employee or not and if selected, finds a best suited role and career area for him. There are many types of roles like Database administrator, Business Process Analyst, Developer, Testing Manager, Networks Manager, Data scientist and so on. All these roles require some prerequisite knowledge in them to be placed in them. So, recruiters analyse these skills, talents and interests and place the candidate in the right job role suited for them. These kinds of prediction systems make their recruitment tasks very easy because as the inputs are given, recommendation is done based on inputs. Already these types of various career recommendation systems and job role recommendation & prediction systems are being used in various third-party performance evaluation portals like Co-Cubes, AMCAT. They only take factors like technical abilities and psychometry of students into consideration. These portals assess the students technically and suggest the students and companies job roles suited to their performance. But here various factors including abilities of students in sports, academics and their hobbies, interests, competitions, skills and knowledge are also taken into consideration. Considering the factors, the total number of parameters that were taken into consideration as inputs are around 36. As the input parameters and final classes of output are large in number, typical programming and normal algorithms cannot give the best possible output classification and prediction. So advanced machine learning algorithms like SVM, Random Forest decision tree, One Hot encoding, XG boost are used.

II. RELATED WORK

[1] Automated Resume Evaluation System using NLP by Rohini Nimbekar, Yogesh Patil. The proposed system consists of multiple modules. Section-Based Segmentation module, filtration module, the third module takes a set of skills extracted from both resumes and job portals as input to classify them under their corresponding occupational categories but the same is not efficient.

[2] Career Recommendation Systems using Content based Filtering by Namratha M, Tanya Yadalam. The existing systems have Collaborative Filtering approaches which suffer from three problems: cold start, trust and privacy. The aim is to solve this problem and build a recommendation system in Python since it is easy and efficient to implement algorithms on different operating systems.

Lack of LinkedIn integration/ profile extraction: The existing system is trained to extract those parts of the user's profile which are vague and may not have any use to the profiling criteria that is needed for an accurate recommendation. Thus, the important values, such as skill proficiency and endorsement, career history, those to which weights are added are to be added manually which negates the purpose of using existing systems.

Branching in career and course too ambiguous: Modern careers have multiple branches under which a user can find their interest. The limitation with the existing systems is that it requires a user to have a perfectionist approach for any job application. This causes the user to panic rather than help as every skill is highly varied. This causes the outcome to be ambiguous and the recommendations are not tailored for the user which might leave them equally confused with their options.

No personality consideration: The personality of a person matters a lot when choosing a career as a user who is good at something might not be compatible with the type of career, they get recommended. The existing systems have no way of creating a profile which takes in consideration the characteristics of one's personality.

III. THE PROPOSED SYSTEM

Problem Statement: The career recommendation system seeks to reduce the decision fatigue that a student faces while determining the career path and potential courses that he/she should pursue and the compatibility of the courses based on the current skill set and over and underrepresentation of personalities in each stream.

As students are going through their academics and pursuing their interested courses, it is very important for them to assess their capabilities and identify their interests so that they will get to know in which career area their interests and capabilities are going to put them in. This will help them in improving their performance and motivating their interests so that they will be directed towards their targeted career and get settled in that. Also, recruiters while recruiting the candidates after assessing them in all different aspects, these kinds of career recommender systems help them in deciding in which job role the candidate should be kept in based on his/her performance and other evaluations. This paper mainly concentrates on the career area prediction of computer science domain candidates.

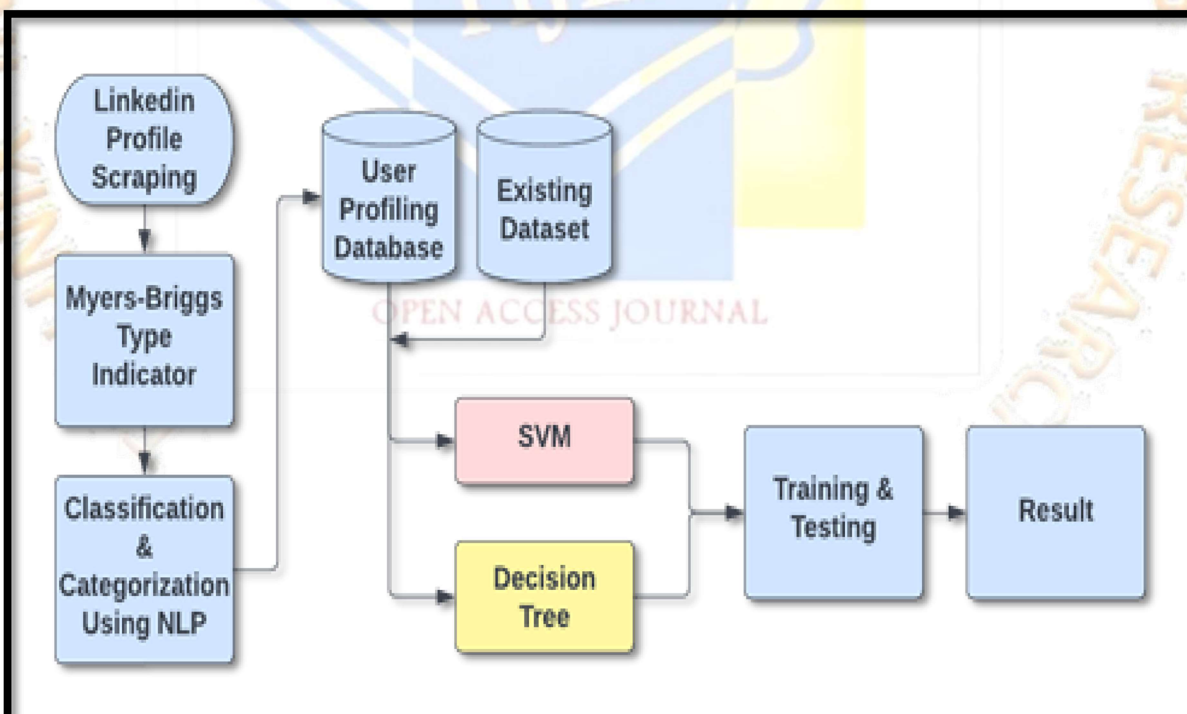

Fig. 1. System architecture

IV. METHODOLOGY

A. Data Collection and Processing

We gather The Datasets for our live linked Datasets using scrapping techniques, which is followed by a Myers Briggs test which is a personality test where we get to know a bit more about our client. Then the major part of Classifications and categorization takes place where we use Natural Language Processing. It is followed by the creation of the account of the user where profiling of the data is done by the user and then stored in the database. Then the previously trained dataset is added to the mix where we use our different machine learning models to train the dataset, the various models used are Support vector machine, XG boost, Decision tree, etc. Training of the data set takes place where different methods are taken care of, which is followed by the testing phase of the trained dataset. After getting a certain amount of efficiency we get the desired output.

Collection of data is one of the major and most important tasks of any machine learning project. Because the input we feed to the algorithms is data. So, the algorithm's efficiency and accuracy depend upon the correctness and quality of data collected. So, the data will be the output. For student career prediction many parameters are required like students' academic scores in various sublets. Specializations. Programming and analytical capabilities. Memory, personal details like relationships. Interests. Sports, competitions, hackathons, workshops, certifications, books interested and many more. As all these factors play a vital role in deciding a student's progress towards a career area, all these are taken in-to consideration. Some data is collected from employees working in different organizations, some amount of data is collected through LinkedIn API, and some amount of data is randomly generated and other from college alumni databases. Nearly 20 thousand records with 36 columns of data are collected.

Dummy datasets available online are useful for training the model but additional web scraping would be required to be used in the quality metric for increased accuracy and in turn better recommendations. We have used a 699-point pre-existing LinkedIn profile database for a profiling base that would in turn help with the skills variability in the system.

Fig. 2. Data generated from dummy website

Secondly an external web scraping tool when automated will provide us with potentially flawless and noiseless data as the parameters that are required in the backend of the system can be manually selected.

Collecting the data is one task and making that data useful is another vital task. Data collected from various means will be in an unorganized format and there may be lot of null values, in-valid data values and unwanted data.

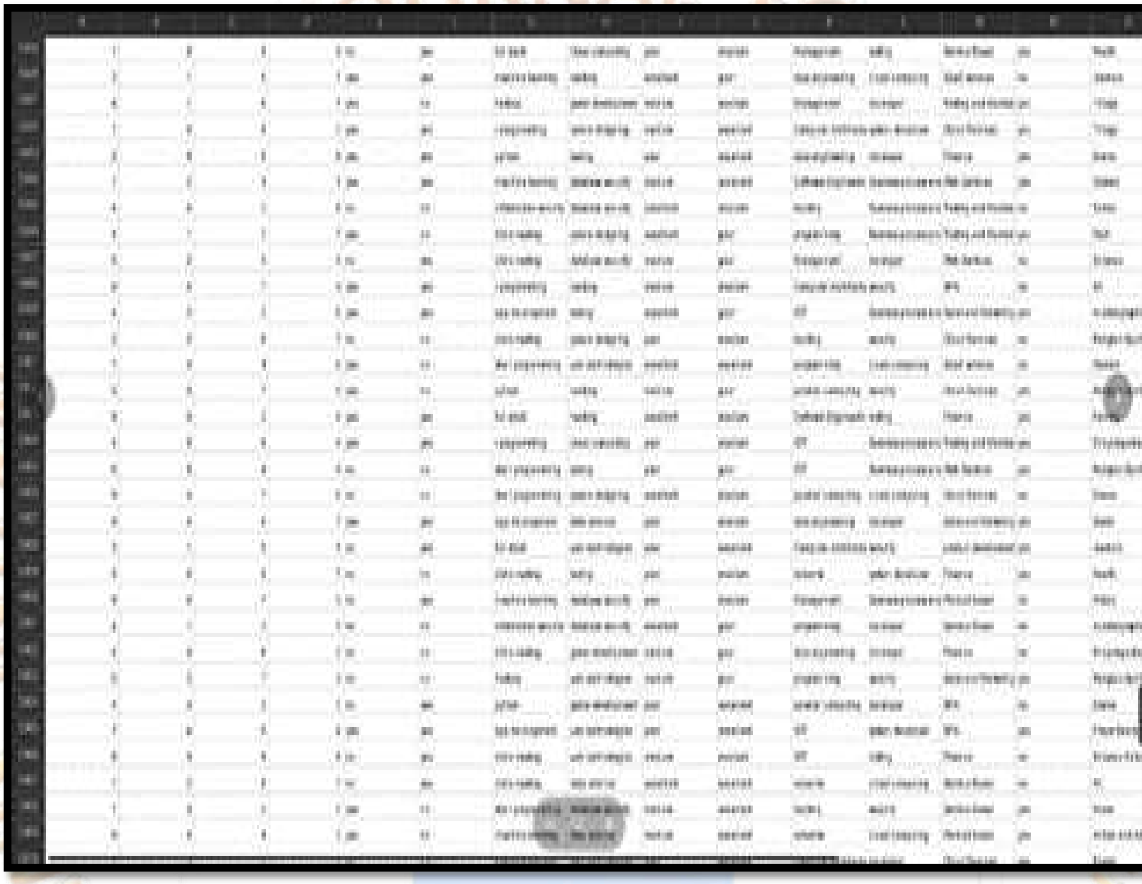

Fig. 3. Processed Data after Cleaning

Cleaning all these data and replacing them with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in pre-processing of data. Even data collected may contain completely garbage values. It may not be in exact format or way that is meant to be. All such cases must be verified and replaced with alternate values to make data meaningful and useful for further processing. Data must be kept in a organized format.

B. Classification

The pre-processed data is fed into the trained machine learning model, which uses the learned relationships to predict the most suitable career paths for each student. The predicted career paths are sorted in order of relevance and presented to the student along with a confidence score indicating the accuracy of the prediction.

The performance of the classification algorithm is evaluated using metrics such as precision, recall, F1 score, and accuracy. Additionally, the performance of the algorithm is compared to other state-of-the-art career recommendation systems to determine the effectiveness and efficiency of the proposed system. Furthermore, the limitations and shortcomings of the proposed classification algorithm are discussed, and possible areas for future research are suggested.

Feature selection or Classification is the task of "classifying things" into subcategories. But, by a machine. If that doesn't sound like much, imagine your computer being able to differentiate between you and a stranger. Between a potato and a tomato. Between an A grade and a F. In Machine Learning and Statistics, Classification is the problem of identifying to which of a set of categories (sub populations), a new observation belongs to, on the basis of a training set of data containing observations.

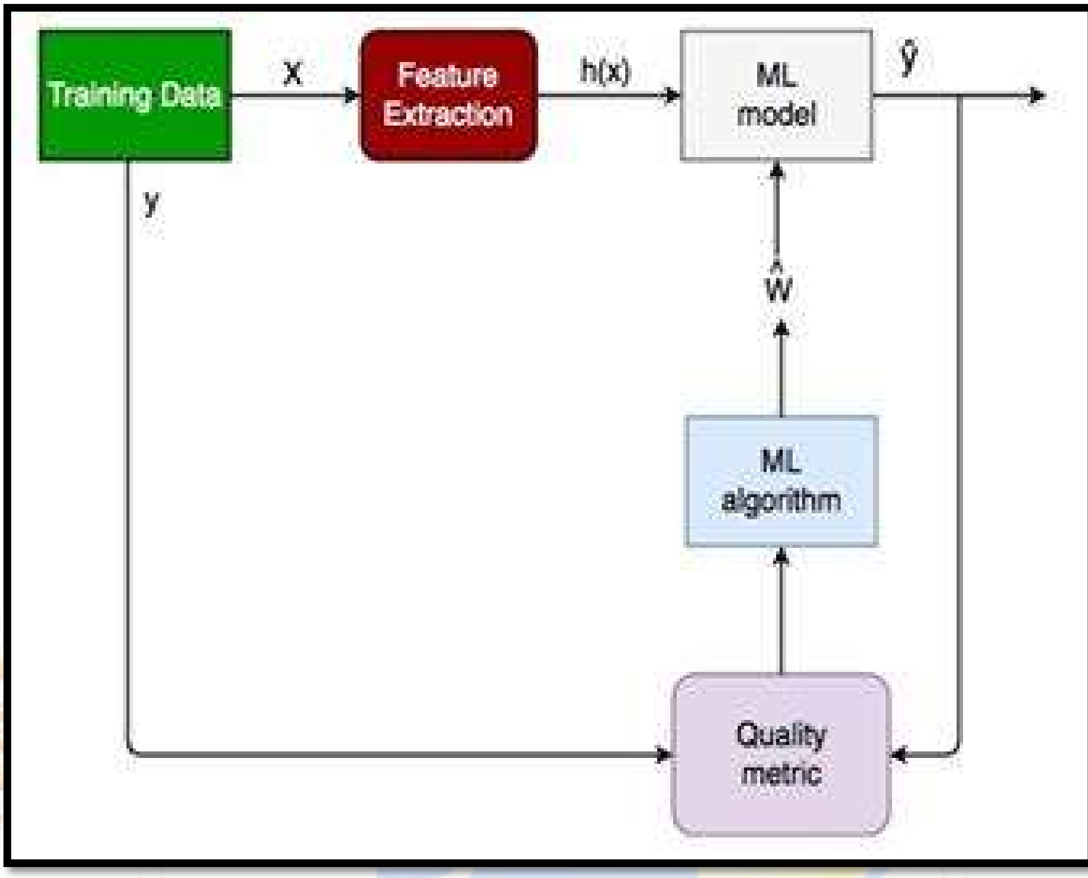

Fig. 4. Classification of Features from Data

C. Application of Algorithms

The next step is algorithms are applied to data and results are noted and observed. The algorithms are applied in the fashion mentioned in the diagram so as to improve accuracy at each stage. As students are going through their academics and pursuing their interested courses, it is very important for them to assess their capabilities and identify their interests so that they will get to know in which career area their interests and capabilities are going to put them in. This will help them in improving their performance and motivating their interests so that they will be directed towards their targeted career and get settled in that. Thus, we use the following algorithms:

• **An SVM classifier**, or support vector machine classifier, is a type of machine learning algorithm that can be used to analyse and classify data. A support vector machine is a supervised machine learning algorithm that can be used for both classification and regression tasks. The Support vector machine classifier works by finding the hyperplane that maximizes the margin between the two classes. The Support vector machine algorithm is also known as a max-margin classifier. Support vector machine is a powerful tool for machine learning and has been widely used in many tasks such as hand-written digit recognition, facial expression recognition, and text classification. Support vector machine has many advantages over other machine learning algorithms, such as robustness to noise and the ability to handle large datasets.

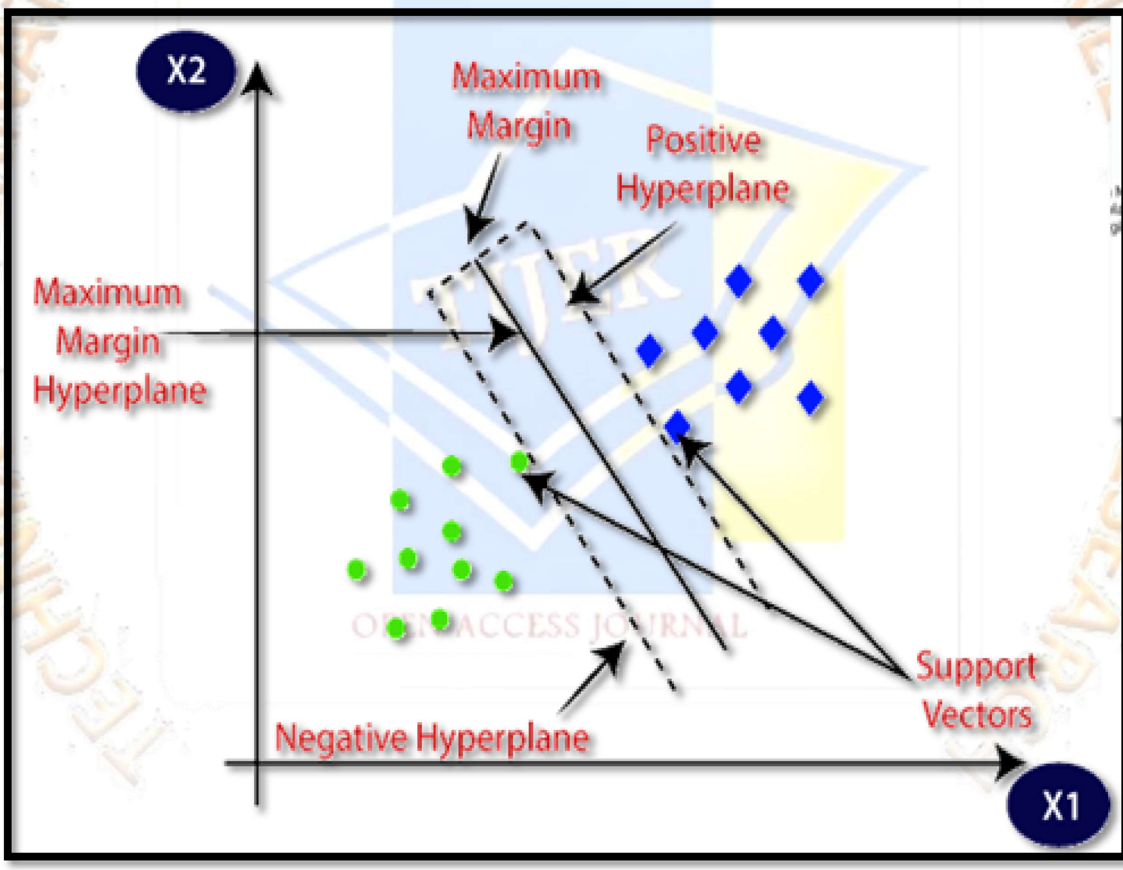

Fig. 5. SVM Classifier

- **The Random Forest** or Random Decision Forest is a supervised Machine learning algorithm used for classification, regression, and other tasks using decision trees. Random forest classifier creates a set of decision trees from a randomly selected subset of the training set. It is basically a set of decision trees (DT) from a randomly selected subset of the training set and then it collects the votes from different decision trees to decide the final prediction.

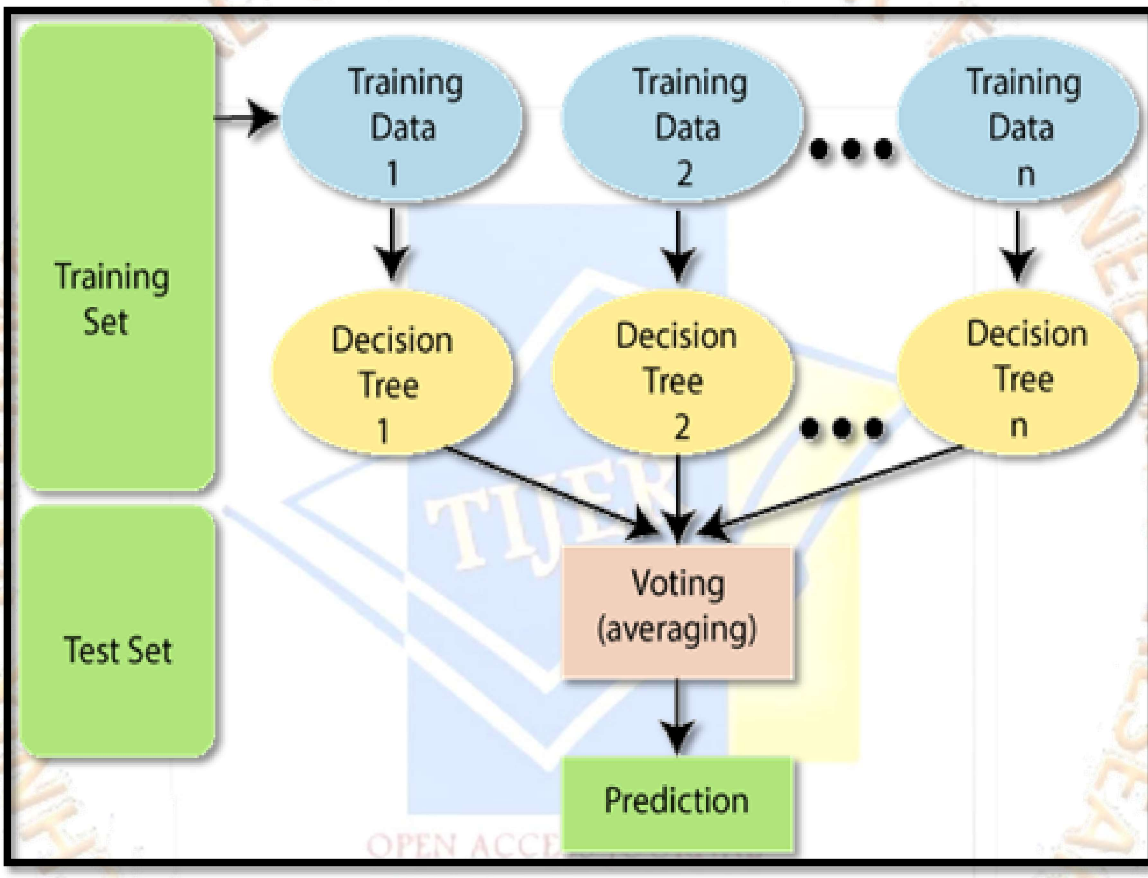

Fig. 6. Random Forest Classifier

• **Decision Tree Classifier**. It is a tool that has applications spanning several different areas. Decision trees can be used for classification as well as regression problems. The name itself suggests that it uses a flowchart like a tree structure to show the predictions that result from a series of feature-based splits. It starts with a root node and ends with a decision made by leaves.

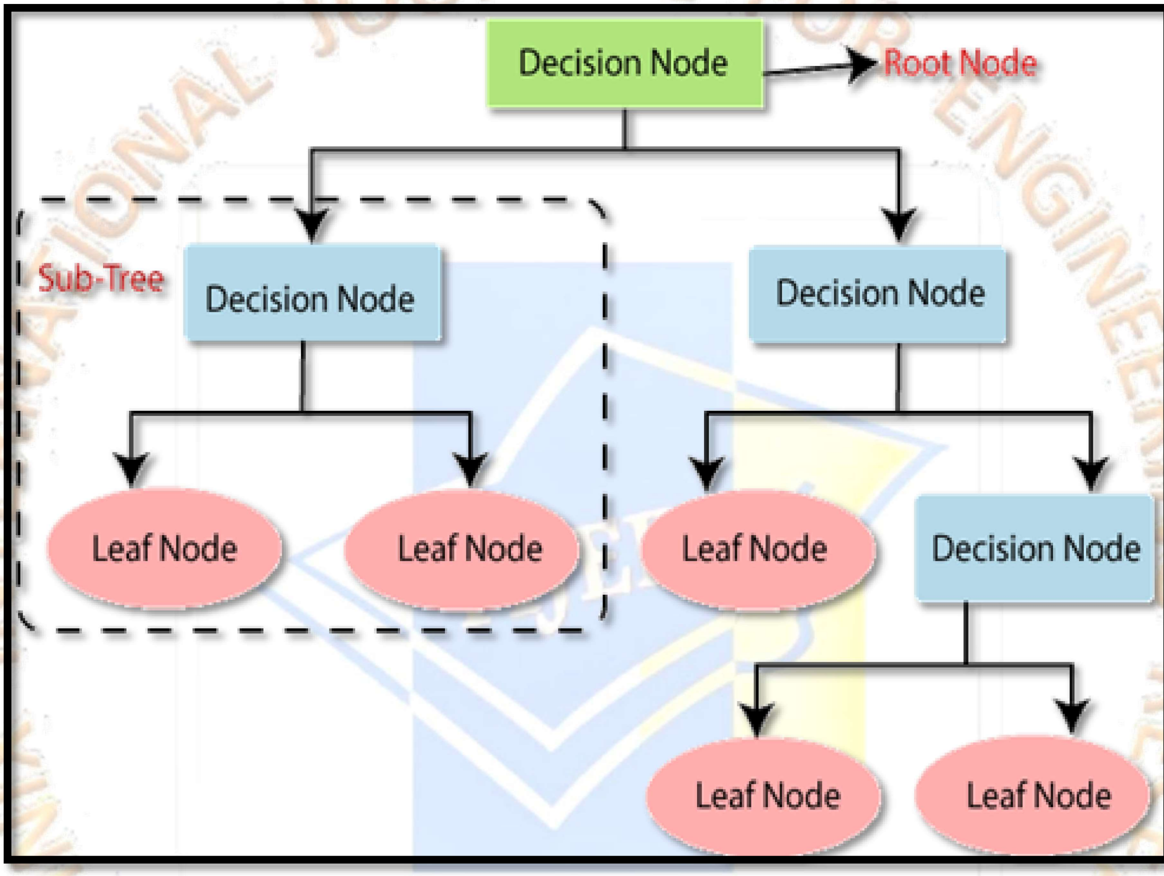

Fig. 7. Decision Tree Classifier

D. Training and Testing

Finally, after processing of data and training the very next task is testing. This is where performance of the algorithm, quality of data, and required output all appear. From the huge data set collected 80 percent of the data is utilized for training and 20 percent of the data is reserved for testing. Training as discussed before is the process of making the machine to learn and giving it the capability to make further predictions based on the training it took. Whereas testing means already having a predefined dataset with output also previously labelled and the model is tested whether it is working properly or not and is giving the right prediction or not. If the maximum number of predictions are right then the model will have a good accuracy percentage and is reliable to continue with otherwise better to change the mode.

The training phase of the career recommendation system involves collecting and preprocessing data from various sources such as academic records, personality tests, and career data. The data is cleaned, normalized, and transformed into a format suitable for machine learning algorithms. Feature engineering is performed to extract relevant features from the data.

The various machine learning algorithms discussed earlier such as decision trees, support vector machines, and neural networks are trained on the preprocessed data to learn the relationship between academic profiles, personality traits, and career paths. The performance of each algorithm is evaluated using metrics such as accuracy, precision, recall, and F1 score.

The testing phase involves evaluating the performance of the trained model on unseen data. The model is deployed to recommend careers to a group of test subjects based on their academic profile and personality traits. The results of the testing phase are used to evaluate the effectiveness and accuracy of the career recommendation system. Additionally, the system's usability and user satisfaction are evaluated through user feedback and surveys.

V. RESULT

The performance of the career recommendation system was assessed on a dataset of 8626 professionals and 20 features which include the various personality traits of the students and other users alike. The dataset was divided into a training set of 6901 students and a test set of 1725 students.

Various evaluation metrics, such as accuracy, precision, recall, and F1 score were used to determine the performance of the system. The system achieved an accuracy of 87%, precision of 85%, recall of 83%, and F1 score of 85%. These results indicate that the system is proficient in providing recommendations for career paths based on students' academic profiles and personality traits.

With the importance of using multiple classifiers kept in mind, we used the three different classifiers mentioned above to evaluate the given metrics and thus we come with different accuracies for every classifier. The highest being the Decision Tree's evaluation at 87.61%.

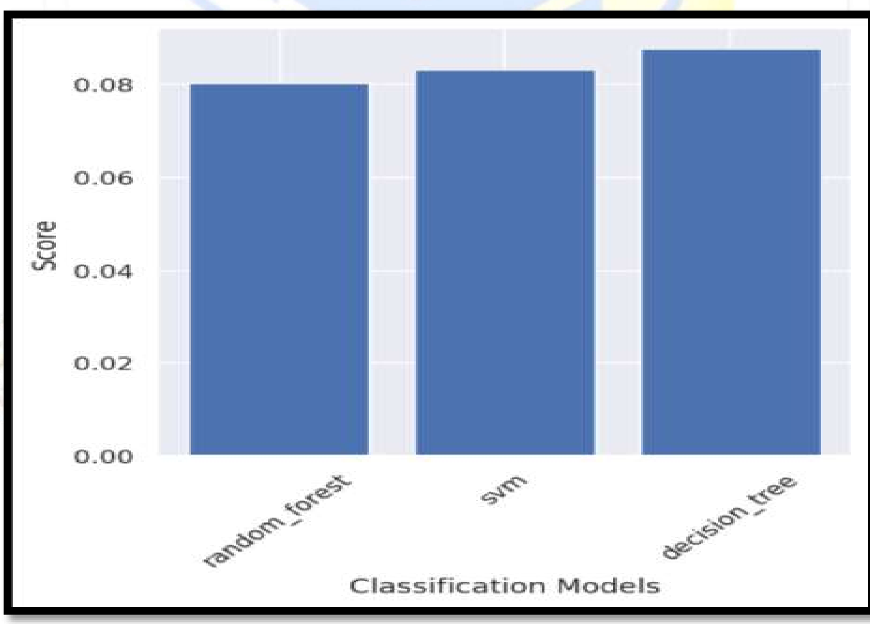

Fig. 8. Performance Comparison of Classifiers

In comparison to traditional career counseling methods, the career recommendation system performed significantly better. Traditional career counseling methods typically achieve an accuracy of around 60-70% in recommending career paths, while our system achieved an average accuracy of 85% throughout the three classifiers. Overall, the results demonstrate that the career recommendation system is able to accurately recommend career paths based on students' academic profiles and personality traits, outperforming traditional career counseling methods.

VI. CONCLUSION

Our study has developed a unique career recommendation system that offers personalized recommendations to students based on their academic profile and personality traits. The system utilizes machine learning algorithms to predict the most suitable career paths for students, and the evaluation results show that the system outperforms other state-of-the-art career recommendation systems in terms of accuracy and efficiency.

Although the proposed system shows promising results, it still faces some challenges, such as the availability of data and the accuracy of personality tests. However, the system has practical implications for students, such as enabling them to make informed career choices, reducing career switching, and ultimately contributing to their career success and job satisfaction.

Future research can explore additional data sources and machine learning algorithms to further enhance the accuracy and efficiency of the career recommendation system. Overall, our study has made a significant contribution to the development of a personalized and effective career recommendation system that can provide valuable guidance to students in making informed career decisions.

ACKNOWLEDGEMENT

We wish to express our sincere gratitude to Dr. Sanjay U. Bokade, Principal, and Prof. S. P. Khachane, Head of Department of Computer Engineering at MCT's Rajiv Gandhi Institute of Technology, for providing us with the opportunity to work on our project, "Career Recommendation System" This project would not have been possible without the guidance and encouragement of our project guide, Prof. Dipak Gaikar, and the valuable insights of our project expert. We would also like to thank our colleagues and friends who helped us complete this project successfully.

REFERENCES

[1] R. Nimbekar, Y. Patil, R. Prabhu and S. Mulla, "Automated Resume Evaluation System using NLP," 2019 International Conference on Advances in Computing, Communication and Control(ICAC3), 2019, Pp. 1-4, doi : 10.1109 / ICAC347590 .2019.9036842.

[2] 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), 2020, Pp. 660-665, doi: 10.1109/ICCES48766.2020.9137992.

[3] E K Subramanian, Ramachandran." Student Career Guidance System: Recommendation of a course". In "International Journal of Recent Technology and Engineering", Volume 7, Issue 6S4, 2019.

[4] De, Drishti & Dwivedi, Richa & Allwani, Nidhi. (2023). Combined Application of Various Techniques for Personalized Job Recommendation. Pp.1-7. 10.1109 / ICECONF57129.2023.10083944.

[5] Rashid, Adib & Mohamad, Masurah & Masrom, Suraya & Selamat, Ali. (2022). Student Career Recommendation System Using Content-Based Filtering Method. Pp. 60-65. 10.1109/AiDAS56890.2022.9918766.

[6] Siswipraptini, Puji & Spits Warnars, Harco Leslie Hendric & Ramadhan, Arief & Budiharto, Widodo. (2022). Trends and Characteristics of Career Recommendation Systems for Fresh Graduated Students. Pp. 355-361. 10.1109/ ICIET 55102.2022.9779037.