Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Career Interests and Personality Traits of Indian Engineering Students: A Dataset and Comprehensive Descriptive Analysis

Minakshi Roy1, Kalpana Sharma*2, and Palash Ghosal3
1,2 Department of CSE, Sikkim Manipal Institute of Technology, Sikkim Manipal University,
3 Department of IT, Sikkim Manipal Institute of Technology, Sikkim Manipal University
**Corresponding author: Kalpana Sharma (e-mail: kalpana.s@smit.smu.edu.in)*

ARTICLE INFO

Received: 02 Nov 2025

Revised: 22 Dec 2025

Accepted: 03 Jan 2026

ABSTRACT

The recent trends in choosing careers among the students from a private engineering college of Sikkim are examined using Big five inventory. The behavioural data is collected from 1st semester onwards with total sample size 1863, comprising participants of 1167 males and 696 females. The main aim is to develop a career interest scale based on Five factor model (FFM) using the standard BFI-10 questionnaire. The data collected through a structured instrument using a Microsoft form as major part of this non-experimental design. Several Descriptive statistics and frequency distribution analysis is carried out to find which parameters have direct major impact on career preferences including gender-based personality traits where men prefer leadership, risk taking and physically demanding jobs, on the other side female's choice inclines to people oriented, stable and creative professions. This personality-based career choices datasets provides insights into the career aspirants engineering students and shows a base for further analytical approaches which leads to career prediction and counselling guidance. The dataset can be explored for several research works which may give important insights for technical institutes to create specialized skill-based courses and career counsellor for career advice. Additionally, it can be used to create particular skill-building exercises, like workshops for engineering or design students that emphasize technical proficiency, communication, creativity, and analysis. This information can be used to track student development and enhance institutional support for all-around student development.

Keywords: behavioural data, career dataset, career prediction, qualitative data

INTRODUCTION

This personality based career interest dataset was collected as part of our research work which will act as a bridge between personality traits of a learner and his job satisfaction. The major role is to determine how personality has influence on the career goals of a student based on demographics along with job interest using BFI-10 questionnaire (Batista & Gondim, 2022; Rammstedt & John, 2007). The dataset comprised of key academic and regional information like name, age, region, semester and career interest, taken by the student's answers through the 10 BFI questionnaire (Fossati et al., 2011; Rammstedt et al., 2013; Weisberg et al., 2011). The factors of the personality based FFM model, viz. openness, conscientiousness, extraversion, agree- ableness, and neuroticism serve as a theoretical basis for this research (Mudhar et al., 2020; Quwaider et al., 2023; Soto & Jackson, 2013). The major role of this dataset is to educate how behavioral traits has large influence on professional choice and decision making by combining all these with career interest (Schmitt et al., 2007; Zell & Lesick, 2022). Several steps which are shown in Figure 1, namely data collection, cleaning, analysis, clustering, and insights are involved in processing the career dataset collected from the students of a private engineering college of Sikkim, India (Erbay et al., 2024). The dataset embodies the major goal. The five important traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experiences are measured by the validated psychometric tool the Big Five Inventory-10 (BFI-10) to

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

connect the career choices with student personality profiles (Semeijn et al., 2020). It offers a broad picture of latest trends in career interests of the institute by collecting responses from different departments and academic years.

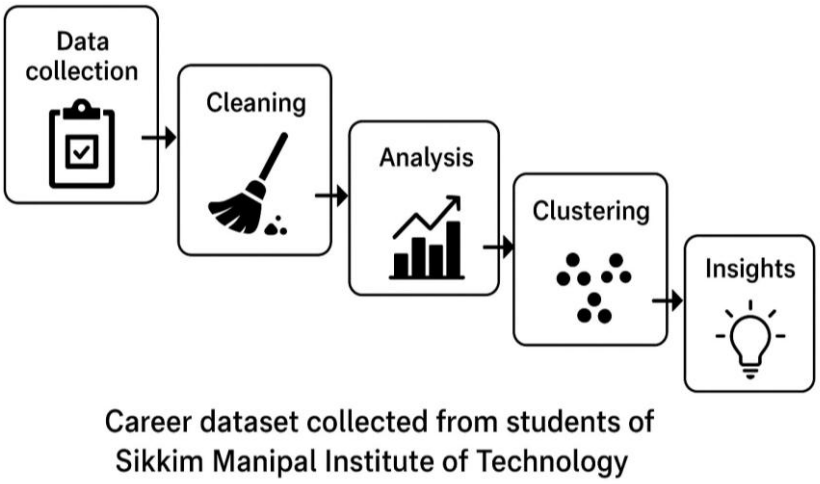

FIGURE 1. Steps involved in processing the career dataset: Data Collection, Cleaning, Analysis, Clustering, and Insights.

The dataset also gives an insightful empirical direction for investigating the link between technical students' personal disposition and vocational inclination by fusing BFI-10 personality assessment with career aspiration measures. These analysis report may help the career development programs, institutional policy, and industry-academia alignment initiatives as a guidance ultimately leading to more satisfied and successful employability enhancement programs at this private engineering college. There are several critical research gaps found in the current literature although extensive research oriented studies carried out on career interests:

1) **Lack of multidimensional, hybrid career- interest datasets.** Existing datasets mostly concentrate on a single dimension, for eg., either career interest, behavirol traits or academic performance indicators. Very few integrated datasets simultaneously capture demographic attributes, BFI-10 personality responses, RIASEC-aligned career interests, academic semester levels, and region wise variations. Such hybrid datasets are essential for understanding how demographic and psychological factors jointly influence career decision-making. The proposed dataset addresses this gap by integrating all these at- tributes into a single, coherent psychometric and behavioral resource.
2) **Limited datasets from Indian engineering institutions.** The majority of publicly available career and personality datasets originate from Western or East Asian populations. Prior studies indicate that cultural and regional contexts significantly influence personality expression and career preferences, making cross-cultural generalization unreliable. Indian (South Asian) engineering students remain severely underrepresented in existing open datasets, resulting in a substantial contextual and cultural research gap.
3) **Scarcity of large-scale psychometric datasets focused on engineering students.** Engineering represents one of the largest educational domains in India, yet publicly accessible large- sample datasets linking psychological traits, demographic variations, and career aspirations are almost non-existent. With data collected from 1,863 undergraduate engineering students, this dataset fills a major gap by offering one of the largest structured psychometric datasets from an Indian technical institution.
4) **Few datasets supporting both descriptive and machine-learning analyses.** Most prior datasets are limited in their analytical scope and do not support comprehensive exploration using both traditional statistical methods and advanced machine-learning techniques. In contrast, this dataset enables descriptive statistics, ANOVA across gender, region, and semester, correlation analysis, unsupervised clustering (e.g., K-means and hierarchical clustering), and supervised machine-learning prediction (e.g., SVM and XGBoost). Consequently, it provides a robust foundation for behavioural modelling and predictive career analytics.
5) **career- interest datasets.** Most existing datasets examine only a single dimension, such as personality traits, career preferences, or academic indicators in isolation. Very few integrated datasets simultaneously

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

capture demographic attributes, BFI-10 personality responses, RIASEC-aligned career interests, academic semester levels, and region wise variations. Such hybrid datasets are essential for understanding how demographic and psychological factors jointly influence career decision-making. The proposed dataset addresses this gap by integrating all these at- tributes into a single, coherent psychometric and behavioral resource.

6) **Limited datasets from Indian engineering institutions.** The majority of publicly available career and personality datasets originate from Western or East Asian populations. Prior studies indicate that cultural and regional contexts significantly influence personality expression and career preferences, making cross-cultural generalization unreliable. Indian (South Asian) engineering students remain severely underrepresented in existing open datasets, resulting in a substantial contextual and cultural research gap.
7) **Scarcity of large-scale psychometric datasets focused on engineering students.** Engineering represents one of the largest educational domains in India, yet publicly accessible large- sample datasets linking psychological traits, demographic variations, and career aspirations are almost non-existent. With data collected from 1,863 undergraduate engineering students, this dataset fills a major gap by offering one of the largest structured psychometric datasets from an Indian technical institution.
8) **Few datasets supporting both descriptive and machine-learning analyses.** Most prior datasets are limited in their analytical scope and do not support comprehensive exploration using both traditional statistical methods and advanced machine-learning techniques. In contrast, this dataset enables descriptive statistics, ANOVA across gender, region, and semester, correlation analysis, unsupervised clustering (e.g., K-means and hierarchical clustering), and supervised machine-learning prediction (e.g., SVM and XGBoost). Consequently, it provides a robust foundation for behavioural modelling and predictive career analytics.

There is a significant scarcity of structured datasets that simultaneously capture the relationship between learner personality traits, demographic diversity, and career choices among Indian engineering students. This gap limits the ability of:

- Educational institutions to design evidence- based and student-centric career counselling interventions.
- Researchers to develop and validate models explaining personality–career relationships.
- Policymakers to understand emerging career trends within engineering education; and
- Data scientists to build robust predictive career- recommendation systems for young Indian learners.

In the absence of such integrated datasets, institutions are unable to meaningfully analyse gender- based differences, regional variations, semester-wise developmental trajectories, or the psychometric factors influencing students' career decisions.

The proposed career–personality dataset offers several distinguishing contributions that make it a unique and valuable resource for behavioural, psychological, and machine-learning research.

1) **One of the largest structured datasets combining career interests with BFI-10 (N = 1863).** The dataset includes demographic at- tributes, declared career preferences, and BFI- 10 personality responses collected from 1,863 undergraduate engineering students. This sample size is substantially larger than those used in comparable studies, providing strong statistical reliability and rich behavioural insights.
2) Cross-cultural psychological understanding with vast demographic diversity.

The students participated from different parts of India: Northeast India, East India, West India, North India, South India and few from outside India. Demographic ANOVA analysis shows how region based background influences the BFI traits demonstrating the importance of cultural responses captured by the questionnaire.

3) Special importance on the Indian engineering students career interests
There is not enough psychometric related data available for engineering students although it's a major academic area in India. This dataset fulfil this limitation created in higher education of this country by offering personality traits, job interests, demographic information, academic aspects and gender based inequalities in document.
4) **Support for advanced multi-level analytics beyond standard psychometric datasets.** As demonstrated in this study, the dataset supports:

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- descriptive statistics (mean, standard deviation, variance);
- correlation analysis;
- ANOVA across gender, region, and semester;
- unsupervised clustering using K-means and hierarchical methods; and
- predictive modelling using SVM and XG- Boost.

Very few datasets in this domain offer such comprehensive analytical capabilities.

5) **An open access, reusable dataset for the future research group.**
In future the researcher can extend the dataset by reusing it as it is openly available on Mendeley data. It inspires comparative, demographic based long term research work. Such transparent dataset about psychometric personality based career interest is limited for Indians.
6) **Career counsellor and policy maker.** The dataset acts as a documentary application initiative for career development. The major roles are gender identification and demographic diversity leads to student's job interests, business demand and resource for career counselling. This also focuses on substantial usefulness of practical aspect.

TABLE 1. Description of the dataset

Field	Value
Name of the dataset	Student Career Interest and Demographics Survey Dataset.
Total respondents	1863 students
Type of data	Table, Chart, Graph, Raw, Analysed, Filtered, Processed
Attributes collected	Gender, Age, Region, Semester, Career Interest, 10 RIASEC-style questionnaire responses
Data collection	Participants responded to ten statements based on the BFI-10 Personality Assessment, each rated on a five-point Likert scale ("Disagree Strongly" to "Strongly Agree"), categorizing personality into Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience.
Data source location	A private engineering college, Sikkim, India
Coordinates	Latitude: 27◦36′07.56′′ N, Longitude: 88◦28′00.12′′ E
Data accessibility	Repository name: Mendeley Data
Data identification number	10.17632/4chz76p5yx.1
Direct URL to data	https://data.mendeley.com/datasets/4chz76p5yx/1

TABLE 2. Attributes with descriptions and examples form the dataset

Attribute	Description	Example
ID	Unique identifier for each participant	2
Gender	Gender of the participant (Male/Female)	Female
Age (Years)	Age of participant in years	20
Region/State	Geographic location of the participant	West India
Semester	Current semester of study	5th Semester
Career Interest	Declared career preference (e.g., Developer, Designer, Researcher)	Design / Human Interaction (empathetic, cooperative, user-focused)
BFI-10 Q1-Q10	Responses to 10 BFI-10 items on a five-point Likert scale	(2, 2, 2, 3, 1, 2, 4, 1, 2, 2)

This dataset, which is derived from undergraduate engineering students, combines personality characteristics determined by the Big Five Inventory-10 (BFI-10) with information about employment interests. Conscientiousness,

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

neuroticism, extraversion, agreeableness, and openness are all examined in this assessment. The dataset offers insights on patterns in career interest within the university and contains a variety of responses from various academic years and departments. The dataset offers a solid foundation for investigating the relationship between personal characteristics and employment preferences by tying students' personalities to their career objectives. The details about the dataset is given in Table 1. The sample of the collected dataset value is shown in Table 2. The findings from this research could enhance employability improvement initiatives by guiding career development programs, influencing institutional policies, and fostering alignment be- tween academia and industry. The contribution of this study is presented as follows: This technical student's psychometric based career interest dataset can be used as a complementary resource for any original research article as it offers raw and structured data for further analysis. Providing new research directions in behavioural traits-based career development, it allows researchers to explore additional patterns beyond the scope of the original study.

RELATED WORK

According to research, Holland's RIASEC model and the Big Five personality characteristics work well together to explain profession choices (Adlya & Zola, 2022). Extraversion is associated with social and enterprising traits, whereas openness is associated with artistic and investigative pursuits. When background and achievement are taken into account, RIASEC interests have a greater influence on students' school choices than the Big Five attributes. Gender differences in personality and career choices are less noticeable in South Asian and Indian cultures because of social expectations and educational opportunities (Shetty et al., 2023). The Big Five and Holland's RIASEC model have been used in several research to examine the relationship between personality and occupational preferences. The Big Five traits and RIASEC dimensions are significantly correlated, despite their theoretical overlap.

While openness is frequently connected to creative and research pursuits, extraversion is associated with social and enterprising domains. Choices for educational paths are more accurately predicted by RIASEC interests than by personality traits alone. When Murwani et al. (2020) looked at the career interest patterns of 981 Indonesian high school students, they found that investigative domains were least popular (3.98) and conventional fields were preferred (42.3) (Mudhar et al., 2020). Gender disparities were evident, with men choosing realistic, artistic, and adventurous careers and women choosing social and traditional ones. According to Nie et al. (2020), the study highlights the importance of occupational interests for students' academic and professional development.

DATA DESCRIPTION

Five basic personality traits—extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experiences—are used to categorize individuals using the Big Five Inventory-10 (BFI-10). A systematic questionnaire was used to gather the dataset from 1863 undergraduate students.

TABLE 3. BFI-10 questionnaire items with corresponding traits and reverse-scoring information. Responses are collected on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree)

Item	Trait	Wording	Reverse Scored
Q1	Extraversion	I see myself as extraverted, enthusiastic.	No
Q2	Agreeableness	I see myself as critical, quarrelsome.	Yes
Q3	Conscientiousness	I see myself as dependable, self-disciplined.	No
Q4	Neuroticism	I see myself as anxious, easily upset.	No
Q5	Openness	I see myself as open to new experiences, complex.	No
Q6	Extraversion	I see myself as reserved, quiet.	Yes
Q7	Agreeableness	I see myself as sympathetic, warm.	No
Q8	Conscientiousness	I see myself as disorganized, careless.	Yes
Q9	Neuroticism	I see myself as calm, emotionally stable.	Yes
Q10	Openness	I see myself as conventional, uncreative.	Yes

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

As seen in Table 3, it includes personality characteristic responses based on the Big Five Inventory-10 (BFI-10), reported job interests, and demographic data. Ten BFI-10 items covering extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experiences were given to each participant on a five-point Likert scale. Data was collected using structured questionnaires, which contained demographic data and responses on the ten scales listed (Steyn & Ndofirepi, 2022). The survey was carried out in this technical college. The repository contains the raw dataset and the questionnaires that go with it for additional study and analysis. The raw data set is presented at **https://data.mendeley.com/datasets/hrbxpk4dh9/1**.

BFI-10 questionnaire items with corresponding traits and reverse-scoring information. Responses are collected on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree).

EXPERIMENTAL DESIGN, MATERIALS AND METHODS

A. DATA ACQUISITION AND EXPERIMENTAL DESIGN

1) Survey Methodology

Stratified sampling was used to choose the participants (Serdiuk & Bazyma, 2021). The dataset was collected using an online form. The survey was distributed to the students of this technical college. This study closely adhered to the ethical standards for research involving human beings. Prior to data collection, all participants were informed of the study's goals, restrictions, and voluntary nature. Each student provided informed consent before responding to the questionnaire. Participation was entirely voluntary, and students were free to leave at any moment without suffering any consequences in their personal or academic lives. Before conducting the surveys, the director of the college and the head of the department (HOD) of computer science and engineering gave their approval.

2) Survey Instruments

- **BFI-10 Questionnaire**: To assess personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism), the survey used the Big Five Inventory-10 (BFI-10) (Zell & Lesick, 2022).
- **Demographic and Career-Related Questions**: Other inquiries gathered data on student names, gender, age, area, semester, and professional interests.
- **Microsoft Form:** Using Microsoft Forms, an online survey was made and sent to students using WhatsApp groups for the semester. This approach guaranteed a broader audience and decreased errors in manual data entering.

3) Data Entry & Processing

- **Responses Export:** For data processing, Microsoft Forms responses were exported straight to CSV format.
- **Data Cleaning:** Before the dataset was finalized, inconsistencies and missing values were examined using R (tidyverse, dplyr) and Python (Pandas, NumPy).

4) Software & Tools Used

- Survey Platform: Microsoft Forms
- Data Entry & Storage: Microsoft Excel
- Data Processing & Analysis:
  -- Python: Pandas, NumPy for data cleaning and structuring
  -- R: tidyverse, dplyr for statistical analysis
  -- SPSS: Used for descriptive and inferential statistical calculations

B. MATERIALS

The Big Five Inventory-10 (BFI-10), developed by Rammstedt & John in 2007, was used in this study to assess the impact of personality factors on engineering students' interests and job inclinations. The BFI-10, which has five personality components, was standardized for comprehensive student examinations. Holland's RIASEC model, a popular framework for classifying professions, was employed in the study to modify the BFI-10 for career interest.

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

This made it possible to have a better knowledge of how personality traits connect to various vocational interest groups and the relationship between personality psychology and occupational behavior. Previous research showing strong cross-cultural validity and satisfactory reliability supports the use of the measure with Indian students. Thirty students from the same university participated in a pilot research to assess the item clarity and internal consistency of a modified scale. With The instrument demonstrated sufficient reliability and met earlier BFI-10 standards, with Cronbach's alpha values ranging from 0.68 to 0.79 across the five personality variables. Construct validity was further ensured by preserving the theoretical relationship between RIASEC career domains and the Big Five traits.

1) Questionnaire:

- Five dimensions of personality traits were evaluated using the Big Five Inventory-10 (BFI-10): extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experiences (Zell & Lesick, 2022).
- Along with other questions on future career objectives and pay expectations, the questionnaire also included demographic questions like age, gender, class, academic stream, and percentage in the previous class.

2) Ethical Considerations:

Informed consent was given by participants, who acknowledged that their answers would be utilized for study. As a result, all participants gave their voluntary agreement to participate, and data collection adhered to ethical standards. This organized data gathering strategy ensured precision and dependability in capturing student responses, enabling further research on personality factors and occupational satisfaction. The objective is to examine the relationships between personality traits, professional goals, and demographic factors in order to understand students' career inclinations and adaptability. This approach ensures dependable data collection and analysis to achieve the study's goals and allows the dataset to be used again in the future for longitudinal or comparative research.

DATA ANALYSIS

This study presents a comprehensive statistical analysis of a dataset from a student personality survey with 1,863 respondents. The study examines the demographic distributions, job field inclinations, and personality trait patterns of Indian students form various academic levels and regions.

A. PERSONALITY ITEMS DESCRIPTIVE STATISTICS

Key personality patterns were observed in the responses and shown in the following tables Table 4:

- **Highest Mean Score:** Q7 (3.26) – indicating this trait is most endorsed by respondents.
- **Lowest Mean Scores:** Q2 (2.05) and Q10 (2.04) – suggesting these traits are least common.
- **Highest Variability:** Q9 (SD = 1.19) – shows the most individual differences.
- **Lowest Variability:** Q8 (SD = 0.94) – indicates more consistent responses.

Scale properties:

- **Response Range:** 1–5 (Likert scale)
- **Quartile Distributions:** Most items clustered around the 2–3 range
- **Central Tendency:** Generally moderate scores across all dimensions

TABLE 4. Descriptive Statistics for Personality Traits

Item	Mean	Standard Deviation	Variance
Q1	2.38	0.98	0.96
Q2	2.05	0.98	0.96
Q3	2.85	1.13	1.28
Q4	2.42	1.06	1.12
Q5	2.53	1.13	1.27

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Q6	2.59	1.09	1.20
Q7	3.26	1.03	1.07
Q8	2.34	0.94	0.88
Q9	2.86	1.19	1.42
Q10	2.04	0.99	0.98

B. REGIONAL DISTRIBUTION OF RESPONDENTS

This is shown in the following Table 5:

TABLE 5. Regional Distribution of Survey Respondents

Region	Students	Percentage
Northeast India	577	30.97%
East India	435	23.35%
West India	388	20.83%
North India	354	19.00%
South India	35	1.88%
Outside India	24	1.29%

C. ACADEMIC LEVEL DISTRIBUTION

This is shown in Table 6.

TABLE 6. Academic Level Distribution

Semester	Students	Percentage
1st Semester	549	29.47%
3rd Semester	526	28.23%
5th Semester	502	26.95%
7th Semester	260	13.96%
6th Semester	8	0.43%
2nd Semester	8	0.43%
8th Semester	5	0.27%
4th Semester	5	0.27%

D. CAREER FIELD PREFERENCES

The data as shown in Table 7 suggests strong preferences for:

- Technical/Engineering fields (37.36%) – aligning with analytical traits.
- Research and Innovation (23.99%) – corresponding to creative and curious traits.
- Specialized technical roles (e.g., Cybersecurity) – reflecting focused expertise.

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

TABLE 7. Career Interest Distribution

Career Field	Students	Percentage
Developer/Engineer (Organized, detail-oriented, disciplined)	696	37.36%
Research/Innovation (Curious, creative, loves learning)	447	23.99%
Cybersecurity Analyst (Calm under pressure, resilient)	211	11.33%
Manager/Coordinator (Outgoing, social, communicator)	209	11.22%
Design/Human Interaction (Empathetic, cooperative, user focused)	151	8.11%
Other variations of above categories	149	8.00%

E. CORRELATION ANALYSIS

It is used for checking the relationship between students' RIASEC scores and the career that they are interested in (or their academic performance) (Dierks et al., 2016).

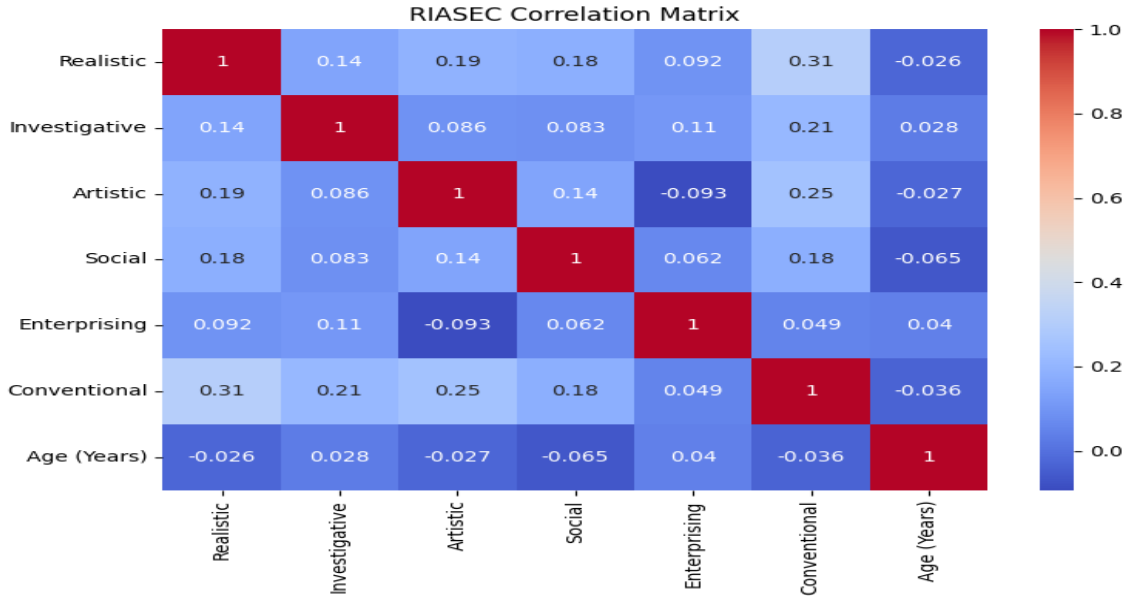

FIGURE 2. RIASEC correlation matrix

Pearson correlation coefficient is used on the dataset whose results are shown in following table, it helps us to identify the changes in one trait that will affect another trait. The result of the correlation analysis shown in Figure 2 depicts that most of the relationships be- tween personality traits are weak (r values lies mostly between 0.09 and 0.31). The strongest relation is there between Realistic and Conventional (r = 0.31), pointing that students who are more practical and like hands-on work also value more structured and routine oriented. Other relationships, like Investigative–Conventional (r = 0.21) and Artistic–Social (r = 0.14), were mild. Age has negligible correlation with any of the personality traits, which suggests that age didn't play a noticeable role in the resulting personality scores in this dataset. The correlation matrix is shown in Table 8.

TABLE 8. Correlation Matrix for RIASEC Domains and Age

Variable	Realistic	Investigative	Artistic	Social	Enterprising	Conventional	Age
Realistic	1.000	0.144	0.190	0.180	0.092	0.310	-0.026

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Investigative	0.144	1.000	0.086	0.083	0.106	0.210	0.028
Artistic	0.190	0.086	1.000	0.142	-0.093	0.249	-0.027
Social	0.180	0.083	0.142	1.000	0.062	0.177	-0.065
Enterprising	0.092	0.106	-0.093	0.062	1.000	0.049	0.040
Conventional	0.310	0.210	0.249	0.177	0.049	1.000	-0.036
Age (Years)	-0.026	0.028	-0.027	-0.065	0.040	-0.036	1.000

F. ANOVA ANALYSIS

It gives the result which helps us analyse if there were significant differences in the resulting career interest scores as shown in Table 7 across different academic years, genders, or departments.

1) ANOVA by Gender

Gender has huge effect on Personality traits which is shown using the following Table 9:

Table shows that gender has a notable effect on personality traits.

TABLE 9. ANOVA Results for RIASEC Traits Across Groups

Trait	**F-value**	**p-value**	**Significance**
Realistic	4.4253	1.2097e−02	Significant
Investigative	25.7246	9.5431e−12	Highly Significant
Artistic	3.1861	4.1559e−02	Significant
Social	4.9854	6.9284e−03	Moderately Significant
Enterprising	14.5445	5.3986e−07	Highly Significant
Conventional	17.0651	4.5281e−08	Highly Significant

2) ANOVA by Semester

All traits except Investigative and Social are strongly affected by semesters. The results are shown in the Table 10. It shows that during the whole of academia Stu- dents' personality traits vary (Sarvottam et al., 2020).

3) ANOVA by Region

TABLE 10. ANOVA Results for Personality Traits

Trait	**F-value**	**p-value**	**Significance**
Realistic	6.6279	9.1980e-08	Highly Significant
Investigative	2.0121	5.0291e-02	Not Significant
Artistic	4.8191	2.1584e-05	Highly Significant
Social	2.0936	4.1229e-02	Significant
Enterprising	3.3137	1.6486e-03	Moderately Significant
Conventional	7.6573	3.8585e-09	Highly Significant

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The ANOVA results indicate that region has a statistically significant effect on all personality traits except the Social trait. Very strong significance levels were observed, particularly for the following traits:

- **Conventional** (p = 1.32e-40) – extremely strong evidence of regional differences.
- **Realistic** (p = 7.07e-25) – highly significant influence of region.

These results highlight that regional background plays a major role in shaping most of the RIASEC traits in the dataset, with the Social trait being the only dimension not significantly affected by region. It shows that geography strongly influences student personality scores. The results are shown using the following Table 11.

TABLE 11. ANOVA Results for Regional Differences in RIASEC Traits

Trait	F-value	p-value	Significance
Realistic	13.3620	7.0769e-25	Highly Significant
Investigative	4.4803	1.0461e-06	Highly Significant
Artistic	7.8722	1.7019e-13	Highly Significant
Social	1.1611	3.0939e-01	Not Significant
Enterprising	5.7039	4.1899e-09	Highly Significant
Conventional	20.9826	1.3244e-40	Highly Significant

G. CLUSTERING ANALYSIS

Unsupervised machine learning method "cluster analysis" is utilized to find patterns or organic groups in the student personality dataset. Based on the students' answers to the ten personality attribute questions, this study seeks to identify unique personality profiles among them.

1) Feature Selection

The analysis focused exclusively on the 10 personality trait variables (Q1–Q10), which were extracted from the complete dataset.

```
# Selecting personality trait variables (Q1-Q10)

personality_data = dataset[['Q1','Q2','Q3','Q4',

                'Q5','Q6','Q7','Q8', 'Q9','Q10']]
```

Listing 1. Personality Data Extraction

2) Data Standardization

Given the importance of equal variable weighting in clustering algorithms, the personality trait data was standardized using z-score normalization.

```
# Standardizing data using z-score normalization
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler() personality_data_scaled =
scaler.fit_transform(
   personality_data)
```

Listing 2. Data Standardization

3) Optimal Cluster Determination

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The elbow method was implemented to identify the optimal number of clusters by examining the within- cluster sum of squares (inertia) across different k values.

```
# Elbow Method for optimal cluster determination
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

inertia = []
K = range (1, 11)
for k in K:
   kmeans = KMeans(n_clusters=k, init='k-means++ ',
      random_state=42)
   kmeans.fit(personality_data_scaled)
   inertia.append(kmeans.inertia_)

plt.plot(K, inertia, 'bx-')
plt.xlabel('Number of clusters (k)')
plt.ylabel('Inertia')
plt.title('Elbow Method for Optimal k')
plt.show()
```

Listing 3. Elbow Method Implementation

4) Final K-Means Implementation

Based on the elbow method analysis, k = 3 was selected as the optimal number of clusters. We are created 3 distinct career interest clusters based on the RIASEC pattern created resulting in group of students with similar vocational profiles. The results are shown in Table 12. This is displayed in Figure 3. Figure 4 shows cluster visualization as per RIASEC traits.

TABLE 12. Distribution of Students Across Career Interest Clusters

Cluster	Count
0	231
1	779
2	853

The resulting cluster sizes are uneven — Cluster 0 small, Cluster 1+2 large. Clustering suggests that we have natural grouping of cluster, but they are not uniform. This result supports that RIASEC traits form meaningful sub-populations.

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

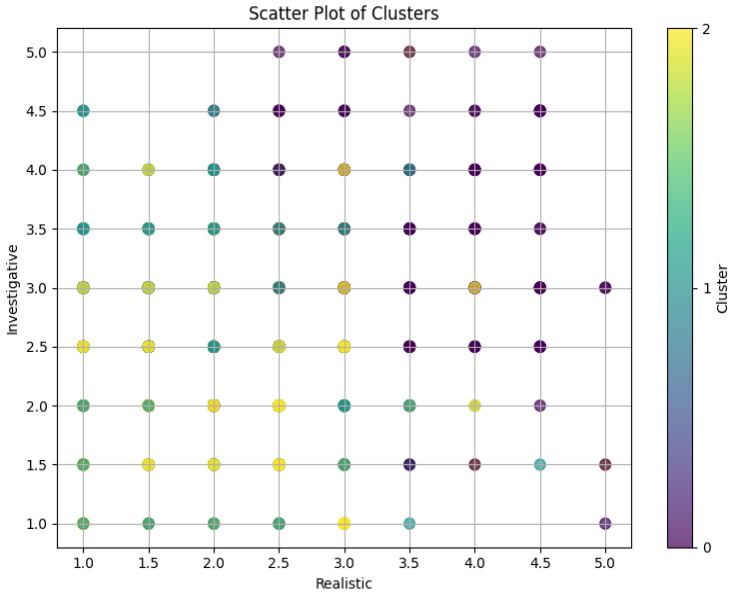

FIGURE 3. K-Means Cluster Separation for RIASEC Dimensions

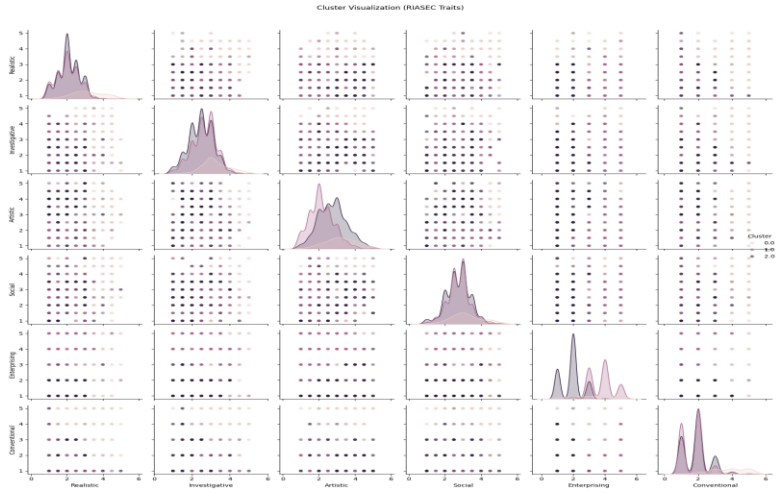

FIGURE 4. Cluster visualization (RAISEC traits)

```
# Final K-Means implementation with k=3

from sklearn.cluster import KMeans


kmeans = KMeans(n_clusters=3, init='k-means++',n_init=10,
     random_state=42)

clusters = kmeans.fit_predict(personality_data_scaled)


# Adding cluster labels to dataset
dataset['Cluster'] = clusters
```

Listing 4. Cluster visualization (RIASEC traits)

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5) Hierarchical Clustering

Dendrogram Visualization: Hierarchical clustering was performed using Ward's linkage method to create a dendrogram showing the hierarchical relationship between observations.

```
# Creating dendrogram for hierarchical clustering

import matplotlib.pyplot as plt
import scipy.cluster.hierarchy as sch


plt.figure(figsize=(8, 5))

dendrogram = sch.dendrogram(
    sch.linkage(personality_data_scaled,
    method=' ward')

)

plt.title('Dendrogram')
plt.xlabel('Students')
plt.ylabel('Euclidean distances')
plt.show()
```

Listing 5. Dendrogram Creation

6) Agglomerative Clustering Implementation

Following dendrogram analysis, agglomerative clustering was applied to create discrete cluster assignments whose values are tabulated in Table 13.

```
from sklearn.cluster import agglomerativeClustering


agg_clustering = AgglomerativeClustering(
    n_clusters=3,
    metric='euclidean',
    linkage='ward'

)


cluster_labels = agg_clustering.fit_predict(data)
```

Listing 6. Agglomerative Clustering Implementation

TABLE 13. Agglomerative Clustering Parameters and Configuration

Parameter	Description
Number of Clusters	3
Linkage Method	Ward

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Distance Metric	Euclidean
Connectivity	None

H. PREDICTIVE MODELLING

We have implemented Machine learning models such as SVM and XGBoost to predict career interest categories based on input attributes shown in Table 14. Target Classes (5):

1) Support Vector Machine (SVM)

The following tables Table 15 and Table 16 show the classification report for career category prediction and performance metrics for SVM classifier:

2) XGBoost

The following tables Table 17 and Table 18 show the classification report for career category prediction and performance metrics for XGBoost classifier:

The result shows that XGBoost performed better than SVM but accuracy still remains low which is shown using the following Table 19 and also displayed in Figure 5:

TABLE 14. Career Domain Preferences Among Students

Career Domain	**Percentage**
Developer/Engineer	38.7%
Research/Innovation	26.0%
Cybersecurity Analyst	13.7%
Manager/Coordinator	11.2%
Design/Human Interaction	10.4%

TABLE 15. Classification Report for Career Category Prediction using SVM

Category	**P**	**R**	**F1**	**Sup.**
Cybersecurity (Resilient)	0.25	0.18	0.21	51
Research / Innovation	0.30	0.26	0.28	97
Design / HCI	0.17	0.33	0.23	39
Developer / Engineer	0.37	0.21	0.27	144
Manager / Coordination	0.16	0.38	0.23	42

TABLE 16. Performance Metrics for SVM Classifier

Metric	**Value**
Accuracy (SVM)	0.2493
Macro F1 (SVM)	0.2415
Macro Avg (P / R / F1)	0.25 / 0.27 / 0.24
Weighted Avg (P / R / F1)	0.29 / 0.25 / 0.25

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Total Samples	373

TABLE 17. Classification Report for Career Category Prediction using XGBoost

Category	**P**	**R**	**F1**	**Sup.**
Cybersecurity (Resilient)	0.21	0.12	0.15	51
Research / Innovation	0.29	0.33	0.31	97
Design / HCI	0.25	0.18	0.21	39
Developer / Engineer	0.36	0.44	0.39	144
Manager / Coordination	0.16	0.12	0.14	42

TABLE 18. Performance Metrics for XGBoost Classifier

Metric	**Value**
Accuracy (XGBoost)	0.3029
Macro F1 (XGBoost)	0.2402
Macro Avg (P / R / F1)	0.26 / 0.24 / 0.24
Weighted Avg (P / R / F1)	0.29 / 0.30 / 0.29
Total Samples	373

TABLE 19. SVM vs XGBoost Performance (Compact Summary

Model	**Acc.**	**MF1**	**Best Trait**
SVM	0.2493	0.2415	Creative (0.28)
XGBoost	0.3029	0.2402	Organized (0.39)

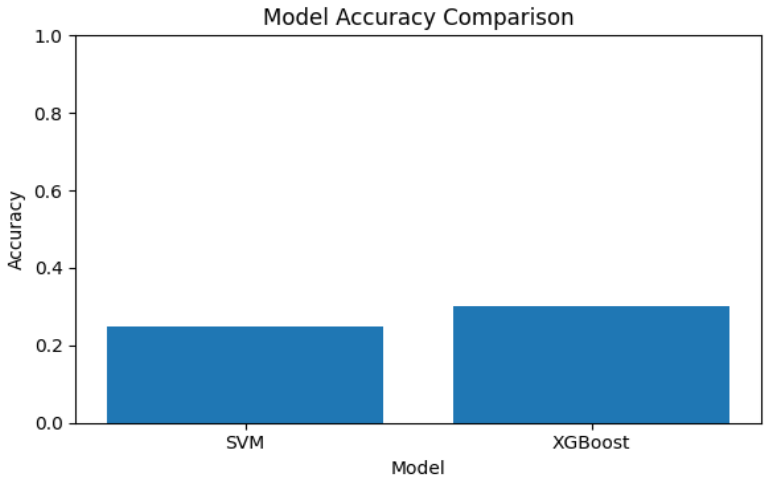

FIGURE 5. Model accuracy comparison

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

I. DESCRPTIVE FINDING

Using the following bar graphs or pie charts as shown in Figure 6, Figure 7, Figure 8 and Figure 10 displaying domain preferences by gender. Presenting comparative statistics such as t-tests or chi-square tests to show how significant gender is for this analysis.

1) T-TEST: Gender Differences in RIASEC Scores

Independent samples t-tests were conducted to ex- amine whether gender has a significant effect on RIASEC personality traits. The results reveal that multiple dimensions show statistically significant differences between male and female students, as summarized in Table 20.

TABLE 20. Gender-Based t-Test Results for RIASEC Dimensions

RIASEC	t-value	p-value	Sig.
Realistic	-2.7928	5.2969e−03	Mod. Sig.
Investigative	-6.8950	8.0078e−12	High Sig.
Artistic	-1.9881	4.6990e−02	Sig.
Social	-2.9078	3.6906e−03	Mod. Sig.
Enterprising	4.9728	7.3904e−07	High Sig.
Conventional	-5.4076	7.6731e−08	High Sig.

2) CHI-SQUARE: Gender vs. Career Domain

Table 21 shows the distribution of career domain preferences across gender groups. These frequencies were further using chi-square tests to determine whether gender significantly influences domain selection as shown in Table 22.

Table 23 presents the domain-wise t-test results evaluating whether gender has a significant effect on specific career domains.

TABLE 21. Career Domain Distribution by Gender

Gender	Resil.	Create.	User-Foc.	Org.	Outg.
Female	97	154	110	236	82
Male	155	323	82	483	124
Other	4	7	1	2	3

TABLE 22. Chi-Square Test: Gender vs Career Field Preference

Statistic	Value
Chi2	51.8961
p-value	1.7627e−08
Dof	8
Significance	High Sig. (p < 0.001)

TABLE 23. Domain-Wise Gender Significance Based on t-Tests

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Domain	t-value	p-value	Significance
Domain 1	-0.928	0.3537	Not Sig.
Domain 2	-3.297	0.0010	Sig.
Domain 3	0.547	0.5845	Not Sig.
Domain 4	-9.521	0.0000	Sig.
Domain 5	-0.858	0.3910	Not Sig.
Domain 6	-2.168	0.0303	Sig.
Domain 7	-2.934	0.0034	Sig.
Domain 8	-0.890	0.3738	Not Sig.
Domain 9	4.973	0.0000	Sig.
Domain 10	-5.408	0.0000	Sig.

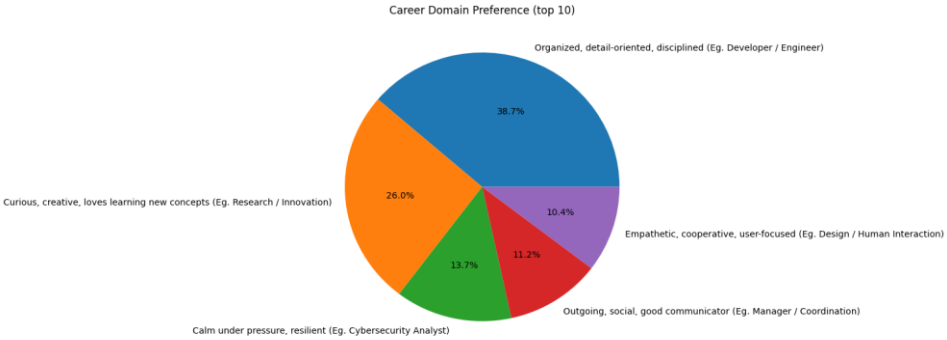

FIGURE 6. Career domain preferences

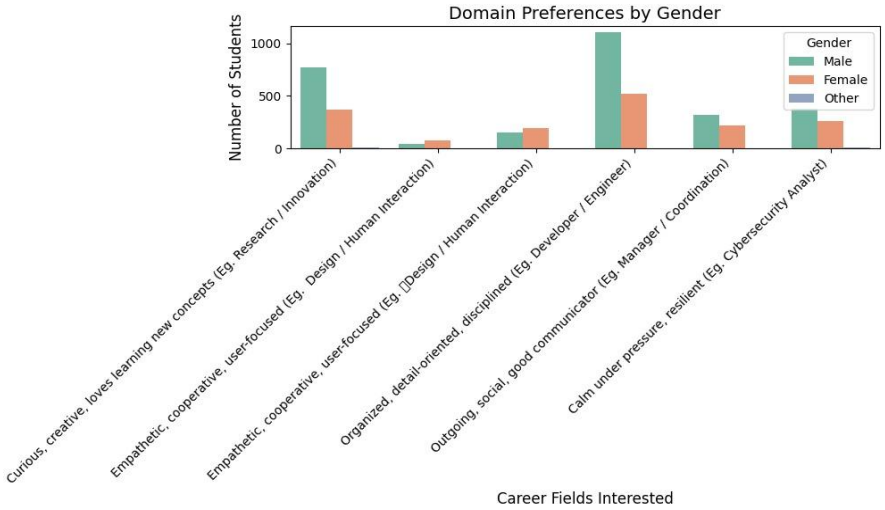

FIGURE 7. Gender wise domain preferences

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

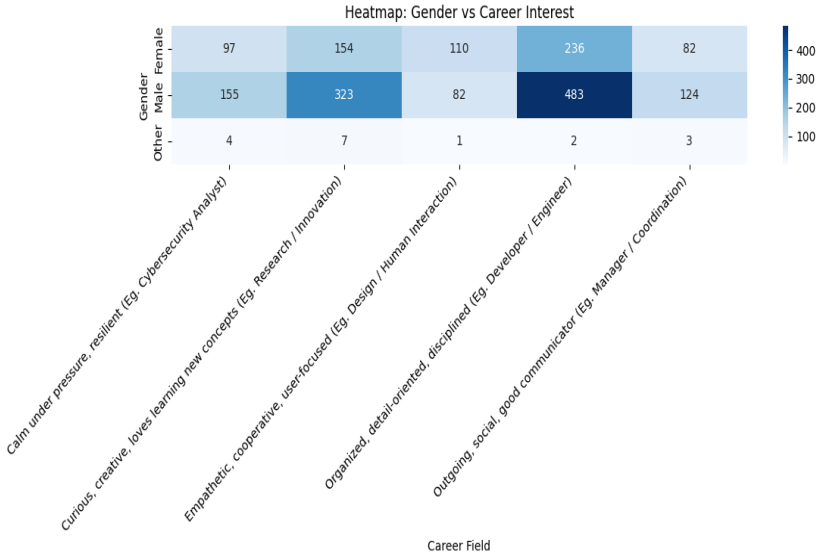

FIGURE 8. Hitman: Gender Vs. Career interest

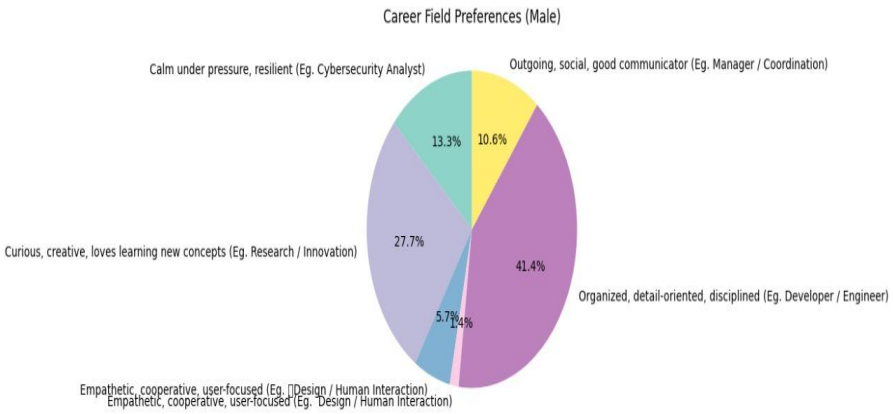

FIGURE 9. Career preferences by male

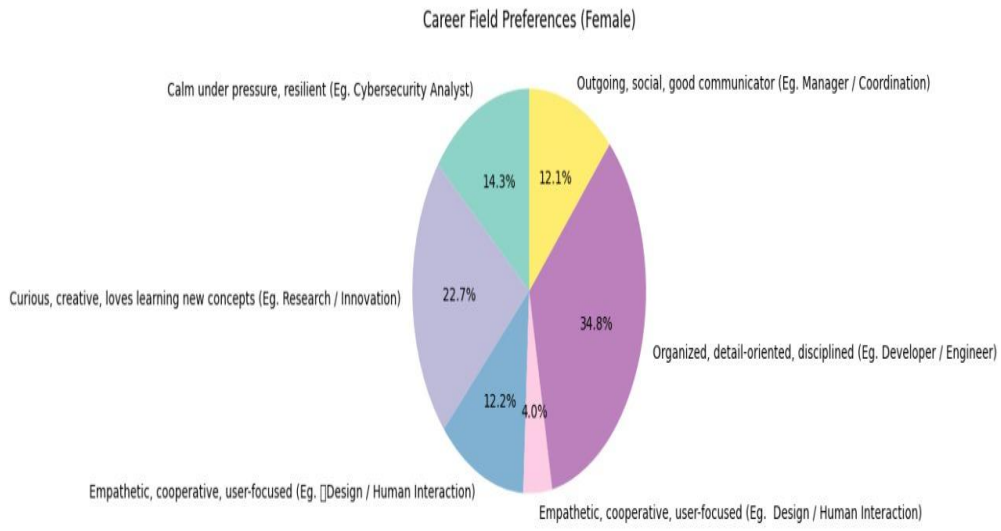

FIGURE 10. Career preferences by female

Journal of Information Systems Engineering and Management
2026, 11(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CONCLUSION

The sample is limited to one institute, and responses are self-reported, which may affect generalizability. Researchers and professionals working in career counselling and educational advice can use this dataset to examine how personality and vocational preferences relate to one another. Significant correlations between job inclinations and the Big Five personality traits are found in preliminary study, indicating the possibility of predictive modelling and tailored career advice. The dataset supports machine learning-based methods in vocational analysis as well as data-driven career counselling. Future work may explore predictive models using personality and demographic data for individualized career counselling.

ACKNOWLEDGEMENT

The authors sincerely thanked the students of the engineering college for their cooperation and involvement in supplying the data needed for this study. We sincerely thank the director of the institute and the head of the Department of Computer Science and Engineering for their cooperation and approval during the data gathering procedure. Their support and direction were crucial to the accomplishment of this study.

REFERENCES

[1] Adlya, S. I., & Zola, N. (2022). Holland's Theory to Guiding Individual Career Choices. Jurnal Neo Konseling, 4(4), 30. https://doi.org/10.24036/00698kons2022

[2] Batista, J. S., & Gondim, S. M. G. (2022). Personality and Person-Work Environment Fit: A Study Based on the RIASEC Model. International Journal of Environmental Research and Public Health, 20(1), 719. https://doi.org/10.3390/ijerph20010719

[3] Dierks, P. O., Höffler, T. N., Blankenburg, J. S., Peters, H., & Parchmann, I. (2016). Interest in science: a RIASEC-based analysis of students' interests. International Journal of Science Education, 38(2), 238–258. https://doi.org/10.1080/09500693.2016.1138337

[4] Erbay, H., Yurttakal, A. H., Dağıstanlı, Ö., & Kör, H. (2024). Advising career choice through tweeter data. Multimedia Tools and Applications, 84(26), 31351–31367. https://doi.org/10.1007/s11042-024-20440-3

[5] Fossati, A., Borroni, S., Marchione, D., & Maffei, C. (2011). The Big Five Inventory (BFI). European Journal of Psychological Assessment, 27(1), 50–58. https://doi.org/10.1027/1015-5759/a000043

[6] Mudhar, Murwani, F. D., Hitipeuw, I., & Rahmawati, H. (2020). Career interest data trends in era information technology of high school students at Surabaya, Indonesia. Data in Brief, 30, 105480. https://doi.org/10.1016/j.dib.2020.105480

[7] Nie, M., Xiong, Z., Zhong, R., Deng, W., & Yang, G. (2020). Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students. Applied Sciences, 10(8), 2841. https://doi.org/10.3390/app10082841

[8] Quwaider, M., Alabed, A., & Duwairi, R. (2023). Shooter video games for personality prediction using five factor model traits and machine learning. Simulation Modelling Practice and Theory, 122, 102665. https://doi.org/10.1016/j.simpat.2022.102665

[9] Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

[10] Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). Methods, Data, Analyses, 7(2), 233–249.

[11] Sarvottam, K., Ranjan, P., & Yadav, U. (2020). Age group and gender-wise comparison of obesity indices in subjects of Varanasi. Indian Journal of Physiology and Pharmacology, 64, 109. https://doi.org/10.25259/IJPP_103_2020

[12] Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The Geographic Distribution of Big Five Personality Traits. Journal of Cross-Cultural Psychology, 38(2), 173–212. https://doi.org/10.1177/0022022106297299

[13] Semeijn, J. H., van der Heijden, B. I. J. M., & De Beuckelaer, A. (2020). Personality Traits and Types in Relation to Career Success: An Empirical Comparison Using the Big Five. Applied Psychology, 69(2), 538–556. https://doi.org/10.1111/apps.12174
[14] Serdiuk, O. O., & Bazyma, B. O. (2021). Адаптація скринінгового опитувальника п'яти факторів особистості BFI-10 та перевірка його діагностичних властивостей на прикладі осіб, які вживають наркотики. Law and Safety, 83(4), 100–110. https://doi.org/10.32631/pb.2021.4.10
[15] Shetty, T., Thomas, N., & Munoli, R. N. (2023). The fundamentals of Indian personality: An investigation of the big five. Indian Journal of Psychiatry, 65(10), 1052–1060. https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_577_23
[16] Soto, C. J., & Jackson, J. J. (2013). Five-Factor Model of Personality. In Psychology. Oxford University Press. https://doi.org/10.1093/obo/9780199828340-0120
[17] Steyn, R., & Ndofirepi, T. M. (2022). Structural validity and measurement invariance of the short version of the Big Five Inventory (BFI-10) in selected countries. Cogent Psychology, 9(1). https://doi.org/10.1080/23311908.2022.2095035
[18] Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender Differences in Personality across the Ten Aspects of the Big Five. Frontiers in Psychology, 2. https://doi.org/10.3389/fpsyg.2011.00178
[19] Zell, E., & Lesick, T. L. (2022). Big five personality traits and performance: A quantitative synthesis of 50+ meta-analyses. Journal of Personality, 90(4), 559–573. https://doi.org/10.1111/jopy.12683