Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 6, pp. 4477–4492, DOI 10.12694/scpe.v25i6.3278

# IN THE EDUCATIONAL NEXUS: UNDERSTANDING THE SEQUENTIAL INFLUENCE OF BIG FIVE PERSONALITY TRAITS, MAJOR IDENTITY, AND SELF-ESTEEM ON ACADEMIC OUTCOMES THROUGH CLUSTERING ALGORITHMS

REBAKAH GEDDAM[*], PIMAL KHANPARA[†], HIMANSHU GHIRIA[‡], AND TVISHA PATEL[§]

**Abstract.** This study investigates the relationship between the Big Five personality traits, major identity, self-esteem, and academic outcomes in education. It uses clustering techniques to examine the impact of these factors on students' academic performance. The research reveals unique patterns when considering personality traits, major identity formation, and self-esteem. The findings highlight the importance of considering these factors when understanding academic attainment trajectories. The study uses popular clustering methods like K-means, DBSCAN, and Hierarchical clustering to reveal latent clusters and provide unique profiles with different combinations of major identity orientations, personality traits, and self-esteem levels. The performance of clustering algorithms is also evaluated using standard assessment metrics. The findings offer insights into the sequential influence of these factors on academic outcomes, guiding the design of student-centric learning materials and providing a framework for promoting successful academic results through an all-encompassing strategy for student development.

**Key words:** Big Five personality traits, K-means Clustering, DBSCAN Clustering. Hierarchical Clustering, Academic Performance.

**1. Introduction.** Personality psychology is a discipline dedicated to examining human personality and individual variations. Emotional, behavioral, and cognitive patterns that are persistent, distinct, and consistent are characteristics that define an individual's attitude [11] [5]. These features not only help us understand human behavior in everyday situations better, but they also help us categorize others who have similar tendencies [22] [4].

Comprehending the complex dynamics of individual differences has become essential to optimizing learning outcomes in education [39]. The realization that learners possess a variety of cognitive and behavioral inclinations has sparked a paradigm shift towards personalized learning strategies, whereas traditional pedagogical approaches have concentrated on standardized methodologies. The comprehensive framework of the Big Five personality traits [27], a psychological foundation famous for clarifying the complex facets of human nature, is essential to enhancing students' learning abilities [36].

Academic achievement correlates with mental health and influences how a student's role changes when they move from student life to a professional career. According to the research [27], students who performed poorly academically also worried about not doing well on examinations and not graduating, which led to increased stress, worry, and sleeplessness and had a long-term effect on their mental health [24]. On the other hand, pupils who excelled in school made a smoother transition from school to the workforce, giving them an advantage in the job market [41].

The framework for categorizing and understanding human personality is the Five-Factor Model (FFM), commonly known as the Big Five personality traits [33]. These traits are broad aspects of personality that cover a variety of particular attributes and actions. The following five factors are:

1. *Openness to Experience (O):* This characteristic demonstrates a person's creativity, curiosity, and openness to new experiences. Individuals with low openness tend to be more conventional, pragmatic, and averse to change, whereas individuals with high openness tend to be creative, daring, and curious [13].

[*]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, India
[†]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, India. (Corresponding author, pimal.khanpara@nirmauni.ac.in)
[‡]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, India
[§]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, India

2. *Conscientiousness (C):* Conscientiousness is a person's level of responsibility, organization, self-control, and goal-directed behavior. While those with low conscientiousness may be more impulsive, disorganized, and irresponsible, those with high conscientiousness are typically organized, dependable, goal-oriented, and conscientious [10].
3. *Extraversion (E):* Extraversion measures how gregarious, confident, gregarious, and extroverted a person is in social situations. Generally speaking, extraverted individuals are gregarious, energetic, and enjoy forming social bonds. They also seek out stimuli in their environment. Conversely, introverts are more quiet, contemplative, and reserved; they enjoy smaller social events or solo pursuits [12].
4. *Agreeableness (A):* Agreeableness is the level of warmth, empathy, cooperation, and care a person displays for others. Individuals with high agreeableness tend to be cooperative, kind, and trustworthy. They also value harmony and preserving healthy relationships. On the other hand, those with low agreeableness levels could interact with others in a more resentful, doubtful, and competitive manner [28].
5. *Neuroticism (N):* Anxiety, sadness, rage, and susceptibility to stress are unpleasant emotions associated with neuroticism (also known as emotional stability). People who score low on neuroticism are typically emotionally stable, calm, composed, and even-tempered. In contrast, those who score high on neuroticism are more likely to experience anxiety, mood swings, and emotional instability [30].

**1.1. Importance of Big Five Personality Traits in the Education Domain.** It is impractical to undervalue the significance of the Big Five personality trait analysis in education because it is an effective tool for comprehending and improving the learning process [7] [16]. An assessment of these traits is essential in an educational setting for the following principal reasons:

- *Personalized Learning:*
  The Big Five personality trait analysis's capacity to identify and consider learners' unique differences is one of its most important achievements. Educators can better meet their students' requirements, preferences, and strengths by customizing assignments, learning environments, and teaching methods based on their understanding of each student's unique personality profile [19].
- *Academic Performance Prediction:*
  Studies have repeatedly demonstrated links between specific personality qualities and successful academic performance. For instance, conscientiousness is frequently linked to improved study habits and grades, while being open to new experiences may indicate the capacity for innovative problem-solving. By exploring these correlations, instructors can predict which students might need extra help or extracurricular activities [29].
- *Increasing Student Engagement:*
  Students' approaches to assignments, interactions with classmates, and involvement in course assignments are all influenced by their personality traits [32]. Teachers aware of these variations can design course contents that enhance students' motivations, interests, and learning preferences, boosting involvement and engagement levels in the classroom [38].
- *Enhancing Social Dynamics:*
  The Big Five personality traits significantly shape interpersonal communications and social interactions in educational environments [3]. With this information, educators may build harmonious peer relationships, settle disputes, and establish an inclusive learning environment that values diversity in opinion and expression.
- *Career Path and Personal Development:*
  Knowing one's personality qualities can help one make important decisions about job inclinations, vocational interests, and personal growth paths. This goes beyond simply succeeding academically [6]. Teachers can assist students in choosing their future paths and utilizing their special skills and talents by incorporating personality tests into career counseling and assistance programmes.
- *Promoting Mental Health and Well-Being:*
  Certain personality qualities, like neuroticism, may make a person more vulnerable to mental health issues like stress and anxiety [8]. Teachers aware of these elements can put them into practice by encouraging students to exercise self-care, supporting their emotional resilience, and giving them access

to the right tools and services [25].

In a nutshell, analyzing the Big Five personality traits provides a comprehensive framework for comprehending the complex relationships among individual deviations, educational outcomes, and learning processes. Educators can use this information to design more effective, individualized, and inclusive learning environments that enable students to succeed academically, socially, and personally.

**1.2. Research Contributions.** The following are the major contributions of this article:

- The Big Five personality traits and their effects on academic success are explored, along with the introduction of personality psychology in the context of educational achievement.
- Comprehensive review and analysis of existing research in personality psychology and academic achievement is presented.
- A comprehensive framework outlining the sequential influence of Big Five personality traits, major identity, and self-esteem on academic outcomes is proposed.
- Popular clustering techniques (K-means, DBSCAN, Hierarchical clustering) are utilized to uncover complex relationships among personality traits, major identity, and self-esteem.
- Based on the obtained results, distinct personality profiles among student participants are identified and characterized.
- Discussion of existing challenges, suggested personalized educational approaches, and future research directions are presented.

**1.3. Taxonomy of the Paper.** The rest of the paper is organized as follows. Section 2 delves into prior research on analyzing personality traits and discussing the key contributions of the existing solutions. Section 3 outlines the objectives of the proposed model, experimental setup, data acquisition process, data cleaning and preprocessing, and key features of our proposed framework to analyze the relations among Big Five personality traits, major identity, and self-esteem for the participating students through statistical analysis. Different Clustering algorithms and their implementations are discussed in Section 3.5. The results are analyzed to derive the linkages between the education outcome and other psychological factors in section 3.6. This section also evaluates the performance of different clustering outcomes using standard metrics. Finally, Section 4 concludes the study by explaining the effect of analyzing student personality attributes on academic achievements.

**2. Related Works.** Incorporating machine learning (ML) approaches into educational research has established novel pathways for comprehending the complex connections between individual differences and educational achievements. This section summarizes the latest developments in machine learning algorithms' application to analyze the Big Five personality traits in educational settings. A thorough overview of state-of-the-art techniques and their impact on educational practices and policies is presented by investigating major studies, methodologies, findings, and implications.

To predict students' academic achievement, the authors of [35] created a prediction model that combines demographic and personality traits. Partial least squares and mathematical modeling of structural equations were used to collect and analyze data from 305 students studying at Al-Zintan University in Libya. Research presented in [17] analyzing the Big Five personality traits of 1735 female and 565 male teacher candidates found that teacher candidates are more extraverted than non-teaching counterparts, highlighting the importance of considering personality group differences in teacher recruitment and training.

The work in [18] used the Big Five Factor model and temporal difference learning analytics to investigate how personality variables affected learning. At the same time, students with high neuroticism had mood swings, and those with higher conscientiousness, openness to experience, and emotional stability performed better. Academic improvement Success requires both conscientiousness and extraversion. Based on the OCEAN big five personality theories, the work in [37] examined how well machine learning algorithms such as SVM, Random Forest, and Neural Network classified students' personalities. With an accuracy of 76%, the Neural Network approach was the most accurate, followed by Random Forest and SVM, with an accuracy of 56% and 40%, respectively. This work aimed to determine if machine learning algorithms may use personality traits to predict students' academic success.

The work in [40] describes a decision tree, gradient boosting decision tree (GBDT), and cat boost approach for personality trait analysis. The Big Five attributes offer dimensional criteria for characterizing people's behaviors and traits through a statistical and semantic combination. The technique uses preprocessed data trained to predict personality types through heatmaps and broken lines analysis. Because personality types are predictable when employing conventional algorithms, the results demonstrate the viability of this approach in psychometric analysis. Another work presented in [21] explores the connection between academic procrastination and personality qualities in young adults. Using a correlational research methodology, it measures procrastination tendencies and the Big Five qualities. The results may inform the creation of individualized programs to help young people attain better academic results and time management skills. Educational institutions, counselors, and lawmakers may find this information helpful in developing interventions that will lessen procrastination.

Age, gender, ethnicity, the Big Five personality traits, and children's self-efficacy were all linked to academic cheating behaviors that authors examined in [42]. According to the findings, boys cheated more than girls, and youngsters cheated less as they grew older. However, no significant correlation was found between cheating and either of the Big Five personality traits or self-efficacy. The results indicate that academic cheating is a problem that emerges in early to middle childhood and that personal traits must develop further before they have strong correlations. [1] examined personality traits' impact on biology students' academic performance in Makurdi, Nigeria. The study involved 384 students and found no significant difference in personality traits based on gender or offered biology. This suggests that gender does not influence personality traits and academic performance. The study recommends improving biology performance and providing male and female students equal opportunities.

The study of the four-dimensional Dark Tetrad personality model in Arab society and its connection to cyber-fraudulent trolling were investigated in [2]. 1093 fourth-year university students majoring in science and literature were involved. The model mediated the relationship between traits of the personality and cyber-fraudulent trolling, and the results indicated a stable model with correlational relationships between the model and the Big Six personality factors. The correlation between gender, academic specialization, and cyber-fraudulent trolling scores was insignificant. This study, [34], investigated the connection between math students' attitudes and teachers' personalities. Teachers in a public school were found to have high levels of conscientiousness, openness, and agreeableness based on data obtained from 118 students. Students expressed poor self-efficacy, moderate curiosity, anxiety, self-concept, and high levels of extrinsic motivation in mathematics. The study's findings, which indicated the significance of instructors in fostering positive surroundings and positive personalities to support learning, showed a somewhat favorable link between teachers' personalities and students' attitudes. Table 2.1 summarizes the input methods, machine learning or deep learning techniques utilized, performance metrics, and personality traits (O = Openness to Experience, C = Conscientiousness, E = Extraversion, A = Agreeableness, N = Neuroticism) for the existing state-of-the-art solutions using the Big Five personality traits in the academics.

An increasing quantity of research in the academic world supports the conclusion that personality traits significantly impact students' academic success. How these traits have an effect is twofold: first, the new human capital theory suggests that certain personality traits boost academic achievement; second, matching personality traits to specific majors creates psychological and motivational incentives. Through their self-efficacy, personality traits within these two pathways impact students' final academic achievements, though the precise nature of this relationship is still unclear. Moreover, there is a major difficulty in the complex matching model between personality traits and professional traits. Consequently, our research avoids the difficulties of matching personality traits with majors by focusing on a specific set of professional students. The emphasis is rather on examining how a student's personality traits affect their core identity in a certain professional context. Adding a new, factual foundation to this research domain is the aim of investigating the dual mediating chain impact between major identity and self-efficacy.

**3. Materials and Methods.** The main objective of our proposed framework is to analyze the impact of Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) among undergraduate college students, especially on their learning outcomes. We used various clustering techniques to identify the personality attributes of the participating students. This section describes the experimental setup, data collection, and preprocessing steps, highlighting our proposed framework to meet the objectives.

Table 2.1: Analysis of the Existing Approaches using Big Five Personality Traits for Education Domain

| Paper | Input Method | Statistical Method | ML/DL Model | Performance Metrics | Personalities |
|---|---|---|---|---|---|
| [35] | Survey questionnaire with Likert scale | Partial Least Squares (PLS), Structural Equation Modeling (SEM) | - | Accuracy, Efficiency of Model Performance | O, C, E, A, N |
| [17] | Computer-assisted telephone interviews | MANOVA, Confidence Intervals | - | Internal Consistency, Retest Reliability | O, C, E, A, N |
| [18] | Questionnaire, Aptitude Tests | Structural Equation Modeling (SEM), Descriptive and Analytical Approach | Random Forest, J48, Naive Bayes | Personality traits correlation, prediction accuracy | O, C, E, A, N |
| [37] | 50-item Likert Scale | Descriptive Statistics, Machine Learning Algorithms | Support Vector Machine (SVM), Random Forest (RF), Neural Network (NN) | Accuracy (76%, 56%, 40%) | O, C, E, A, N |
| [40] | Pre-processing, Data Conversion, Data Cleansing | Analysis in broken-lines and Heatmap | Decision Trees, GBDT, Cat Boost | Predictive accuracy: Decision Trees - 0.52, GBDT - 0.68, Cat Boost - 0.78 | O, C, E, A, N |
| [21] | Big Five Inventory-10 (BFI-10), General Procrastination Scale | Pearson Correlation Coefficients, T-tests | - | - | O, C, E, A, N |
| [42] | Zoom Recruitment | Correlation Analysis, T-tests | - | - | O, C, E, A, N, Self-efficacy |
| [1] | Five-Factor Inventory Questionnaire (FFIQ), Biology Performance Test (BPT) | Mean, Standard Deviation, ANOVA, T-tests | - | - | O, C, E, A, N |
| [2] | Dark Tetrad four-dimensional personality scale, Cyber Fraudulent Trolling scale, Big Six personality factors scale | Correlational Analysis, Mediation Analysis | Linear Regression, Mediation Analysis | Correlation Coefficients, Mediation Effects | O, C, E, A, N |
| [34] | Questionnaires from 118 randomly selected Students | Correlation Analysis | - | Correlation Coefficient | O, C, E, A, N |

**3.1. Experimental Setup.** During the data collection process, 1016 undergraduate students from an engineering college in Gujarat volunteered to participate. These students were between 18 and 24 years old. Of the 1016 participants, 642 were male, and 374 were female. All participants were from the same technical background and thus had the same fundamental knowledge of that technical domain.

The participants were given a standardized questionnaire based on the Big Five Inventory (BFI) and similar

Table 3.1: Questionnaire Used in the Proposed Framework

| | **Extraversion** | **Neuroticism** | **Agreeableness** | **Conscientiousness** | **Openness to Experience** |
|---|---|---|---|---|---|
| Q1 | I am a party animal. | I don't often feel down. | I make fun of people. | I focus on the details. | I have really good ideas. |
| Q2 | I'm not a big talker. | I get upset easily. | I have no interest in the troubles of other people. | I refuse to do my work. | I have a lot on my mind. |
| Q3 | Being among people makes me feel at ease. | I'm easily stressed out. | I don't really care about other people. | I finish chores immediately. | My imagination is not very strong. |
| Q4 | Being the center of attention doesn't bother me. | My emotional swings are frequent. | I am sensitive to the feelings of others. | I adhere to a schedule. | I take some time to think things through. |
| Q5 | I strike up discussions. | I quickly get annoyed. | My care for other people is little. | I'm prepared at all times. | I have trouble grasping abstract concepts. |
| Q6 | At gatherings, I converse with a wide range of people. | A lot of things bother me. | I understand the emotions of others. | I work meticulously. | Abstract Ideas don't appeal to me. |
| Q7 | I prefer to remain unnoticed. | I get depressed often. | People stimulate my curiosity. | I screw up a lot of stuff. | My vocabulary is extensive. |
| Q8 | I don't have much to say. | I get mood swings a lot. | My heart is gentle. | I enjoy discipline and organization. | My understanding of things is swift. |
| Q9 | I keep quiet around new people. | Most of the time, I am relaxed. | I make people comfortable. | I leave my stuff around. | I make use of tricky words. |
| Q10 | I prefer not to be the center of attention. | I'm easily agitated. | I make time for other people. | I'm terrible at putting things back in their place | My imagination is powerful. |

scales. The questionnaire consisted of 50 questions/statements in total. There were 10 questions/statements for each personality trait from the Big Five model. Each Big Five personality trait is evaluated using a sequence of statements or questions in the questionnaire. Participants were required to respond to each statement or question using a Likert scale (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree). Participants were made aware of the purpose of the study and the fact that their answers would be recorded before the exam started. They were asked to affirm their consent after finishing the test.

The questionnaires were circulated via email or other digital channels, accompanied by a clear explanation outlining the survey's purpose and significance. Upon receiving student responses, the data was stored on our computer for thorough analysis. Furthermore, to ensure accessibility and future reference, the results were also stored in the cloud. This dataset now serves as the fundamental repository for training and testing the machine learning model, promising valuable insights into the determinants of undergraduate academic performance. The questions included in the questionnaire are shown in Table 3.1.

**3.2. Data Preprocessing.** The steps involved in preparing data from a microscopic perspective for statistical analysis are explained in this subsection. The feature selection reasoning is described in detail, focusing on the standards for limiting the selection to particular columns. We investigate the conversion of categorical values into numerical representations and discuss how this could affect reliable clustering analyses. The crucial significance that data preparation plays in maintaining the integrity of subsequent analyses is highlighted in this section. The procedure includes converting and modifying unprocessed data to guarantee its quality, relevance, and compatibility for further investigation.

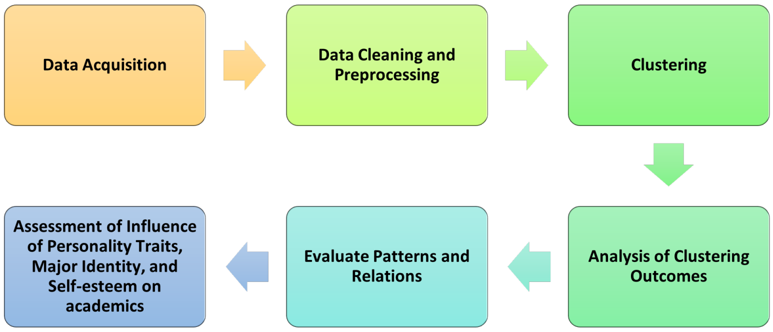


Fig. 3.1: Proposed Model Architecture

After the file has been downloaded from the cloud server, the data is scrutinized to ensure it is in an easily processed format. This procedure involves removing redundant or irrelevant columns and unwanted columns. The data is also cleaned up to remove outliers, inconsistent data, and missing values. The category responses are converted into numerical form so clustering algorithms can utilize them. The data is normalized as needed to guarantee that each feature has the same scale.

We finally divide the columns based on a question list to categorize the data according to particular questions or themes. After grouping and categorizing the questions, the extraversion attribute was assigned to questions 1 through 10 in our form. Likewise, questions 11–20 assess neuroticism, 21–30 assess agreeableness, 31–40 assess conscientiousness, and 41–50 assess openness. We remove any missing values from the data collection to ensure the data is accurate and complete.

**3.3. Proposed Framework.** This section examines the complex relationships between critical psychological factors and how they affect academic performance. We analyze the association between personality traits and academic performance based on the Big Five personality traits model: neuroticism, agreeableness, extraversion, and conscientiousness. We integrate major identity and self-esteem into the paradigm to fully comprehend their mediating roles. We aim to identify unique patterns and groupings in the data using advanced clustering techniques. We offer insights into how combinations of personality traits, major identity, and self-esteem influence different academic outcomes. Our framework aims to provide detailed insights into the complex connection between psychological traits and academic success in the educational setting.

Based on the synthesis of prior research, it has been observed that conscientiousness, as one of the Big Five personality traits, often exhibits a positive direct impact on students' academic performance. Additionally, it can indirectly enhance academic performance by fostering improved self-esteem. Other dimensions of personality traits may have different effects depending on certain contextual elements like circumstances, cultural variations, and professional backgrounds. Major identity, on the other hand, tends to contribute to elevated self-esteem and positively affects academic achievement. The mechanism through which personality traits affect major identity is unclear. Still, there is a potential to measure the alignment between personality attributes and majors for students in the given academic disciplines within particular environments.

Building upon these insights, as shown in our proposed architecture in Fig.3.1, aims to examine the effects of personality attributes on major identity, academic self-esteem, and academic performance. Moreover, the combined impact of both factors' chain mediating effects will be analyzed to observe the possible implications of assessing how personality traits influence academic success.

Here, the direct impact of personality attributes on academic performance is denoted by $D_{pa}$; the independent mediating effects of major identity and self-esteem are denoted by $I_i$ and $I_e$, respectively; the joint chain mediating effect of major identity and self-esteem is denoted by $I_{ie}$; and the total impact of personality attributes on academic performance is denoted by $T_{pa}$. The coefficient vector of attributes associated with a

Table 3.2: Statistical Significance of Samples Aggregated based on Personality Traits

| Personality Traits | Mean | Median | Standard Deviation |
|---|---|---|---|
| Extraversion | 32.30556 | 32 | 3.111979 |
| Neurotic | 29.27778 | 30 | 4.865474 |
| Agreeable | 33.97434 | 34 | 4.637148 |
| Conscientious | 31.91667 | 31.5 | 3.555397 |
| Openness | 31.15278 | 30 | 3.928739 |

personality type is represented as $\beta$. $x_i$ represents the vector of independent variables related to an individual $i$.

$$D_{pa} = F(\beta_{pa}, x_i) \tag{3.1}$$

$$I_i = F(\beta_{pi} \cdot \beta_{ia}, x_i) \tag{3.2}$$

$$I_e = F(\beta_{pe} \cdot \beta_{ea}, x_i) \tag{3.3}$$

$$I_{ie} = F(\beta_{pi} \cdot \beta_{ie} \cdot \beta_{ea}, x_i) \tag{3.4}$$

$$T_{pa} = D_{pa} + I_i + I_e + I_{ie} \tag{3.5}$$

**3.4. Statistical Analysis.** To find the relationship between the variables in our proposed assessment approach, we first constructed a structural equation model for each of the following: academic performance, self-esteem, major identities, and the Big Five personality traits. These effect estimation results are then used to compute the chain mediation effects produced by both variables and the mediation effects of the significant identity and self-esteem factors. Initially, there were a total of 50,800 samples in the dataset. After removing 49 inadequate samples and 67 outliers, 50,684 observations remained. A more thorough analysis of the psychometric characteristics of the Big 5 Personality traits might be conducted using these data. The statistics of the aggregated personality traits are shown in Table 3.2.

**3.5. Clustering Techniques.** Using various clustering techniques, grouping students according to their academic achievement indicators and personality trait profiles is essential. Clustering allows for examining students' academic performance metrics and personality traits, gaining valuable insights that can be used to improve academic advising, prevent dropouts, personalize curriculum, support research and policy development, and allocate resources optimally in educational settings [9]. Our proposed framework implements popular clustering techniques such as K-means, DBSCAN, and Hierarchical clustering.

**3.5.1. K-means Clustering.** K-means clustering is a useful tool in student support systems and educational research when it comes to investigating the Big Five personality traits for academic achievement [23]. With this approach, different student profiles or clusters were found according to their academic performance and personality characteristics. The dataset was divided into five clusters, each representing a grouping of students with related attributes. This enables educators to better understand the relationship between academic success and personality traits. Students were placed into groups using K-means clustering based on shared personality qualities, including neuroticism, agreeableness, extraversion, conscientiousness, and openness, as well as the academic achievement indicators accompanying them. This segmentation made academic advice, early intervention, dropout prevention, and curriculum modification possible, making targeted support techniques and personalized interventions possible. However, while using this approach to analyze the Big Five personality traits for academic achievement, it's important to consider the constraints of K-means clustering, such as sensitivity to initial centroid location and the requirement to provide the number of clusters beforehand.

Table 3.3: Statistical Parameters of Student Cluster groups

| | Cluster 1 | | Cluster 2 | | Cluster 3 | | Cluster 4 | | Cluster 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Extraversion | 32.2142857 | 2.85803557 | 32.2068966 | 3.17728262 | 33.3571429 | 2.91809975 | 30.7692308 | 2.45431633 | 31.6722312 | 2.7965321 |
| Neurotic | 23.7857143 | 4.42684298 | 31.2068966 | 3.7174269 | 31.4285714 | 4.38690072 | 27.9230769 | 0.5756396 | 27.5423365 | 3.1973822 |
| Agreeable | 34.7857143 | 3.50873545 | 33.8965517 | 2.15510345 | 34.8571429 | 4.06829453 | 30.3846154 | 2.40315375 | 30.9427364 | 2.8753924 |
| Conscientious | 29.7857143 | 3.36321639 | 31.1724138 | 2.90147365 | 34.7857143 | 3.42633856 | 32.6153846 | 3.1265233 | 31.2788427 | 3.1635921 |
| Openness | 30.92857 | 3.514547 | 29.89655 | 2.368541 | 35 | 3.70328 | 28.61538 | 2.338259 | 29.7146581 | 2.697812 |
| N = 50684 | 9859 | | 13896 | | 11291 | | 8846 | | 6792 | |
| Percentage | 19.45 | | 27.42 | | 22.28 | | 17.45 | | 13.4 | |

**3.5.2. DBSCAN Clustering.** An effective method for comprehending the complex relationships between students' personalities and their academic success is to apply DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering to the analysis of Big Five personality traits for academic performance [26]. DBSCAN is very useful for handling noise in the data and finding clusters of any shape, in contrast to conventional clustering techniques. Even in datasets with irregularly shaped or overlapping clusters, DBSCAN may identify groups of students with comparable personality trait profiles and academic achievement measures by identifying clusters based on the density of data points rather than fixed centroids. This makes it possible for researchers and educators to find subtle links and patterns that other clustering techniques might miss.

Based on their academic performance and personality features, students were categorized into clusters by DBSCAN clustering. This enabled the development of individualized support plans and targeted interventions catering to each cluster's unique needs. However, it's crucial to remember that DBSCAN may not function as well in datasets with different densities or high-dimensional spaces. It requires careful adjustment of its parameters. Despite these limitations, DBSCAN clustering effectively reveals significant insights into the relationship between personality traits and academic achievement in learning environments.

**3.5.3. Hierarchical Agglomerative Clustering (HAC).** When the Big Five personality traits are analyzed for academic performance, hierarchical clustering provides a holistic approach to comprehending the complex relationships between students' personalities and their academic accomplishments [14]. The dendrogram, a hierarchical tree-like structure of clusters created by hierarchical clustering instead of other clustering techniques, shows the nested interactions between clusters at various granularities. This enables educators and academics to systematically investigate the variety of personality profiles and academic performance measures among the student community. Hierarchical clustering finds student clusters with comparable psychological trait profiles and academic achievement metrics by iteratively merging or dividing clusters based on similar metrics. This makes it possible to identify common traits and behaviors among students.

By using hierarchical clustering, teachers may better understand the diversity of their learners and create individualized support plans and focused interventions suited to the requirements of various groups. On the other hand, hierarchical clustering can be computationally demanding for large datasets and may necessitate careful consideration of linking criteria and distance measurements. Although all of this, hierarchical clustering effectively offers significant insights into the complex connection between academic success and personality traits in educational settings.

We have used global and native clustering for K-means, DBSCAN, and Hierarchical models. Native clustering involves grouping within smaller, localized regions, accounting for spatial or temporal variation in the data distribution, whereas global clustering concentrates on dividing the entire dataset into clusters. We have trained the global model on the Big Five personality dataset and the native model on the dataset generated through our survey. The assignment of participants to different clusters based on their responses is depicted in Table 3.3.

**3.6. Result Analysis.** This section analyzes the results obtained for K-means, DBSCAN, and Hierarchical clustering techniques employed to categorize participants based on their personality traits as extroverted, neurotic, agreeable, conscientious, and open to experience.

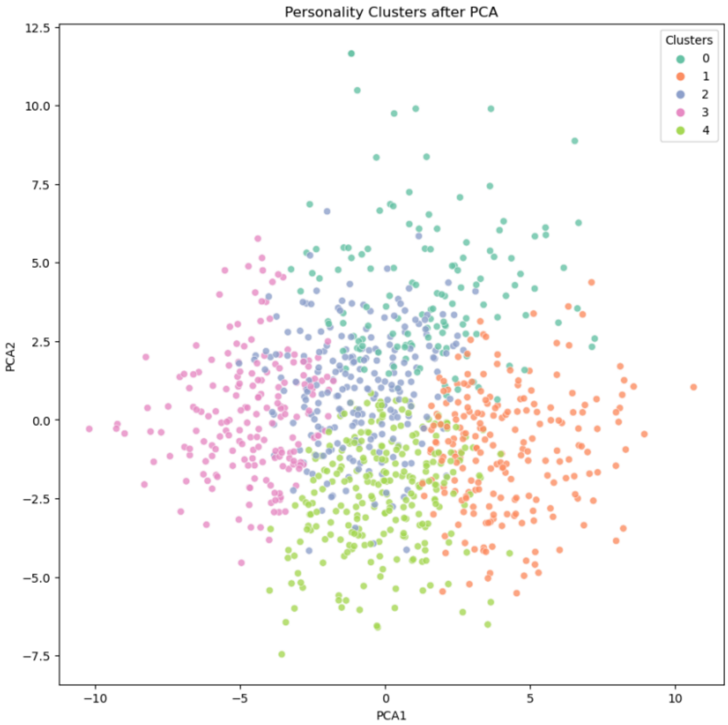


Fig. 3.2: Visualization of Clusters using Principal Component Analysis

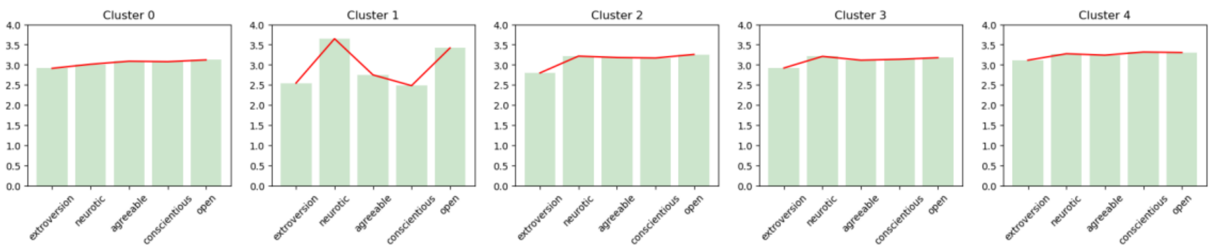


Fig. 3.3: Visualization of Global Model Clusters using K-means

**3.6.1. Principal Component Analysis (PCA).** Principal Component Analysis (PCA) [20] is a popular dimensionality reduction technique used in data analysis and machine learning. Here, the objective of performing PCA is to preserve the most significant information while converting high-dimensional data into a lower-dimensional space. Principal components, or the orthogonal directions in the data that capture the most variation, are found by PCA to accomplish this. These major components are arranged according to how much variance they explain to reduce dimensionality while preserving as much variance as feasible.

Our proposed framework uses PCA to visualize the data and identify potential clusters. The outcome of the same is shown in Fig.3.2.

Fig.3.3 and Fig.3.4 depict the visualizations of global and native clusters when K-means clustering is applied. Five clusters are generated to reflect the five personality traits. K-means clustering results are assessed using Adjusted Random Index (ARI), Silhouette score, Davies Bouldin Index (DRI), and Calinski-Harabasz Index (CHI) metrics and are described later in this section.

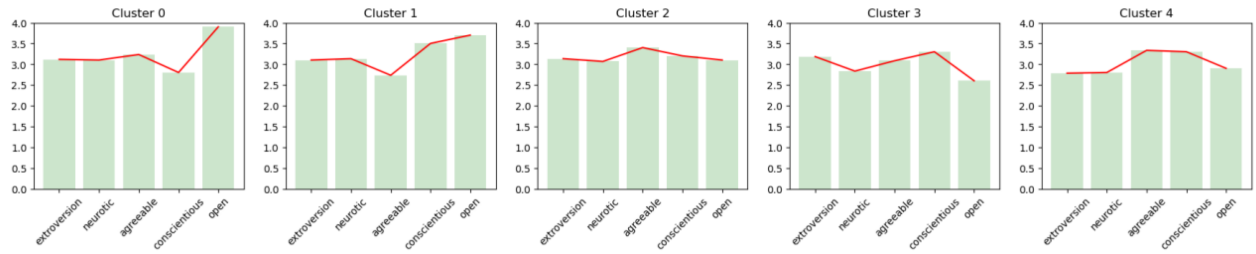


Fig. 3.4: Visualization of Native Model Clusters using K-means

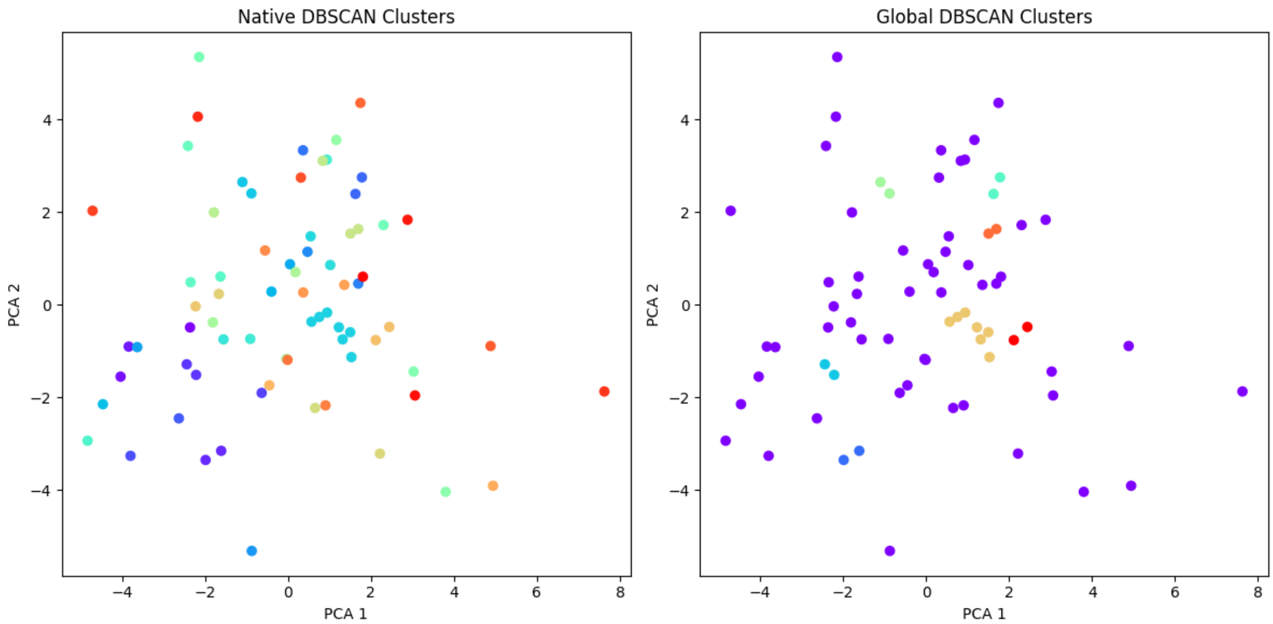


Fig. 3.5: Native and Global DBSCAN Clusters

Fig.3.5 shows the global and native model clusters when the DBSCAN clustering technique is used.

Fig.3.6 represent the global and native model clusters when the hierarchical clustering (HAC) technique is implemented. HAC clusters are visually represented in a hierarchical tree-like structure, dendrogram, and are shown in Fig.3.7.

We use different metrics such as Adjusted Random Index (ARI), Silhouette Score, Davies Bouldin Index (DBI), and Calinski-Harabasz Index (CHI) [14] [31] to assess the performance of the clustering algorithms used in our proposed framework. The following descriptions discuss the reasons behind choosing these performance evaluation metrics and the outcome of the clustering assessment.

**3.6.2. Adjusted Random Index (ARI).** The Adjusted Rand Index (ARI) [15] compares two clusterings of the same dataset in terms of similarity. The agreement between sample pairs concerning their cluster assignments between the two clusterings under comparison is computed. Concerning cluster assignments, the ARI considers agreement and disagreement, yielding a normalized measure from -1 to 1. Strong agreement between the clusterings is indicated by a value around 1. In contrast, random agreement is implied by a value close to 0, and strong disagreement is indicated by a number close to -1. The "adjusted" part of the ARI considers the predicted agreement resulting from chance. Because of this, it is beneficial for comparing

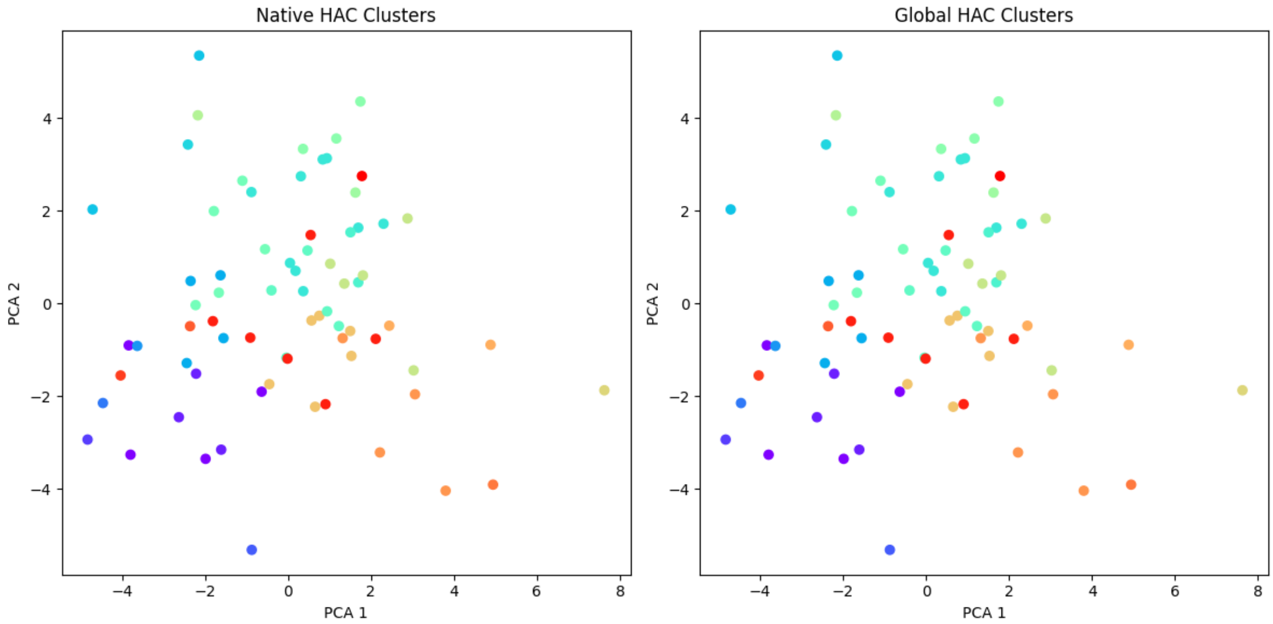


Fig. 3.6: Native and Global HAC Clusters

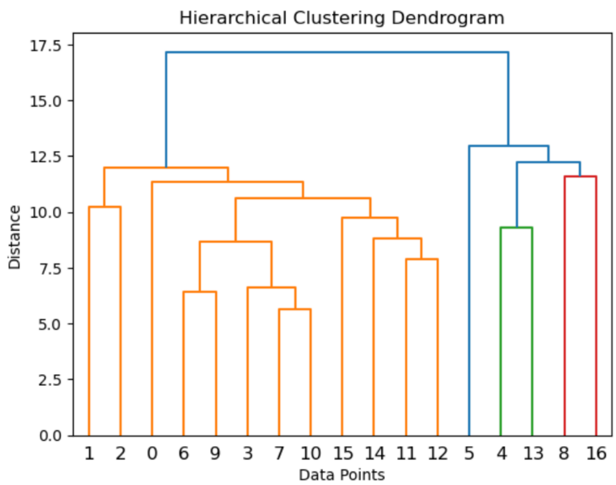


Fig. 3.7: Hierarchical Clustering: Dendrogram

clustering outcomes or assessing clustering algorithms in situations where ground truth labels are unavailable.

**3.6.3. Silhouette Score.** The quality of the clusters created by a clustering algorithm is assessed using a metric called the Silhouette Score [15]. The degree to which each data point, relative to other clusters, fits into the designated cluster is measured. From -1 to 1, the Silhouette Score is a numerical representation of the distance between a data point and other points in the same cluster. A high score denotes a well-clustered data point. If a point's score is almost zero, it may be close to the line dividing two clusters. The average Silhouette Score measures the clustering algorithm's overall performance across all data points. Higher average scores indicate better-defined clusters.

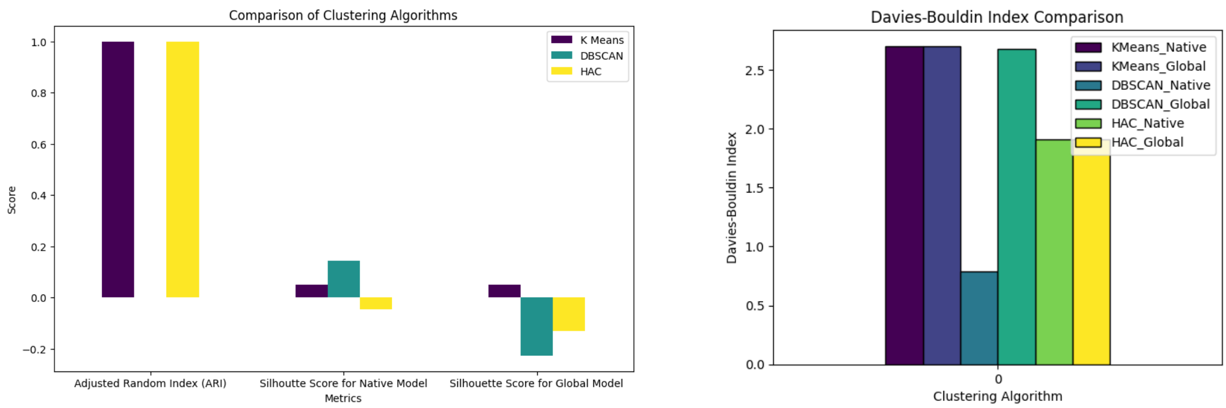


Fig. 3.8: Silhouette Score and Davies Bouldin Index Comparison for Clustering Algorithms

**3.6.4. Davies Bouldin Index (DBI).** The Davies–Bouldin Index (DBI) [15] is a metric used for evaluating the quality of clustering in a dataset. The distance between clusters and the clusters' compactness are quantified. Better clustering, when clusters are closely spaced and densely populated, is indicated by lower DBI values. In addition to considering average cluster size, the index considers the average similarity between each cluster and its most similar cluster. Through the DBI, a single score representing the overall quality of the clustering is generated by computing the ratio of these two parameters across all clusters.

**3.6.5. Calinski-Harabasz Index (CHI).** The Calinski-Harabasz Index (CHI) [15] is a metric used to assess the quality of clustering in a dataset. Higher values indicate better clustering. It measures the ratio of within-cluster dispersion to between-cluster dispersion. To be more precise, the CHI simultaneously assesses both the compactness of clusters and the distance between them. It calculates the ratio of the within-group dispersion to the between-group dispersion; higher values denote more compact and well-defined clusters.

**3.6.6. Evaluation of Clustering Results.** This section shows how various metrics are utilized to evaluate the quality of clusters generated through K-means, DBSCAN, and Hierarchical clustering.

Fig.3.8 depicts the comparison among K-means, DBSCAN, and Hierarchical clustering algorithms for global and native models using Davies Bouldin Index (DBI) and Silhouette scores.

The comparison of clustering algorithms is done using the Adjusted Rand Index (ARI), a metric quantifying the similarity between two clusterings while correcting for chance. ARI is computed for K-Means, DBSCAN, and Hierarchical clustering against ground truth labels, providing a standardized measure of clustering agreement. A high ARI indicates a robust agreement between the predicted and actual clusters. These facts can be seen in Fig.3.8.

As shown in Fig.3.8, the Silhouette Score is used for cluster cohesion and separation in comparative analysis. This metric assesses the results of all three clustering algorithms to provide insights into the internal consistency of clusters. A higher silhouette score indicates distinct, well-defined clusters. For native models, DBSCAN outperforms the other two clustering algorithms. For global models, the K-means algorithm performs better than the other two.

The Davies-Bouldin Index (DBI), a cluster compactness and separation measure, is applied to clustering results. As depicted in Fig.3.8, a lower index signifies better clustering, indicating well-separated and compact clusters. Here, DBSCAN implies better clustering for native models; hierarchical clustering outperforms the other two clustering algorithms for global models.

The Calinski-Harabasz Index (CHI) is used to assess the ratio of between-cluster variance to within-cluster variance, and the values of CHI for the same are included in Fig.3.9 for the three clustering algorithms. This index helps measure how separable and compact a cluster is. The comparison of CHI values provides insights into clustering algorithms' ability to form cohesive and distinct clusters.

| | K Means | DBSCAN | HAC |
|---|---|---|---|
| **index** | | | |
| **Adjusted Random Index (ARI)** | 1.000000 | 0.000000 | 1.000000 |
| **Silhoutte Score for Native Model** | 0.049233 | 0.142936 | -0.046198 |
| **Silhouette Score for Global Model** | 0.049233 | -0.226687 | -0.131080 |
| **Davies Bouldin Index for Native Model** | 2.702828 | 0.791410 | 1.909324 |
| **Davies Bouldin Index for Global Model** | 2.702828 | 2.675913 | 1.909324 |
| **Calinski-Harabasz Index for Native Model** | 4.047798 | 90.728206 | 10.717509 |
| **Calinski-Harabasz Index for Global Model** | 4.047798 | 1.585339 | 10.717509 |

Fig. 3.9: Computational Analysis of Clustering Algorithms

Fig.3.9 represents the values of different assessment metrics for the three clustering algorithms: K-means, DBSCAN, and Hierarchical. This computational analysis clearly shows which clustering algorithm performs better for global and native models.

In summary, comparing the performance of clustering algorithms is crucial for evaluating their effectiveness in organizing data into meaningful groups. The similarity between actual and predicted cluster assignments is measured by the Adjusted Rand Index (ARI), which provides information about the algorithm's accuracy. A higher ARI indicates better agreement between the predicted and actual clusters. A higher silhouette score indicates more significant distinction and clarity of clusters. The silhouette score evaluates the cohesiveness and separation of clusters. To evaluate the algorithm's capacity to form distinct and coherent clusters, it calculates the distance between each point in a given cluster and the points in its neighboring clusters. Lastly, the Davies-Bouldin index quantifies the compactness and separation between clusters, with lower values indicating more optimal clustering. Considering these metrics collectively, one comprehensively understands a clustering algorithm's accuracy, cohesion, and separation performance. This facilitates informed decisions in choosing the most suitable algorithm for identifying Big Five personality traits among the participants.

The clustering analysis results offer insightful information about the complex relationships between the factors in the educational setting. Students are grouped according to shared personality traits, primary identities, and self-esteem; this allows the analysis to reveal patterns and linkages that might not be immediately obvious when looking at individual variables separately. These clusters provide a sophisticated knowledge of how various amalgamations of self-esteem, primary identity, and personality factors might affect academic performance. Educators and researchers can also customize interventions and support systems to match the unique requirements and challenges experienced by various student groups by identifying discrete clusters. This will ultimately promote a more inclusive and productive learning environment.

**4. Conclusion.** This work presents strong evidence supporting the significant effect of personality attributes on the academic achievement of students majoring in computer science within the engineering program. The investigation, employing a chain mediating effects framework, sheds light on the mediating roles of major identity and self-esteem, particularly emphasizing the behavioral efficacy dimension. Notably, self-identity and self-esteem emerge as crucial factors influencing academic success. Based on the Big Five personality attributes, five unique personality groups were identified using K-means, DBSCAN, and Hierarchical clustering analysis.

By offering a deeper awareness of the complex interactions between personality characteristics, psychological variables, and academic performance, the study's findings can enhance the body of literature already in existence and impact psychology and education theory, research, and practice. Longitudinal research would be more appropriate to analyze the progressive impact of self-esteem, major identity, and personality factors on learning results over time. The future scope is to compare the end-of-semester academic results and derive meaningful validation. The results may not be as applicable to students in other academic programmes or institutions or to people with diverse origins in terms of demographics due to their homogeneity.

This work compares clustering algorithms and their performance analysis, proving high accuracy in categorizing students according to their personality traits. Specifically, the models demonstrated that extraversion and conscientiousness are pivotal in positively influencing students' academic achievements. These findings provide valuable insights into the intersection of personality traits and academic performance. They also have implications for interventions and instructional strategies specifically designed to address the special needs of computer science students in the engineering field. Our proposed framework also comprehensively explains a clustering algorithm's accuracy, cohesion, and separation performance, facilitating informed decisions in choosing the most suitable algorithm for a Big Five dataset.

## REFERENCES


[1] E. Achor, V. Adache, E. Ejeh, and B. Kayode, *Perceived impact of personality traits on the academic performance of students in biology*, Innovare Journal of Education, 12 (2024), pp. 38–42.

[2] M. M. Ali, *Modelling the relationship between cyber fraudulent trolling and the dark tetrad personality, considering the big six personality factors among university students*, Kurdish Studies, 12 (2024), pp. 362–381.

[3] N. Alkiş and T. T. Temizel, *The impact of motivation and personality on academic performance in online and blended learning environments*, Journal of Educational Technology & Society, 21 (2018), pp. 35–47.

[4] A. Ashraf, Q. Zhao, W. H. Bangyal, and M. Iqbal, *Analysis of brain imaging data for the detection of early age autism spectrum disorder using transfer learning approaches for internet of things*, IEEE Transactions on Consumer Electronics, (2024), p. 1–13.

[5] É. Audet, S. Levine, P. Dubois, S. Koestner, and R. Koestner, *The unanticipated virtual year: How the big 5 personality traits of openness to experience and conscientiousness impacted engagement in online classes during the covid-19 crisis*, Journal of College Reading and Learning, 53 (2023), pp. 298–315.

[6] N. Babakhani, *The relationship between the big-five model of personality, self-regulated learning strategies and academic performance of islamic azad university students*, Procedia-Social and Behavioral Sciences, 116 (2014), pp. 3542–3547.

[7] C. Cachero, J. R. Rico-Juan, and H. Macià, *Influence of personality and modality on peer assessment evaluation perceptions using machine learning techniques*, Expert Systems with Applications, 213 (2023), p. 119150.

[8] J. Chen, L. Shi, S. Xiao, X. Zheng, Y. Xue, B. Xue, J. Zhang, X. Li, Y. Chen, Y. Wu, and C. Zhang, *The impact of intimate partner violence on depressive symptoms among college students: A moderated mediation model of the big five personality traits and perceived social support*, Journal of Affective Disorders, 350 (2024), p. 203–213.

[9] J. Chi and Y. N. Chi, *Cluster analysis of personality types using respondents' big five personality traits*, International Journal of Data Science, 4 (2023), pp. 116–135.

[10] A. S. Chowdhury, A. H. Mahamud, K. Nur, and H. Z. Haque, *Predicting behavior trends among students based on personality traits*, in Proceedings of the International Conference on Computing Advancements, 2020, pp. 1–5.

[11] D. Darmawan, *The effect of the big five personality on job performance*, Management & Accounting Research Journal, 2 (2017), pp. 36–42.

[12] Y. Ding, Y. Zheng, J. Huang, and T. Zheng, *An online personality traits mining approach based on cluster analysis*, in 2020 International Symposium on Educational Technology (ISET), IEEE, 2020, pp. 258–262.

[13] X. Dong, O. A. Kalugina, D. G. Vasbieva, and A. Rafi, *Emotional intelligence and personality traits based on academic performance*, Frontiers in Psychology, 13 (2022), p. 894570.

[14] M. Fuchs and W. Höpken, *Clustering: Hierarchical, k-means, dbscan*, in Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications, Springer, 2022, pp. 129–149.

[15] S. Gupta, B. Kishan, and P. Gulia, *Comparative analysis of predictive algorithms for performance measurement*, IEEE Access, 12 (2024), p. 33949–33958.

[16] S. Hakimi, E. Hejazi, and M. G. Lavasani, *The relationships between personality traits and students' academic achievement*, Procedia-Social and Behavioral Sciences, 29 (2011), pp. 836–845.

[17] F. G. Hartmann and B. Ertl, *Big five personality trait differences between students from different majors aspiring to the teaching profession*, Current Psychology, 42 (2023), pp. 12070–12086.

[18] V. Hegde and M. Ruthvik, *Prediction of higher education students academic grades based on personality traits*, in 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Oct 2022, pp. 1–6.

[19] N. Herranz-Zarzoso and G. Sabater-Grande, *Monetary versus grade incentives depending on personality traits: A field experiment on undergraduate students' performance*, Heliyon, 9 (2023), p. e15885.

[20] M. Jafarzadegan, F. Safi-Esfahani, and Z. Beheshti, *Combining hierarchical clustering approaches using the pca method*, Expert Systems with Applications, 137 (2019), pp. 1–10.

[21] L. Jesubas, S. Selvaraj, and S. M. Benedict, *Examine how specific personality traits are linked to procrastination tendencies in academics among young adults*, Swetha and Benedict, Sunil Maria, Examine How Specific Personality Traits are Linked to Procrastination Tendencies in Academics Among Young Adults (February 9, 2024), (2024), p. 26.

[22] A. S. Khan, H. Ahmad, M. Z. Asghar, F. K. Saddozai, A. Arif, and H. A. Khalid, *Personality classification from online text using machine learning approach*, International journal of advanced computer science and applications, 11 (2020), pp. 460–476.

[23] P. Lay and A. B. Warsito, *Penerapan algoritma k-means untuk clustering big five personality*, Jurnal Teknik Mesin, Industri, Elektro dan Informatika, 3 (2024), pp. 240–246.
[24] W. Maharani and V. Effendy, *Big five personality prediction based in indonesian tweets using machine learning methods*, International Journal of Electrical and Computer Engineering (IJECE), 12 (2022), p. 1973.
[25] J. Meyer, J. Fleckenstein, J. Retelsdorf, and O. Köller, *The relationship of personality traits and different measures of domain-specific achievement in upper secondary education*, Learning and Individual Differences, 69 (2019), pp. 45–59.
[26] J. Meyer, T. Jansen, N. Hübner, and O. Lüdtke, *Disentangling the association between the big five personality traits and student achievement: Meta-analytic evidence on the role of domain specificity and achievement measures*, Educational Psychology Review, 35 (2023), p. 12.
[27] H. Nassaji, *Personality traits and learner success*, Language Teaching Research, 22 (2018), p. 653–656.
[28] A. V. Nikčević, C. Marino, D. C. Kolubinski, D. Leach, and M. M. Spada, *Modelling the contribution of the big five personality traits, health anxiety, and covid-19 psychological distress to generalised anxiety and depressive symptoms during the covid-19 pandemic*, Journal of affective disorders, 279 (2021), pp. 578–584.
[29] A. J. Oluwadamilare and M. A. Ayanwale, *Partial least square modeling of personality traits and academic achievement in physics*, Asian Journal of Assessment in Teaching and Learning, 11 (2021), pp. 77–92.
[30] Z. Papamitsiou and A. A. Economides, *Exhibiting achievement behavior during computer-based testing: What temporal trace data and personality traits tell us?*, Computers in Human Behavior, 75 (2017), pp. 423–438.
[31] S. Pervaiz, W. H. Bangyal, K. Nisar, and N. U. Rehman, *Population initialization of seagull optimization algorithm with pseudo random numbers for continous optimization*, in 2021 International Conference on Frontiers of Information Technology (FIT), IEEE, 2021, pp. 49–54.
[32] S. Rahmi, A. Anwar, and R. Sovayunanto, *Identify the big five factors of personality for knowing student interpersonal communication in tarakan city junior high school*, International Journal of Educational Narratives, 2 (2024), pp. 129–136.
[33] K. P. Rao, M. C. Sekhara, and B. Ramesh, *Predicting learning behavior of students using classification techniques*, International Journal of Computer Applications, 139 (2016), pp. 15–19.
[34] S. M. Saidi, S. H. Teoh, P. S. A. Singh, and H. Retnawati, *The relationship between teachers' personalities and students' attitudes*, Jurnal Pendidikan Sains dan Matematik Malaysia, 14 (2024), pp. 61–70.
[35] F. S. E. Shaninah and M. H. Mohd Noor, *The impact of big five personality trait in predicting student academic performance*, Journal of Applied Research in Higher Education, 16 (2024), pp. 523–539.
[36] N. Siddiquei and R. Khalid, *The relationship between personality traits, learning styles and academic performance of e-learners*, Open Praxis, 10 (2018), pp. 249–263.
[37] D. Supriyadi, Purwanto, and B. Warsito, *Performance comparison of machine learning algorithms for student personality classification*, in 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Nov 2022, pp. 73–78.
[38] A.-R. Verbree, L. Maas, L. Hornstra, and L. Wijngaards-de Meij, *Personality predicts academic achievement in higher education: Differences by academic field of study?*, Learning and Individual Differences, 92 (2021), p. 102081.
[39] H. Wang, Y. Liu, Z. Wang, and T. Wang, *The influences of the big five personality traits on academic achievements: Chain mediating effect based on major identity and self-efficacy*, Frontiers in Psychology, 14 (2023), p. 1065554.
[40] Y. Wang, *Personality type prediction using decision tree, gbdt, and cat boost*, in 2022 International Conference on Big Data, Information and Computer Network (BDICN), Jan 2022, pp. 552–558.
[41] J. Xu, C. Zhou, and Q. Shao, *Resnet-lstm model based big five personality prediction system*, in Proceedings of the 4th International Conference on Intelligent Science and Technology, 2022, pp. 25–33.
[42] S. Yee, A. Xu, K. Batool, T.-Y. Duan, C. A. Cameron, and K. Lee, *Academic cheating in early childhood: Role of age, gender, personality, and self-efficacy*, Journal of Experimental Child Psychology, 242 (2024), p. 105888.

*Edited by:* Mudasir Mohd
*Special issue on:* Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
*Received:* Jan 31, 2024
*Accepted:* Apr 25, 2024