Mapping and Visual Analysis of Vocational career with personality trait based on Big Five and Holland's code using Machine Learning

MINI PROJECT REPORT

Submitted by

SWATHI S	**210701274**
TEJASHREE D	**210701287**

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

ANNA UNIVERSITY :: CHENNAI 600 025

APRIL 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Report titled " **Mapping and Visual Analysis of Vocational career with personality trait based on Big Five and Holland's code using Machine Learning"** is the bonafide work of **Swathi S (210701274) and Tejashree D (210701287)** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Karthick V
Associate Professor,
Department of Computer Science and Engineering,
Rajalakshmi Engineering College,
Chennai – 602015

Submitted to Mini Project Viva-Voce Examination held on ___________________

Internal Examiner **External Examiner**

ABSTRACT

Choosing the right career path is essential for long-term job satisfaction and personal well-being, yet many individuals end up in careers that do not align with their personality traits and interests. This project addresses this issue by developing a machine learning-based system that maps personality traits to suitable vocational careers using the Big Five Personality traits model and Holland's RIASEC vocational codes. By analyzing personality data with the K-Means clustering algorithm, the system identifies natural groupings of traits and maps them to corresponding career types. The process involves preprocessing the dataset, applying MiniBatch K-Means for efficient clustering, and visualizing the results with heatmaps and scatter plots.

This personalized career counseling tool aims to guide young students in making informed career choices, reducing career mismatches, and promoting greater job satisfaction and personal fulfillment. Many working professionals feel burnt out and frustrated because they are in careers that do not suit their interests, often due to external pressures and a lack of self-awareness about their own strengths and preferences. Children exhibit unique traits and interests from a young age, which, if nurtured, can guide them toward fulfilling careers.

This project helps students choose careers according to their specific traits and personalities by using predictive models trained on repositories of personality traits. By integrating K-Means Clustering, the system analyzes personality data to find patterns useful in psychological research, career counseling, and personal development. The resulting insights help children and young adults make better career choices, aligning their professional paths with their inherent traits and interests.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Kathick V** Associate Professor, Department of Computer Science and Engineering. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

Swathi S -210701274
Tejashree D -210701287

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SVM	Support Vector Machines
RIASEC	Realistic, Investigative, Artistic, Social, Enterprising and Conventional.

CHAPTER 1
INTRODUCTION

1.1 GENERAL

Choosing the right career path is crucial for achieving long-term job satisfaction and personal fulfillment. Unfortunately, many individuals find themselves in careers that do not align with their personality traits and interests, leading to burnout and frustration. This misalignment often results from external pressures and a lack of self-awareness regarding one’s strengths and preferences. Recognizing the unique traits and interests that individuals exhibit from a young age can guide them towards more fulfilling careers. This project addresses this issue by using machine learning to map personality traits to suitable vocational careers, helping individuals make more informed career choices.

1.2 OBJECTIVE

The primary objective of this project is to develop a predictive system that aligns individuals' personality traits with appropriate career paths. By integrating the Big Five Personality traits model with Holland's RIASEC vocational codes, the system aims to provide a personalized career counseling tool. This tool will reduce the incidence of career mismatches, enhance job satisfaction, and promote personal well-being by guiding young students in making informed career choices based on their unique traits and interests.

1.3 EXISTING SYSTEM

Current career counseling systems often fail to adequately consider the individual differences in personality traits, leading to generic and less effective career advice. Traditional methods of career counseling typically rely on self-reported questionnaires and lack the precision offered by advanced data analysis techniques. As a result, many individuals end up in careers that do not match their personal attributes and interests. This mismatch can lead to dissatisfaction and decreased productivity, highlighting the need for more tailored and data-driven approaches in career guidance.

1.4 PROPOSED SYSTEM

The proposed system utilizes machine learning, specifically the K-Means clustering algorithm, to analyze personality data and identify natural groupings of traits. The system involves preprocessing the dataset, applying MiniBatch K-Means for efficient clustering, and visualizing the results using heatmaps and scatter plots. By mapping the Big Five Personality traits to Holland's RIASEC codes, the system provides a clear and personalized career guidance tool. This approach not only helps in identifying distinct personality profiles but also aligns these profiles with suitable vocational paths, thereby offering more accurate and individualized career recommendations.

CHAPTER 2
LITERATURE SURVEY

1. The research paper by Gottfredson and Holland (1996) presents a "Dictionary of Holland Occupational Codes," providing a comprehensive reference for understanding vocational interests and career paths. This seminal work serves as a foundational framework for career counseling and guidance professionals.

2. Hearst et al. (1998) introduced the concept of support vector machines (SVM) in their paper published in IEEE Intelligent Systems. This work elucidates the theoretical foundations and practical applications of SVMs, highlighting their efficacy in various machine learning tasks such as classification and regression.

3. Hicks et al. (2021) developed "mbkmeans," a fast clustering algorithm tailored for single-cell data analysis, as presented in their paper published in PLoS Computational Biology. This algorithm has significant implications for bioinformatics and genomics research, enabling efficient clustering of complex biological datasets.

4. Kazameini et al. (2020) proposed a novel approach for personality trait detection using bagged support vector machines (SVM) over BERT word embedding ensembles. Their paper contributes to the advancement of natural language processing techniques for personality analysis, offering insights into the integration of machine learning and linguistic features.

5. Kess (1992) explores the interdisciplinary field of psycholinguistics in his book, "Psycholinguistics: Psychology, linguistics, and the study of natural language." This seminal work provides a comprehensive overview of the theoretical foundations and empirical findings in psycholinguistics, shedding light on the intricate relationship between psychology and language.

6. Mairesse et al. (2007) conducted pioneering research on using linguistic cues for automatic personality recognition from conversation and text, as presented in their paper published in the Journal of Artificial Intelligence Research. This work demonstrates the potential of computational linguistics in personality analysis, paving the way for future developments in this field.

7. Majumder et al. (2017) employed deep learning techniques for document modeling to detect personality traits from text, as described in their paper published in IEEE Intelligent Systems. This research advances the state-of-the-art in personality detection, leveraging neural networks to extract meaningful insights from textual data.

8. Mehta et al. (2020) investigated the predictive power of psycholinguistic and language model features for personality detection, as presented in their paper published in the IEEE International Conference on Data Mining (ICDM). Their work contributes to the development of data-driven approaches for personality analysis, leveraging linguistic cues and machine learning algorithms.

9. Mehta et al. (2019) provided a comprehensive review of recent trends in deep learning-based personality detection in their paper published in the Artificial Intelligence Review. This survey paper offers valuable insights into the state-of-the-art techniques and methodologies in personality analysis, highlighting emerging trends and future directions.

10. Mikolov et al. (2013) introduced word2vec, an efficient method for estimating word representations in vector space, as described in their paper. This groundbreaking work revolutionized natural language processing, enabling advancements in semantic analysis, document clustering, and other text mining tasks.

11. Poria et al. (2013) enhanced SenticNet with affective labels for concept-based opinion mining, as presented in their paper published in IEEE Intelligent Systems. This research enhances the capabilities of sentiment analysis systems, enabling more nuanced understanding of opinions and emotions expressed in textual data.

12. Steinwart and Christmann (2008) provided an authoritative exposition on support vector machines (SVM) in their book. This comprehensive guide offers a detailed overview of SVM theory, algorithms, and applications, serving as a valuable resource for researchers and practitioners in machine learning.

13. Wang et al. (2024) conducted a systematic review and meta-analysis on the use of natural language processing in emergency medicine health service research, as described in their paper published in Academic Emergency Medicine. This research synthesizes existing literature and provides insights into the potential applications of NLP techniques in healthcare.

14. Wang, Fan, and Niu (2023) investigated the effects of career adaptability and educational identity on the career decision-making of Chinese higher vocational students, as presented in their paper published in the International Journal for Educational and Vocational Guidance. This study contributes to our understanding of factors influencing career choices among student populations, offering valuable insights for educational policymakers and career counselors.

15. Wu (2012) contributed to the advancement of K-means clustering with a comprehensive exploration of data mining thinking in his book. This seminal work provides a detailed overview of K-means clustering algorithms, methodologies, and applications, serving as a valuable resource for researchers and practitioners in data mining and machine learning.

CHAPTER 3
SYSTEM DESIGN

3.1 DEVELOPMENT ENVIRONMENT

3.1.1 HARDWARE SPECIFICATIONS

This project uses minimal hardware but in order to run the project efficiently without any lack of user experience, the following specifications are recommended

Table 3.1.1 Hardware Specifications

Operating System	(Windows/Linux)
RAM	4GB minimum
Secondary Storage	(256 GB Minimum)
Web Camera	Required
Audio speakers	Required

3.1.2 SOFTWARE SPECIFICATIONS

The software specifications in order to execute the project has been listed down in the below table. The requirements in terms of the software that needs to be pre- installed and the languages needed to develop the project has been listed out below.

Table 3.1.2 Software Specifications

Operating System	Windows/Linux
Internet Browser	Chrome/Edge/Mozilla Firefox
Python	Version 3.13.3
Softwares used	Visual Studio Code, Jupyter Notebook

3.2 SYSTEM DESIGN

3.2.1 ARCHITECTURE DIAGRAM

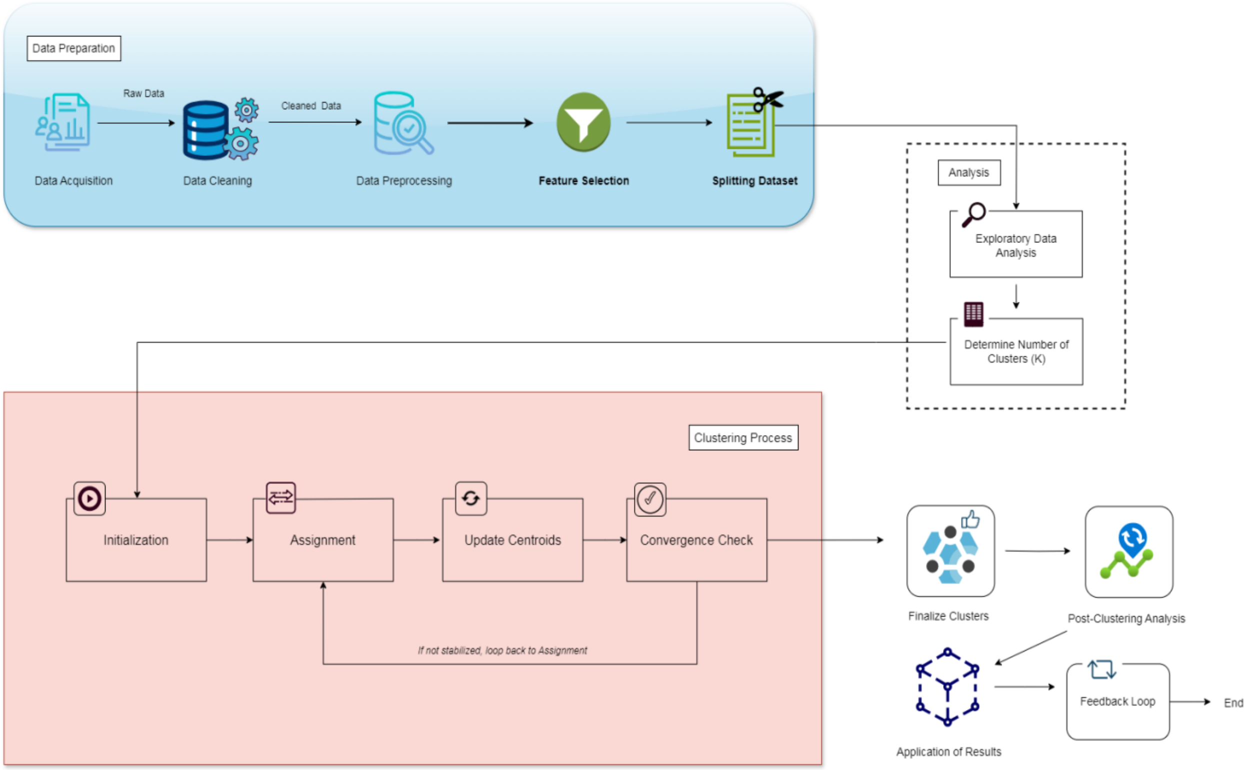

Fig 3.2.1 Architecture Diagram

The system architecture of the vocational career counseling tool is designed to seamlessly integrate data processing, machine learning, and visualization components to provide a robust and intuitive user experience. At the core of the architecture is the data pipeline, which begins with the collection of personality trait data from Kaggle, including the Big Five Personality traits and Holland's RIASEC codes. This data is fed into a preprocessing module where it is cleaned, filtered, and transformed to ensure it is suitable for analysis. The preprocessed data is then passed to the clustering module, where the K-Means algorithm is applied. This module is responsible for identifying natural groupings within the personality data by iteratively assigning data points to clusters and updating centroids until optimal clusters are formed.

The system's evaluation module uses metrics like the Silhouette Score and Inertia to assess the quality of the clustering, ensuring that the resulting clusters are both cohesive and well-separated. Once the clusters are validated, the mapping module links the identified personality traits to corresponding vocational categories using Holland's RIASEC codes. This mapping translates the abstract trait groupings into practical career recommendations.

Finally, the visualization module generates heatmaps and scatter plots to present the clustering and mapping results. Heatmaps display the correlation between personality traits and RIASEC types, while scatter plots visualize the relationship between different clusters and their trait scores. These visual tools help users intuitively understand how their personality profiles align with potential career paths. The entire system is designed to be interactive and user-friendly, facilitating effective career counseling and personal development.

PRE-PROCESSING:

Preprocessing is a crucial step in the development of the vocational career counseling system, ensuring that the data used is clean, consistent, and suitable for the machine learning algorithms. The preprocessing pipeline for this system involves several key steps:

1. Data Collection: The initial step involves collecting datasets from reliable sources such as Kaggle. The primary datasets used are the Big Five Personality traits and Holland's RIASEC vocational codes. These datasets include responses from individuals who have taken online personality and vocational interest tests.

2**. Data Cleaning:** The collected data often contains noise, missing values, and inconsistencies that need to be addressed. This step involves:

- Handling Missing Values: Missing values are either imputed using statistical methods (e.g., mean, median imputation) or removed if they are too numerous and could bias the results.
- Removing Duplicates: Duplicate entries are identified and removed to prevent redundancy.
- Correcting Inconsistencies: Ensuring consistent data formats (e.g., date formats, categorical labels) across the dataset.

3. Feature Extraction: Relevant features are extracted from the raw data. For the Big Five Personality traits, features include scores for extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience. For Holland's RIASEC codes, features include scores for Realistic, Investigative, Artistic, Social, Enterprising, and Conventional interests.

4. Data Transformation: Transforming the data into a suitable format for clustering. This may involve:

- **Normalization:** Scaling the data so that features contribute equally to the distance calculations in the clustering algorithm. Normalization techniques like min-max scaling or z-score standardization are commonly used.
- **Encoding Categorical Variables**: If any categorical data is present, it is converted into numerical form using techniques like one-hot encoding or label encoding.

CHAPTER 4
PROJECT DESCRIPTION

4.1 MODULE DESCRIPTION

1. Data Collection and Preprocessing

The first module focuses on collecting and preprocessing the datasets required for the project. Data is sourced from Kaggle, including the Big Five Personality traits and Holland's RIASEC vocational codes. The preprocessing involves cleaning the data to remove inconsistencies, handling missing values, and extracting relevant features to ensure the datasets are ready for analysis. This step is crucial for maintaining data quality and ensuring accurate clustering results.

2. Clustering with K-Means

The second module involves applying the K-Means clustering algorithm to the preprocessed data. This module starts by selecting an appropriate number of clusters (ten in this case) and initializing random centroids. Each data point is then assigned to the nearest centroid based on Euclidean distance, forming initial clusters. The centroids are iteratively updated, and the assignment of data points is adjusted until the clusters converge. This process results in distinct clusters that represent unique combinations of personality traits.

3. Evaluation Metrics

The third module evaluates the quality of the clustering using metrics such as the Silhouette Score and Inertia. The Silhouette Score measures how similar each data point is to its own cluster compared to other clusters, providing a clear indication of cluster cohesion and separation. Inertia measures the sum of squared distances between data points and their centroids, with lower values indicating better clustering. These metrics help ensure the effectiveness and reliability of the clustering process.

4. Mapping Traits to RIASEC Codes

In the fourth module, the personality traits identified in the clustering process are mapped to Holland's RIASEC codes using a dictionary structure in Python. This mapping links the clusters of personality traits to specific vocational categories, translating the abstract trait groupings into practical career recommendations. This step is essential for aligning personality profiles with vocational interests.

5. Visualization

The final module focuses on visualizing the results of the clustering and mapping processes. Heatmaps are generated to show the compatibility between personality traits and RIASEC types, with color intensity indicating the degree of correlation. Scatter plots are also created to visualize the relationship between clusters and trait scores, where the x-axis represents different clusters and the y-axis represents trait scores. These visualizations provide clear and intuitive insights, making it easier for users to understand how their personality traits align with potential career paths.

By integrating these modules, the system offers a comprehensive tool for career counseling, helping individuals make well-informed career choices that align with their unique personality traits and interests.

CHAPTER 5

IMPLEMENTATION AND RESULTS

5.1 IMPLEMENTATION

For the implementation of the vocational career counseling system, various methods and algorithms were utilized to ensure robustness and efficiency in mapping personality traits to suitable vocational careers. Initially, extensive data was collected from reputable sources, specifically Kaggle, which provided datasets for both the Big Five Personality traits and Holland's RIASEC vocational codes. This data included responses from individuals who had completed personality and vocational interest tests, encompassing various demographic and psychometric attributes.

The collected data underwent meticulous preprocessing, involving several critical steps to prepare it for analysis. Data cleaning techniques were employed to handle missing values through imputation or removal, and duplicate entries were identified and eliminated to maintain data integrity. Categorical variables were encoded using methods like one-hot encoding, and numerical features were normalized and standardized to ensure uniformity and enhance the performance of the clustering algorithm. Additionally, dimensionality reduction techniques, such as Principal Component Analysis (PCA), were applied to reduce the feature space while preserving essential variance.

For the clustering process, the MiniBatch K-Means algorithm was chosen due to its efficiency in handling large datasets. The data was divided into training and validation sets to ensure the model's robustness. The number of clusters (K) was determined to be 10, based on initial exploratory data analysis and domain knowledge. The algorithm was initialized with random centroids, and iterative steps were taken to assign data points to the nearest centroids and update the centroids' positions until convergence was achieved. This process was optimized using metrics like the Silhouette Score and Inertia, which were rigorously calculated to assess the quality of the clustering and ensure well-defined and cohesive clusters.

The next step involved mapping the clustered personality traits to Holland's RIASEC vocational codes. This mapping was conducted using a dictionary structure in Python, which facilitated the translation of abstract personality clusters into practical career recommendations. The mapped data provided insights into how different personality profiles align with specific vocational interests, aiding in the development of personalized career counseling recommendations.

Model evaluation was conducted rigorously, focusing on metrics such as cluster cohesion and separation, as indicated by the Silhouette Score, and the sum of squared distances within clusters, measured by Inertia. These evaluations ensured that the clustering model was both accurate and reliable. Visualization tools, including heatmaps and scatter plots, were generated to intuitively display the compatibility between personality traits and vocational types. Heatmaps showed the degree of correlation between traits and career interests, while scatter plots illustrated the relationship between clusters and trait scores.

Upon finalizing the model, it was deployed into a production environment to enable real-time career counseling. The deployment process involved developing web applications and APIs to integrate the model seamlessly into existing educational and career guidance platforms. Continuous monitoring and maintenance of the deployed model were prioritized to ensure its ongoing accuracy and reliability. Regular updates and retraining sessions were scheduled to adapt the model to new data patterns and evolving career landscapes.

Overall, the implementation of the vocational career counseling system emphasized the use of advanced methods and algorithms to achieve high efficiency and accuracy in predicting suitable vocational careers based on personality traits. This robust implementation ensures that individuals receive personalized career recommendations, enhancing their career satisfaction and personal fulfillment.

5.2 OUTPUT SCREENSHOTS

The accuracy of the clusters is evaluated by 2 metrics namely - ‘Inertia’ and ‘Silhouette. The optimal number of clusters is where the inertia starts to stabilize or decrease at a slower rate. In this case, the inertia decreases from approximately 58 for 2 clusters to around 42 for 10 clusters. The optimal number of clusters could be where the silhouette score is highest. From the table, we can see that it dips at 4 clusters (around 0.084) and peaks around 6 clusters (near 0.092). Based on the results shown below, a reasonable choice might be 6 clusters, where both metrics are relatively favorable.

Number of clusters	Inertia	Silhouette Score
2	~58	~0.095
3	~52	~0.090
4	~48	<0.085
5	>45	>0.085
6	<45	~0.092
7	>45	<0.090
8	<45	>0.087
9	>42	<0.089

Fig 5.2.1 Inertia and silhouette scores for each clusters

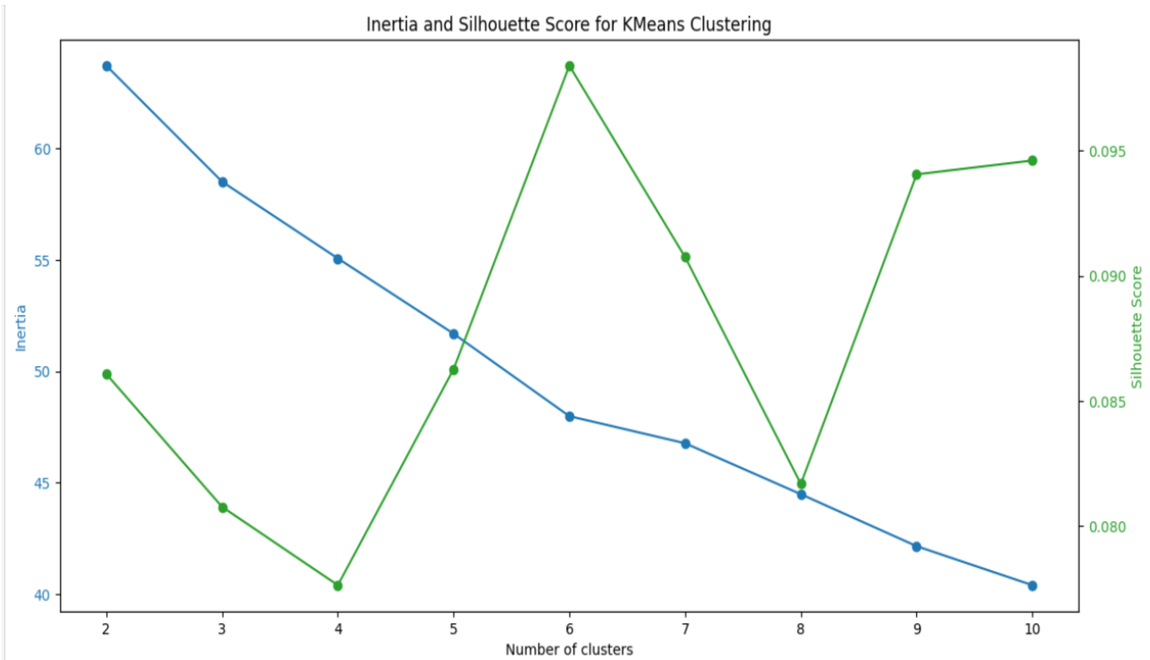

Fig 5.2.2 Inertia and silhouette scores for each clusters

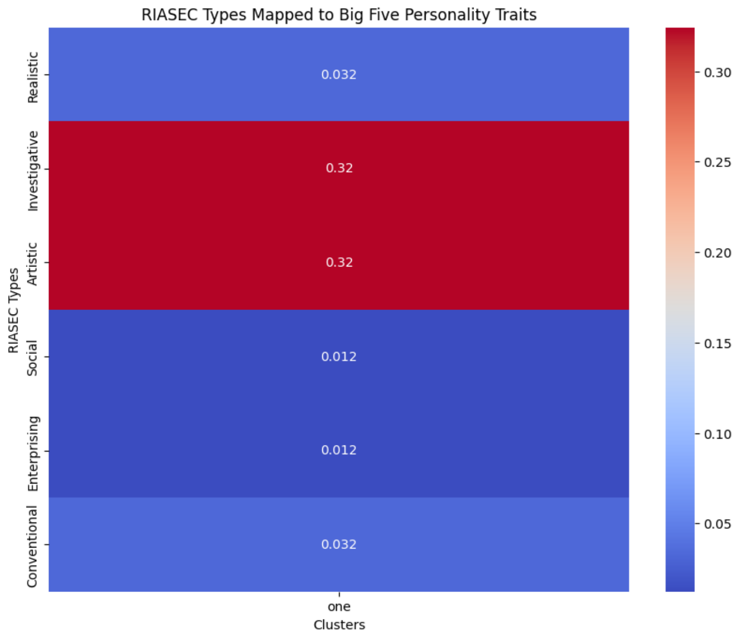

Fig. 5.2.3 RIASEC Types Mapped to Big Five Personality Trait Scores for cluster 1

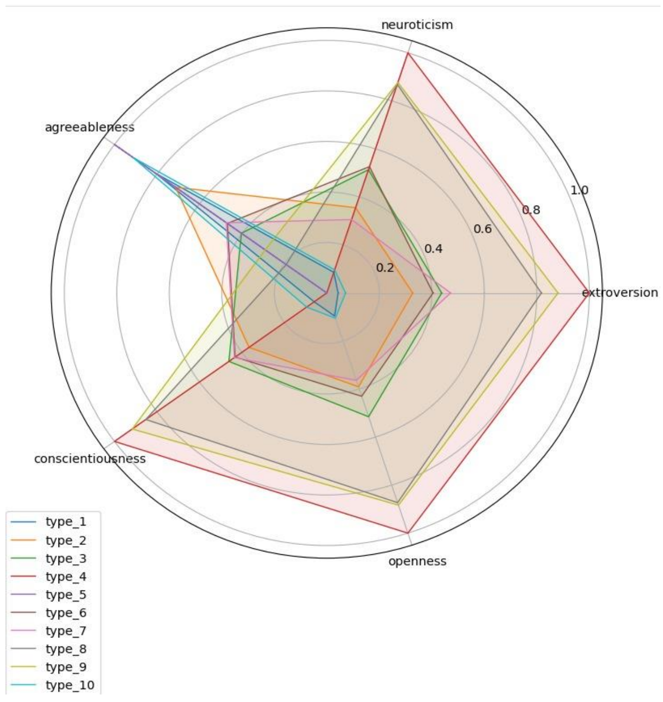

Fig 5.2.4 Radar Chart representing to identify strengths and weaknesses of each trait in each cluster

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENTS

6.1 CONCLUSION

In conclusion, we presented a comprehensive approach to address the challenge of guiding individuals, particularly young students, in making informed career choices aligned with their interests and traits by visualizing their relationship. The project utilized a combination of Holland's RIASEC model and the Big Five Personality traits, using machine learning algorithms to predict vocational career paths.

Through the integration of K-Means Clustering, the system identified natural groupings within personality data. By mapping Big Five Personality traits to Holland's code, the system established relationships between personality traits and vocational categories, offering a structured framework for career guidance. The proposed study offers several contributions. Firstly, it provides a data-driven approach to career prediction, moving beyond traditional methods of career counseling. Secondly, it empowers individuals to make informed decisions by aligning career choices with their unique traits and interests. Thirdly, it enhances the effectiveness of vocational guidance.

Future work could focus on further refining the predictive models through larger and more diverse datasets. Additionally, integrating additional factors such as aptitude tests, personal values, and preferences could enhance the accuracy and relevance of career predictions. In addition, creating a system for parents who can record down their childrens' day-to-day activities which in turn can forecast what coursework would interest their children. Overall, the project represents a significant step towards empowering the next generation to make meaningful and fulfilling career choices.

6.2 FUTURE ENHANCEMENTS

Future enhancements for this project entail several strategic avenues. First, there's a focus on refining predictive models by leveraging expansive datasets to augment accuracy and reliability. Second, the integration of supplementary factors, like aptitude tests and personal preferences, aims to provide a nuanced understanding of career suitability. Third, fostering parental involvement will offer a holistic approach to career exploration, nurturing collaboration and support. Fourth, enhancing the user interface with interactive features aims to elevate engagement and accessibility, ensuring a seamless user experience. Finally, conducting longitudinal studies will enable tracking of career trajectories, facilitating ongoing refinement and alignment with dynamic market demands. These initiatives collectively propel the project towards continuous enhancement and robust adaptability in the ever-evolving landscape of career guidance and decision-making.

.

REFERENCES

[1] **Gottfredson, G.D., and J.L. Holland**, “Dictionary of Holland Occupational Codes,”Psychological Assessment Resources Incorporated, (1996).

[2] **Hearst, M.A., S.T. Dumais, E. Osuna, J. Platt, and B. Scholkopf**, “Support vector machines,” *IEEE intelligent systems*, **13** (4), pp. 18–28 (1998).

[3] **Hicks, S.C., R. Liu, Y. Ni, E. Purdom, and D. Risso**, “mbkmeans: Fast clustering for single cell data using mini-batch k-means,” *PLoS computational biology*, **17** (1), p. e1008625 (2021).

[4] **Kazameini, A., S. Fatehi, Y. Mehta, S. Eetemadi, and E. Cambria**, “Personality Trait Detection Using Bagged SVM over BERT Word Embedding Ensembles,” 2020.

[5] **Kess, J.F.**, “Psycholinguistics: Psychology, linguistics, and the study of natural language,” John Benjamins Publishing, (1992).

[6] **Mairesse, F., M.A. Walker, M.R. Mehl, and R.K. Moore**, “Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text,” *Journal of Artificial Intelligence Research*, **30**, pp. 457–500 (2007).

[7] **Majumder, N., S. Poria, A. Gelbukh, and E.Cambria**, “Deep learning-based document modeling for personality detection from text,” *IEEE intelligent systems*, **32** (2), pp. 74–79 (2017).

[8] **Mehta, Y., S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, and S. Eetemadi**, “Bottom-up and top-down: Predicting personality with psycholinguistic and language model features,” *2020 IEEE International Conference on Data Mining (ICDM),* IEEE (2020).

[9] **Mehta, Y., N. Majumder, A. Gelbukh, and E. Cambria**, “Recent trends in deep learning based personality detection,” *Artificial Intelligence Review*, **53** (4), pp. 2313–2339 (2019).

[10] **Mikolov, T., K. Chen, G. Corrado, and J. Dean**, “Efficient Estimation of Word Representations in Vector Space,” 2013. http://arxiv.org/abs/1301.3781.

[11] **Poria, S., A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay**, "Enhanced SenticNet with affective labels for concept-based opinion mining," *IEEE intelligent systems*, **28** (2), pp. 31–38 (2013).

[12] **Steinwart, I., and A. Christmann**, "Support Vector Machines," Springer Science & Business Media, (2008).

[13] **Wang, H., N. Alanis, L. Haygood, *et al.***, "Using natural language processing in emergency medicine health service research: A systematic review and meta-analysis," *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine*, (2024).

[14] **Wang, Z., C. Fan, and J. Niu**, "Predicting effects of career adaptability and educational identity on the career decision-making of Chinese higher vocational students," *International journal for educational and vocational guidance*, pp. 1–20 (2023).

[15] **Wu, J.**, "Advances in K-means Clustering: A Data Mining Thinking," Springer Science & Business Media, (2012).