

Smart Career Guidance System using Machine Learning

Sanil Gokarn¹, Sanskar Taware², Rujuta Vartak³

¹Department of Artificial Intelligence & Data Science Vidyavardhini's College of Engineering and Technology Mumbai, India
Email: sanil.s211029101[at]vcet.edu.in

²Department of Artificial Intelligence & Data Science Vidyavardhini's College of Engineering and Technology Mumbai, India
Email: sanskar.s211079102[at]vcet.edu.in

³Department of Artificial Intelligence & Data Science Vidyavardhini's College of Engineering and Technology, Mumbai, India
Email: rujuta.vartak[at]vcet.com

Abstract: Numerous post-secondary graduates encounter a substantial quandary when determining their future trajectories. They confront the need for guidance in delineating a fitting vocational journey, amalgamating mental, psychological, and academic facets, along with other pivotal dimensions of their lives. Frequently, they solicit counsel from external sources, precipitating potential confusion or misguidance. To mitigate this predicament, our System was conceived, empowering students to autonomously navigate informed decisions sans third-party influence. Functioning as a pivotal platform, the System enables students to delve into their interests and unearth potential career trajectories congruent with their inclinations. By adeptly collating and scrutinizing pertinent data, the System employs machine learning methodologies to prognosticate students' proficiencies and deficiencies, facilitating the revelation of their domains of mastery.

Keywords: Decision Tree, Support Vector Machine, K Nearest Neighbors

1. Introduction

Career counseling helps individuals reflect on their passions, objectives, competencies, qualifications, and more. It aids in linking educational pathways and professional markets to personal self-awareness. Essentially, career counseling educates us on how to structure and decide on our educational and professional endeavors. It offers support by streamlining, arranging, and making information about job markets and educational opportunities more accessible and convenient for us. In contemporary forms, career counseling integrates insights from diverse academic disciplines such as labor economics, education, sociology, and psychology.

In the contemporary era, selecting a profession isn't merely about earning a living; it's about finding fulfillment and dedicating one's time to something meaningful. Many individuals find themselves in jobs that fail to satisfy their skills and interests, highlighting the need to revamp society's conventional approach to career choices. To facilitate better self-assessment and minimize dilemmas, an online portal has been established for students. Rather than blindly selecting a career path without understanding one's strengths, which often leads to frustration and disillusionment, it's prudent to embrace self-discovery. This initiative aims to empower students to make informed decisions based on their unique abilities and aspirations, free from peer influence or impulsive choices. The system caters specifically to students interested in pursuing careers in information technology and informatics, offering comprehensive assessments covering pertinent topics and subjects in these fields.

2. Literature Review

As described in [14], a career counseling system has been developed to anticipate the most suitable department for an individual by considering their chosen interests and performance in their top two 12th-grade subjects. The system employs K Nearest Neighbor and Support Vector Machine algorithms, attaining classification accuracies of 94% and 89%, respectively.

A web-based application for career counseling has been conceptualized and developed, offering tailored recommendations for candidates to choose the most suitable department. Compared to existing systems, the proposed system provides more precise recommendations. To assess candidate skill sets and determine the best department fit, the system employs the K-Nearest Neighbor algorithm. Additionally, it utilizes the K-Means Clustering algorithm to form clusters based on student scores in specific skill sets, analyzing success rates across various departments within each cluster.

In reference [1], a Career Guidance System specifically designed for students studying Science in the 11th and 12th grades is put forth. This system aims to assist students in selecting suitable career paths post-12th, aligning with their aptitude, emotional quotient, personality traits, and interests. The accuracy rates for the employed algorithms are as follows: The K Nearest Neighbor algorithm yields a classification accuracy of 65%, while Random Forest achieves 81%, and Naive Bayes demonstrates a classification accuracy of 57%.

The suggested system forecasts suitable career paths for students in the 10th and 12th grades by utilizing pertinent input data. Additionally, it offers a quiz feature in

the student's field of interest, allowing them to assess their knowledge. Accessible as both an Android and web application, students can conveniently utilize the system on desktops and mobile devices. Prediction tasks are executed using the Naïve Bayes Algorithm, incorporating a dataset encompassing various attributes such as personal, curriculum, and background details, as well as interests. By providing tailored guidance based on individual interests and skills, this application aims to empower students in making informed career decisions.

The Career Guidance and Recommendation System, as proposed in [2], seeks to anticipate fitting career options for students studying in the 10th and 12th grades by utilizing pertinent input data. Additionally, it offers a quiz feature in the student's field of interest, enabling them to assess their knowledge. Accessible as both an Android and web application, students can conveniently utilize the system on desktops and mobile devices.

A dedicated online platform has been established to aggregate data from applicants, conducting assessments on their aptitude, interests, and personality traits. This dataset serves as the training set for the Machine Learning model, which employs advanced algorithms to generate optimal career path predictions based on the input variables.

As Suggested in Paper [11], refinement of hyperparameters and improvements made to the model, The Student Career Guidance System, incorporating Machine Learning as delineated in [4], deploys the Logistic Regression model to achieve a 74% accuracy rate. While adopting advanced machine learning classification techniques like Decision Trees, Gradient Boosting, and Support Vector Machines may yield more refined models, this endeavor necessitates a substantial expertise level owing to the comprehensive array of features integrated into the model.

1) QUIZ'S based on Skill

As Described in [9] The candidate completes an assessment in this module that consists of both psychological and fundamental skill-oriented questions. HTML 5, CSS 3, and JavaScript are only a few of the web technologies that were used in the development of this module. HTML version 5, often known as Hyper Text Markup Language 5, is a markup language used to create pages that shown within a computer browser. It provides the document with a skeleton structure; if we only use HTML, the document will be static. A style sheet language called Cascading Style Sheets 3 (CSS version 3) is specifically used to add styles and designs to HTML documents in order to improve their display. The client-side scripting language is called JavaScript.

As Suggested in Paper [8] JavaScript is essential to back-end programming, whereas HTML 5 and CSS 3 are essential to front-end development. There will be multiple choice answers for each question displayed separately. JavaScript will be used to help with the validation process, where each option in a question will have a different weight based on which answer is the most appropriate. The final result will be provided according to skill level after

the validation is completed in a skill-wise way.

3. Proposed System

This article introduces a prediction model through a web-based portal, primarily focusing on the CS/IT field. The model evaluates the most straightforward career trajectory for an individual based on a survey filled out by users. The model's primary predictive algorithm, The KNN algorithm leverages advanced technologies like Python, complemented by machine learning methodologies including Decision Trees, Random Forests, and Support Vector Machines. By leveraging LinkedIn profiles, job roles have been mapped according to the requisite skills for specific positions, enhancing the accuracy of our forecasts to align more closely with actual requirements. Upon entering the portal, students are greeted with a welcome screen and landing page providing a concise overview of the site. The subsequent sections detail the two main operational steps of the platform.

Step 1: The first stage involves collecting input from students, who answer questions related to their personal skills and academic grades.

Step 2: Job Role Suggestion

Upon submission of their responses, students receive career recommendations, typically up to a maximum of three top job roles. They are also given the option to explore further details regarding the provided predictions.

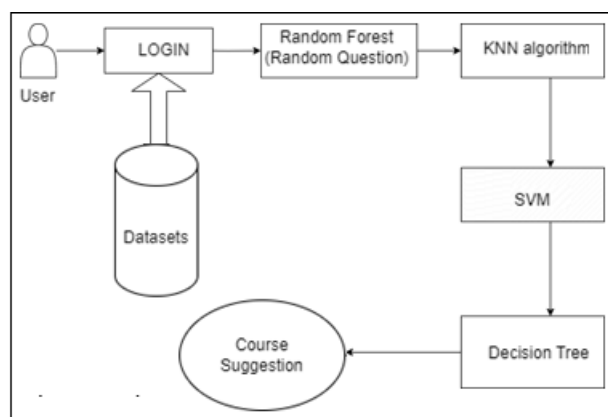


Figure 1: System Block Diagram

4. Methodology Used

Data collection stands as one of the most crucial tasks in any machine learning research, as it serves as the input for algorithms. Consequently, the accuracy and efficiency of these algorithms hinge greatly on the quality and precision of the collected data. In this study, two datasets, denoted as [5] and [6], were merged to form a unified dataset utilized in the

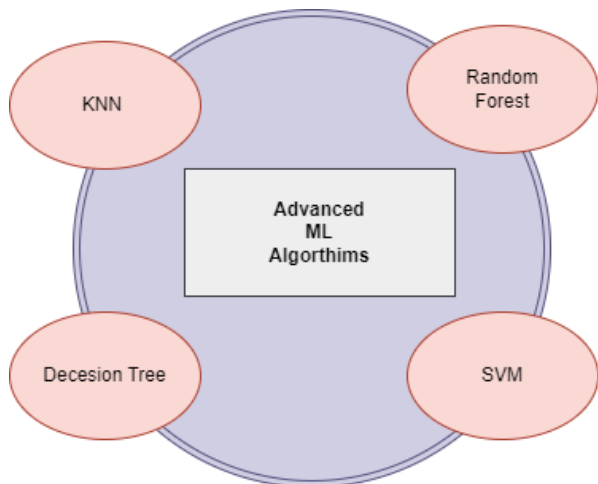


Figure 2: Algorithms Used

prediction model. One dataset encompasses student interests, while the other contains student marks.

Dataset Pre-Processing: The two datasets initially held disparate value types; one featured numeric values within ranges, while the other comprised numeric values represented as percentages. To align them for compatibility, the dataset was normalized, converting the values into categorical data.

Dataset Training and Testing: For our model, the training and testing data are partitioned in a 70:30 ratio to facilitate effective training and evaluation.

Algorithms used for prediction:

As Described in [10] For prediction, the algorithms utilized include the straightforward technique called K-Nearest Neighbor (KNN). KNN is employed to determine the unknown function with adequate accuracy, relying on local minima of the objective function. It calculates the range or distance of the unknown input using metrics such as Euclidean distance, Minkowski distance, cosine similarity measure, chi-square, and correlation properties.

Of all the modules, the prediction module holds paramount importance and serves as the foundation. Various technologies were leveraged in the development of this module, including machine learning techniques, APIs, and datasets for modeling and training. As Described in [13], Python serves as the primary programming language across all applications.

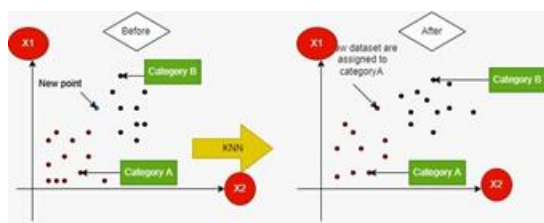


Figure 3: KNN

As Suggested in [7]K-Nearest Neighbor (KNN) excels in handling classification problems, showcasing its proficiency. Despite the utilization of alternative classification algorithms like support vector machine, random forest, and decision tree, K-nearest neighbor

consistently sustains an accuracy exceeding 90%.

This was achieved using a dataset meticulously crafted through a combination of diverse resources and methodologies.

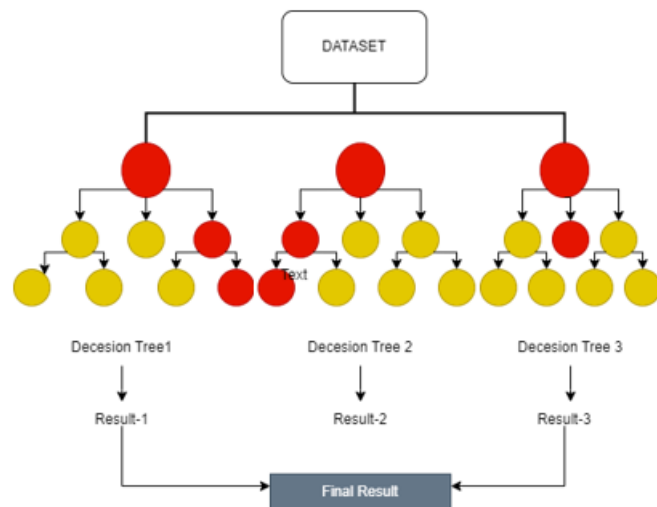


Figure 4: Random Forest

As Described in [12] Random Forest classifier is composed of numerous decision trees created from various subsets of a given dataset. Its predictive accuracy is improved through the aggregation of outcomes across these trees.

Diverging from dependency on an individual decision tree, a random forest amalgamates forecasts generated by numerous trees, employing a majority voting mechanism to ascertain the ultimate outcome. Augmenting the forest's tree count not only enhances predictive accuracy but also serves to alleviate the potential for oversampling.

A decision tree as suggested in [3] serves as a supervised learning technique applicable to both classification and regression tasks, although it's primarily recommended for classification purposes. Structured like a tree, it consists of nodes that represent features of the dataset, branches indicating decision rules, and leaf nodes representing outcomes. The tree includes two key types of nodes: decision nodes, facilitating decisions with multiple branches, and leaf nodes, These trees exclusively consist of terminal nodes devoid of further branches. Decisions or evaluations within the tree are contingent upon the attributes present in the dataset, and it furnishes a visual representation delineating all conceivable solutions or decisions contingent upon specific conditions.

Support Vector Machine (SVM) in [15], stands out as a widely employed supervised learning algorithm applied to address classification and regression tasks, though its primary utility lies in machine learning classification endeavors. SVM focuses on identifying extreme points or vectors to construct the hyperplane. These extreme instances, termed support vectors, delineate the essence of the algorithm, hence its designation as a support vector machine. Consider the

Table I: Accuracy Table

Classification Techniques	Accuracy
KNN	0.9410
Decision Tree	0.8906
SVM	0.8632
Random Forest	0.8105

graph below with two different classes classified using a decision boundary or hyperplane. In the context of a binary classification scenario, the goal of SVM is to find a hyperplane that separates at most two classes. This hyperplane acts as a decision boundary, and the support vectors are the instances closest to this boundary.

We tested out a number before selecting the one that produces the best result using the gridsearch technique. In order to determine the combination of hyperparameter values that yields the greatest performance for a particular model, GridSearch entails creating a grid and doing a methodical search over it. The most common method for doing the search is cross-validation, in which the dataset is partitioned into several subsets, or folds, and the model is trained and assessed several times, using a distinct subset for validation each time.

5. Result

This online portal provided comprehensive recommendations to students using the information gathered from their responses during the assessment process. Predictive output is obtained by mapping these responses to relevant job categories and providing users with personalized career recommendations.

Our primary mission is to help students identify and understand their unique gifts, personalities and talents. With the innovative career guidance system built into our portal, people can make informed decisions about their career. The system becomes a guiding force that not only helps in choosing a career, but also empowers users to direct their choices according to their true interests. The focal point of our project

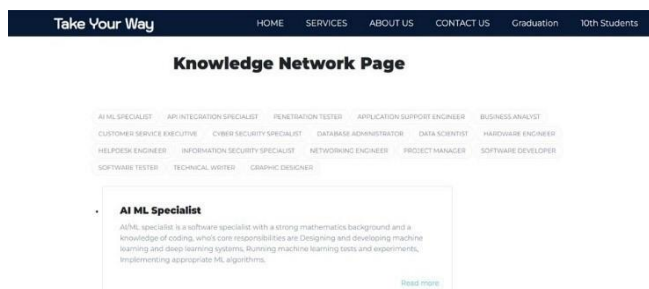


Figure 5: Suggestion Page

lies in providing career recommendations to students. Through our project, we have effectively delivered personalized career guidance to students. The prediction is generated based on the responses provided during the assessment, where the answers are mapped to relevant job categories to derive the outcome.

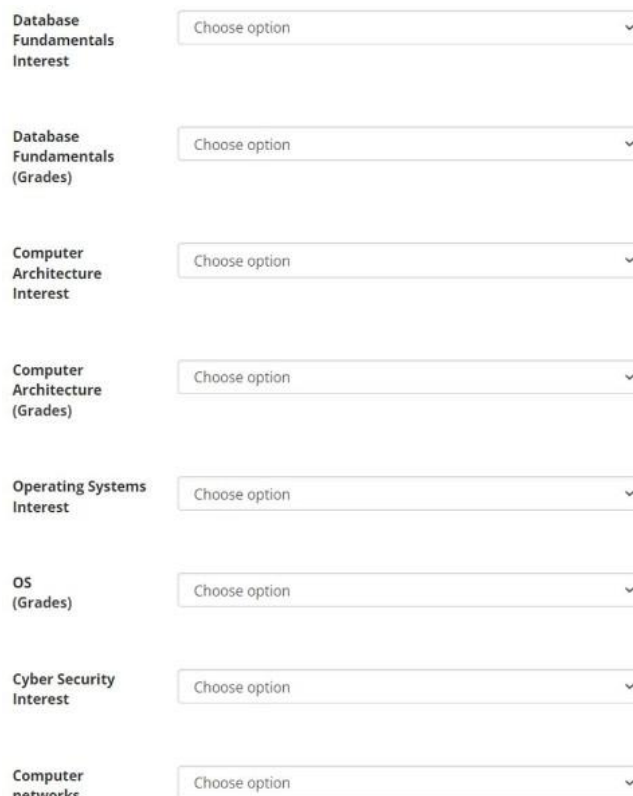


Figure 6: Quiz Assessment

The innovative quiz is meticulously crafted to gauge aptitude and inclination across diverse disciplines such as computer science, mechanical engineering, civil engineering, and more. By leveraging the power of technology, this screening tool aims to provide a comprehensive understanding of an individual’s preferences and strengths, thereby aiding in career guidance and educational counseling. The screenshot showcased in the paper encapsulates the user interface of the quiz, illustrating its user-friendly design and the array of questions formulated to delve into the intricacies of one’s interest in distinct fields. This research marks a significant stride in the realm of personalized career assessment, fostering a data-driven and technology-driven approach to guide individuals towards fulfilling and compatible professional paths.



Figure 7: Prediction result

We’ve been successful in developing an online portal that provides suitable recommendations to aid candidates in choosing the ideal position. We have applied the K-Nearest Neighbor algorithm for categorization. Assisting students in realizing their skills, personalities, and talents is the main objective of the initiative. Students can use a career advising system to help them choose a professional path. It will also allow them to follow their passions. This

approach helps graduates in the CS/IT stream select a career.

6. Conclusion

Incorporating predictive algorithms into job opportunities adds a forward-looking dimension, enabling users to proactively develop their skills and aspirations. The seamless integration of applying directly through our portal not only simplifies the job search process, but also makes the transition from skill acquisition to practical application easier. We've developed a user-centric platform that not only addresses immediate career needs but also cultivates a supportive community and encourages ongoing growth and progress. Alongside providing a chat room for direct communication and offering skill-specific courses tailored to various job requirements, we aim to expand our integration to encompass all relevant departments. Our objective is to incorporate platforms like LinkedIn, Udemy, and others into our services, facilitating connections between users and experts in their respective fields. This integration offers valuable networking opportunities and advice to support users' development and enhance their chances of success in their chosen careers. Our role is to showcase potential positions predicted by our algorithm, enabling candidates to promptly submit applications through our portal.

References

- [1] Dahanke Ajay, Shinde Nilesh, Dhagate Anirudh, Shaikh Huzair, and PV Nagare. An intelligent career guidance system using machine learning. 2022.
- [2] Hmood Al-Dossari, Fawaz Abu Nughaymish, Ziyad Al-Qahtani, Mohammed Alkahlifah, and Asma Alqahtani. A machine learning approach to career path choice for information technology graduates. *Engineering, technology & applied science research*, 10(6):6589–6596, 2020.
- [3] Bahzad Charbuty and Adnan Abdulazeez. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01):20 – 28, Mar. 2021.
- [4] Ankush Daharwal, Sandeep Gore, Aishwarya Bhagwat, Shraddha Detha, and Sunny Chavan. Career guidance system using machine learning for engineering students (cs/it). 2020.
- [5] Samruddhi Deshpande, Pooja Gupta, Nikita Singh, and Deepali Kadam. Prediction of suitable career for students using machine learning. *International Research Journal of Engineering and Technology*, 8(02), 2021.
- [6] Serhii Dolhopolov, Tetyana Honcharenko, Svitlana Anastasiia Dolhopolova, Olena Riabchun, Maksym Delembovskyi, and Oksana Omelianenko. Use of artificial intelligence systems for determining the career guidance of future university student. In *2022 International Conference on Smart Information Systems and Technologies (SIST)*, pages 1–6. IEEE, 2022.
- [7] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In Robert Meersman, Zahir Tari, and Douglas C. Schmidt, editors, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 986–996, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.
- [8] Simon Holm Jensen, Anders Møller, and Peter Thiemann. Type analysis for javascript. In Jens Palsberg and Zhendong Su, editors, *Static Analysis*, pages 238–255, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- [9] Atefeh Khosravirad, Ahmad Akbari, Mohammad Karimi, and Moslem Cherabin. Designing a skill-based education model in schools based on the real needs of society using grounded theory. *Journal of Adolescent and Youth Psychological Studies (JAYPS)*, 5(2):1–13, Feb. 2024.
- [10] Vivekkumar Mourya, Shubham Phatale, Siddhesh Thakur, and Panjab Mane. Career guide application using ml. *International Research Journal of Engineering and Technology (IRJET)*, 7(09), 2020.
- [11] Heungseok Park, Yoonsoo Nam, Ji-Hoon Kim, and Jaegul Choo. Hyper-tendrill: Visual analytics for user-driven hyperparameter optimization of deep neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1407–1416, 2021.
- [12] Aakash Parmar, Rakesh Katariya, and Vatsal Patel. A review on random forest: An ensemble classifier. In Jude Hemanth, Xavier Fernando, Pavel Lafata, and Zubair Baig, editors, *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, pages 758–763, Cham, 2019. Springer International Publishing.
- [13] KR Srinath. Python—the fastest growing programming language. *International Research Journal of Engineering and Technology*, 4(12):354–357, 2017.
- [14] S Vignesh, C Shivani Priyanka, H Shree Manju, and K Mythili. An intelligent career guidance system using machine learning. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 987–990, 2021.
- [15] S.V.M. Vishwanathan and M. Narasimha Murty. Ssvm: a simple svm algorithm. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, volume 3, pages 2393–2398 vol.3, 2002.