# Career Guidance System Using Machine Learning

**Subiddhya Panthee[1], Sudikshya Rajkarnikar[1], Rabiya Begum[1]**

Department of Electronics and Computer Engineering,

Advanced College of Engineering and Management, Nepal

Email: sudikshya.075bct067@acem.edu.np

## ABSTRACT

Career Guidance System is important for students going through academic and pursuing courses to access their capabilities and identify their interests. This paper mainly concentrates on career guidance through personality prediction. It helps them decide which job role suits the best based on their performance and other evaluations. The system will predict the Big Five personality, and VAK learning type of the users through data collected from web-based questionnaires. The system recommends different career choices based on their answers. Machine learning algorithms will be used to check the performance of the system and validate the result.

*Keywords: Career Guidance System, Random forest model, Decision tree, Machine Learning*

## INTRODUCTION

Career Guidance System enables students to choose the career path best suited to their personality, interests, and skills without much effort. New career opportunities are introduced due to changing technological and social trends which have made career guidance more challenging. Therefore, we have proposed to use machine learning algorithms like Random Forest Algorithm and Support Vector Machine Algorithm to predict suitable careers for students based on their skills, interests, and personality so that students can explore and learn about various career paths that best suit them.

Psychology is the branch of science that has integrated the logic and complex emotions of human beings. By studying the various phenomena, psychologists have been able to determine why humans behave in a certain way. It says a lot about human nature and preferences based on the simple questions and answers they provide. The Big Five personality test and Myers Briggs Type Indicator (MBTI) are similar kind of personality assessment that has been used globally for career recommendation. We further wanted to make a more accurate system using personality tests and aptitude tests so that we can predict careers that are also in their intellectual capability so that it can help individuals make more informed decisions about their career paths and increase their chances of success and satisfaction in their chosen field.

We wanted students to be effective learners by helping them figure out their learning styles. For this, we have included VAK learning style cognitive test. Cognitive skills are very important for processing new information. Classifying students on their respective learner like visual, auditory, and kinesthetic type, enables students to reflect on mental and personality dimensions and help them become better learners.

**RELATED WORKS**

The paper, Career Counseling using Data Mining, focuses on finding patterns from datasets and extracting features from the stored or warehoused data which was used to gain insight into aspects of any organization and predict outcomes for future situations. The C5.0 data mining algorithm is used for discovering patterns that categorize data, assemble them into classifiers, and use them to make prediction [1].

The research, An Intelligent Career Guidance System using Machine Learning [2], proposed a system that uses the K-Nearest Neighbor Algorithm to classify the skills of students and the K-Means Clustering Algorithm to create clusters based on skill sets. The k- means clustering algorithm is an unsupervised machine learning algorithm that groups unlabeled datasets in multiple clusters or groups. The departments were grouped using K-Means Clustering Algorithm to map students' performance and provide secondary recommendations. Classification in predictive modeling was also discussed, where class labels are predicted for input data. K-Nearest Neighbor Algorithm was used as a data classification technique. Confusion Matrix was used to determine the performance of the classification model by calculation precision, recall, and accuracy. The confusion Matrix uses tables to represent four different values present in the matrix.

An Intuitive Implementation of Course Recommendation System Based on Learner's Personality Asad Jatoi, Memoona Sami, Muzamil Nawaz and Junaid Baloch [3]. The study focuses on the classification of learners into numerous groups based on learners' capabilities and interests. Naive Bayes Method, Support Vector Mechanism (SVM) and Linear Regression algorithms were used in determining learning style and recommended career choices that will best suit them. Naive Bayes Algorithm was used to determine the degree of excellence in their recognized learning styles. Navie Bayes algorithm enables data filtering using the Bayes Probability Theorem . This method was proposed to be more accurate than other algorithms such as Linear Regression, SVM (Support Vector Machine) because Naive Bayes Method uses concepts of dependency. It uses probabilities of categories and adds newer types into categories that are nearly like the probabilistic values.

The research paper, Use of Artificial Neural Network in Developing a Personality Prediction Model for Career Guidance [4], uses Artificial Neural Network technique to build a predictive model. MBTI (Myers-Briggs Type Indicator) physiological assessment test was used in the manual system to help users choose the best career. Data from the MBTI categorization questionnaires was used in machine learning for training the input traits to develop a better classification system. Confusion Matrix and K-Folds Cross Validation Methods were used to measure the performance of the model.

The study, Career Guidance through Admission Procedure Using Artificial Neural Networks [5], uses Artificial Neural Networks in building a career guiding system for students. The system is based on Multilayer Perceptron Topology for analysis and training of student data. Socio economic, biological, environmental and academic factors are taken into consideration to categorize the input for the model. Those factors were then transformed into a suitable format for neural network analysis.

**METHODOLOGY**

**Machine Learning Algorithms**

**Random Forest Algorithm**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique which can be used for both Classification and Regression problems. It employs numerous decision trees which are trained on different subsets of a given dataset. By averaging the predictions of these trees, random forest improves the accuracy of predictions. It uses prediction from each tree and makes a final output prediction based on the majority vote. Increasing the number of trees in the random forest increases the accuracy of the model and mitigates the issue of overfitting. To decide a number of nodes in the decision tree branch we use Gini parameter which is given as,

$$Gini = 1 - \sum_{i=1}^{c} \left(p_i\right)^2$$

**Support Vector Machine**

Support Vector Machine is a supervised learning platform that is used for classification as well as regression. It aims to create an optimal line or decision boundary, known as a hyperplane, which effectively separates data points in an n-dimensional space into different classes. The algorithm selects critical points known as support vectors, which play a vital role in defining the hyperplane. The kernel in SVM is used to transform input data into higher dimensions feature space which is given as,

$$f(x1, x2) = e^{\frac{-||x1-x2||^2}{2\sigma^2}}$$

Where 'σ' is the variance and hyperparameter and ||x1-x2|| is the Euclidean distance between two points x1 and x2.

**Data Collection and Preparation**

The first step in building and developing a machine learning model is to acquire a relevant dataset. We used two datasets in our project Big Five Personality and the VAK Learning Style dataset. We imported the Big Five Personality dataset from the Kaggle dataset. The dataset contains 1,015,342 questionnaire answers collected online by Open Psychometrics. We imported VAK Learning Style Dataset from Google Dataset Search.

**Data Pre-Processing for Big Five Dataset**

Data preprocessing is a crucial step in machine learning. It enables us to enhance the quality of data to promote the extraction of meaningful insights from the data. We used three core libraries for data preprocessing: NumPy, Pandas, and Matplotlib. We removed unnecessary columns from the data set and calculated the mean of the row that contained the missing values and replaced the result for those missing values.

**Data Pre-Processing for VAK Dataset**

During the data preprocessing of the VAK dataset, we dropped the "Age" column from the dataset, converting categorical variables to numerical values using the LabelEncoder class, and normalizing the data using the Normalizer class from sklearn. The data is then split into training and testing sets using the train_test_split function.

**RESULT AND ANALYSIS**

**Comparison of Accuracy of Algorithms For Big Five Dataset**

Decision Tree and Random Forest machine learning algorithm were used to determine the accuracy of the predicted output. The Random Forest Algorithm was ultimately selected to for the system as it demonstrated higher accuracy than Decision Tree Algorithm.

*Table 1 Comparison table for accuracy for Big Five Dataset*

| ALGORITHM | ACCURACY |
|---|---|
| Decision Tree Algorithm | 78.25% |
| Random Forest Algorithm | 91.14% |

**Comparison Of Accuracy Of Algorithms For Vak Dataset**

For the VAK dataset machine learning algorithms including Decision Tree algorithm, Random Forest Algorithm, XGBOOST and SVM were applied. The accuracy of each algorithm was evaluated, with the SVM machine learning algorithm ultimately providing the most accurate result for VAK Learning Dataset. As a result of the comparison, we decided to apply the SVM algorithm in our final system.

*Table 2 Comparison table for accuracy for VAK Dataset*

| ALGORITHM | ACCURACY |
|---|---|
| Decision Tree Algorithm | 77.68% |
| Random Forest Algorithm | 83.82% |
| XGBOOST | 88.42% |
| SVM | 90.50% |

**Confusion Matrix**

A confusion matrix is an *N x N matrix* used for summarizing the performance of a classification model. It compares the actual target values with those predicted by the machine learning model and evaluates the performance of a classification model through the calculation of performance metrics like accuracy, precision, recall, and F1-score.
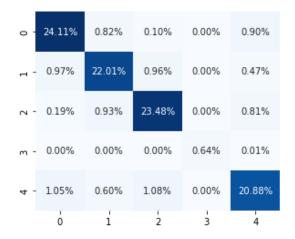
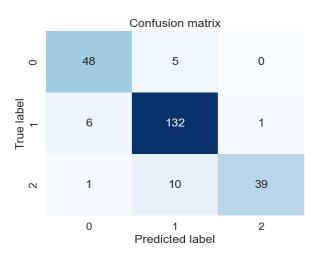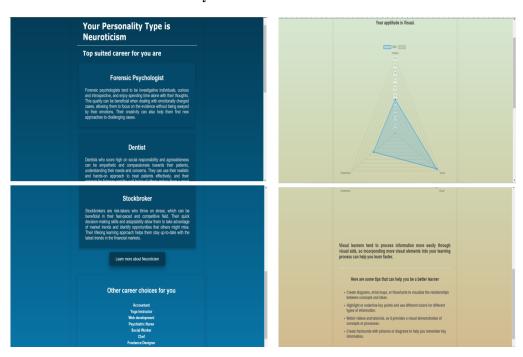*Figure 1: Confusion matrix for Random Forest Model in Big Five Dataset*



*Figure 2: Confusion Matrix for SVM in VAK Dataset*

**Website for Career Guidance System**

## CONCLUSION

Overall, our project aimed to provide a reliable system to recommend career choices based on an individual's personality and learning style. With this system, we hope to help people make informed decisions about their career paths and improve their chances of achieving professional success. Our project focused on predicting the career path of an individual based on their MBTI and Big5 test results. We collected and preprocessed the data using libraries such as NumPy, Pandas, and Matplotlib and employed K-Means clustering and Random Forest algorithms to classify the data into different clusters and improve the predictive accuracy of the model. We used platforms such as Kaggle for datasets and My SQL for databases. The Random Forest model was chosen for the Big5 test and SVM algorithm for VAK learning. Since our model is a multi-class classification model. It is suitable for the classification of individuals into different personality types. It combines multiple decision trees to improve the accuracy of the predictions which is particularly useful for personality prediction as it allows for the identification of complex non-linear relationships between the input variables and the personality types. Also, Random Forest is less prone to overfitting compared to other algorithms such as Decision Trees, making it more robust and reliable for prediction tasks.

## LIMITATIONS

- The model's accuracy heavily depends on the quality and bias of the data used for training. The quality of the dataset of the VAK learning model was very poor.

-  The VAK model is overfitting on the training data of age which makes it difficult to perform well on new data.

- The datasets rely on a limited set of input features so other possible features that could have a significant impact on the accuracy of the model are missed.

- Personality is not a static trait and can change over time due to various factors such as life experiences, environment, and personal growth so the output may not be valid in years to come.

## FUTURE ENHANCEMENTS

- Instead of simply providing users with a personality profile and career options, the model could also offer personalized feedback and recommendations such as books, movies, music, and even job opportunities based on their results.

- Incorporate machine learning explainability techniques, such as LIME or SHAP to understand the model's decision-making process.

- Rather than a simple questionnaire, an interactive personality test could be created that engages the user and provides a more interesting and enjoyable experience.

- The model can be improved by implementing a feedback system that allows users to report any inconsistencies in the predictions. This feedback can be used to retrain the model and improve its accuracy.

**REFERENCES**

[1]     N. Gorad, I. Zalte, A. Nandi, D. Nayak, and A. Professor, "Career Counseling using Data Mining," 2017. [Online]. Available: http://ijesc.org/

[2]     D. Ajay, S. Nilesh, D. Anirudh, S. Huzaif, and P. v Nagare, "AN INTELLIGENT CAREER GUIDANCE SYSTEM USING MACHINE LEARNING," *www.irjmets.com @International Research Journal of Modernization in Engineering*, 2471, [Online]. Available: www.irjmets.com

[3]     A. A. Jatoi, M. Sami, M. Nawaz, J. Baloch, and A. Jatoi, "An Intuitive Implementation of Course Recommendation System Based on Learner's Personality Intuitive Implementation of Alpha Beta pruning using Tic-Toc-Toe View project Recommendation system for personalized learning of students View project An Intuitive Implementation of Course Recommendation System Based on Learner's Personality," 2022. [Online]. Available: https://www.researchgate.net/publication/360121661

[4]     A. Rao, S. Chowdhury, A. S. Rao, B. Sunil Kamath, S. A. Pattan, and R. Krishna Kundar, "Use of Artificial Neural Network in Developing a Personality Prediction Model for Career Guidance: A Boon for Career Counselors," *Article in International Journal of Control and Automation*, vol. 13, no. 4, pp. 391–400, 2020, [Online]. Available: https://www.researchgate.net/ publication/341592047

[5]     A. Kumar Sinha and R. Ranjan, "Career Guidance through Admission Procedure Using Artificial Neural Networks in Different Universities of State Jharkhand." [Online]. Available: www.joics.org