NIET Journal of Management (NIETJOM)
Volume 14, Issue 2 Winter / Summer 2023 ISSN: 0975-7643 (Print)

Unveiling Personality Profiles: A Data-Driven Approach to Big Five Traits, Sentiment, and Textual Self-Description

Krishna Pratap Rao[1], Ramanand Roy,[2] Rampravesh Kumar, Nishant Kumar, Nisha Kumari

(Noida Institute of Engineering & Technology (MCA Institute), Greater Noida, India)

Abstract: *This study presents a data-driven framework for understanding personality traits, emotional expression, and self-description using Natural Language Processing (NLP) and psychometric analysis. The research integrates the Big Five Inventory (BFI), sentiment analysis, semantic embeddings, clustering, and regression modeling to investigate behavioral and emotional patterns in textual self-expression. Data was collected through structured survey responses containing personality inventory items, subjective happiness indicators, lifestyle satisfaction measures, and open-ended textual responses describing self-identity and workplace rewards. Big Five personality dimensions were computed using standardized psychometric scoring with reverse-coded items. NLP techniques including Sentence Transformers, emotion classification, and zero-shot classification were used to analyze textual responses. K-Means clustering and UMAP dimensionality reduction were applied to identify latent behavioral groups. Regression analysis was conducted to examine the relationship between personality dimensions and happiness outcomes. The findings demonstrate significant associations between personality traits and subjective well-being while highlighting the effectiveness of textual analysis in behavioral profiling. The study contributes to computational psychology and behavioral analytics by combining psychometric frameworks with modern machine learning approaches for personality interpretation.*

Keywords :- Behavioral Analytics, Big Five Personality Traits, Natural Language Processing, Sentiment Analysis, Text Mining

1. Introduction

The increasing availability of behavioral and textual data has created new opportunities for understanding human personality using computational methods. Traditional personality assessment methods rely heavily on psychometric questionnaires, while recent advancements in Natural Language Processing (NLP) allow researchers to infer psychological characteristics from textual expression and semantic behavior.

The Big Five personality model remains one of the most widely accepted frameworks for personality assessment. The five dimensions—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness—provide a comprehensive structure for understanding behavioral tendencies and emotional dispositions. Simultaneously, machine learning and NLP techniques have enabled automated interpretation of human emotions, values, and identity expression through language patterns.

This study attempts to bridge psychometric assessment and computational text analysis by integrating structured personality scores with open-ended textual responses. The research investigates how individuals

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

define themselves, how workplace rewards are interpreted emotionally, and how personality traits influence happiness and satisfaction outcomes.

The primary objectives of the study are:

1. To compute and evaluate Big Five personality traits using psychometric scoring.
2. To analyze emotional patterns within textual self-descriptions.
3. To identify dominant human values from workplace-related textual responses.
4. To cluster behavioral identities using semantic embeddings.
5. To examine the relationship between personality traits and happiness through regression analysis.

The study contributes to behavioral analytics by integrating psychology, machine learning, and semantic analysis into a unified analytical framework.

2. Literature Review

2.1 Big Five Personality Theory

The Big Five personality framework has been extensively used in psychology to measure human behavioral tendencies. Research indicates that personality traits significantly influence emotional stability, professional satisfaction, and social interaction patterns. Extraversion and conscientiousness are often positively associated with happiness and workplace success, while neuroticism is negatively associated with emotional well-being.

2.2 Natural Language Processing in Behavioral Research

Recent developments in NLP have enabled researchers to analyze personality and emotional states through textual data. Semantic embedding models such as Sentence Transformers capture contextual meaning within language, enabling clustering and similarity analysis.

Transformer-based architectures such as BERT and RoBERTa have improved sentiment classification and emotion recognition accuracy. Zero-shot learning models further enable flexible classification without domain-specific training data.

2.3 Computational Personality Analytics

Computational personality analytics combines psychometrics, machine learning, and behavioral modeling to identify psychological patterns in structured and unstructured data. Existing research demonstrates that language usage patterns correlate strongly with emotional disposition, values, and personality traits.

However, limited research integrates Big Five psychometrics with semantic clustering and value classification simultaneously. This study addresses that gap.

3. Research Methodology

3.1 Data Collection

The dataset was collected using a structured survey consisting of:

1. Big Five personality inventory items.
2. Demographic information.
3. Happiness and satisfaction indicators.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

4. Open-ended self-description responses.
5. Open-ended workplace reward descriptions.

Responses with missing values were removed to ensure analytical consistency.

3.2 Data Preprocessing

Column names were standardized and cleaned to remove formatting inconsistencies. Reverse-coded personality items were transformed using reverse scoring techniques to maintain directional consistency in personality computation.

The reverse scoring equation used was:

$$X_{reversed} = 6 - X$$

Personality trait scores were calculated as the mean of relevant inventory items.

3.3 Big Five Personality Computation

The five personality dimensions computed were:

1. Extraversion
2. Agreeableness
3. Conscientiousness
4. Neuroticism
5. Openness

Cronbach's Alpha was used to evaluate internal consistency and reliability of personality measurements.

3.4 Textual Analysis

Open-ended textual responses were analyzed using transformer-based NLP models.

3.4.1 Semantic Embeddings

Sentence embeddings were generated using the "all-mpnet-base-v2" Sentence Transformer model. These embeddings enabled semantic comparison and clustering of identity-related textual responses.

3.4.2 Emotion Detection

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

Emotion classification was performed using a DistilRoBERTa-based transformer model. Joy scores were extracted from self-description responses to measure positive emotional expression.

3.4.3 Value Classification

Zero-shot classification using BART-Large-MNLI was applied to workplace reward responses. The candidate value labels included:

- Achievement
- Security
- Benevolence
- Power
- Self-direction
- Stimulation

The dominant predicted value was assigned to each response.

3.5 Clustering and Dimensionality Reduction

K-Means clustering was used to identify latent behavioral groups within semantic embedding space.

The clustering objective minimized within-cluster variance:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} || x - \mu_i ||^2$$

UMAP dimensionality reduction was applied to visualize behavioral clusters in two-dimensional space.

3.6 Regression Analysis

Multiple linear regression was conducted to examine the relationship between personality traits and happiness outcomes.

The regression model was defined as:

$$Happiness = \beta_0 + \beta_1 E + \beta_2 A + \beta_3 C + \beta_4 N + \beta_5 O + \epsilon$$

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

Where:

- E= Extraversion
- A= Agreeableness
- C= Conscientiousness
- N= Neuroticism
- O= Openness

Additional control variables included age and satisfaction indicators.

4. Results and Analysis

4.1 Personality Trait Distribution

Descriptive statistics indicated moderate variation across all five personality dimensions. Extraversion and openness showed relatively higher mean scores, suggesting active self-expression and exploratory tendencies among respondents.

Table 1: Descriptive statistics of Big Five personality traits

	Mean	Standard Deviation	Minimum	Maximum
extraversion	2.94	0.7	1.12	5
agreeableness	2.55	0.82	1.11	4.56
conscientiousness	2.79	0.76	1	4.67
neuroticism	3.04	0.74	1.25	4.75
openness	2.6	0.65	1.3	4.3

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

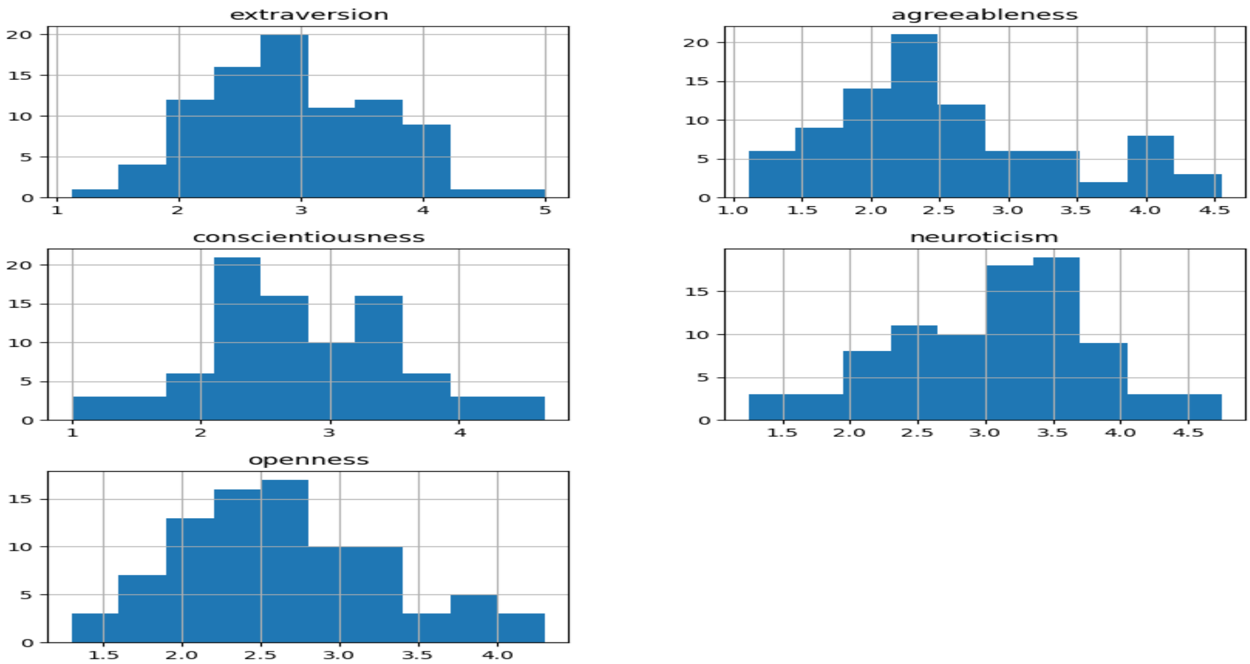

Fig. 2: Distribution of Big Five personality trait scores

4.2 Reliability Analysis

Cronbach's Alpha demonstrated acceptable reliability across personality constructs, indicating consistency in survey responses and psychometric validity.

4.3 Correlation Analysis

Correlation analysis revealed meaningful relationships between personality traits and happiness indicators. Extraversion and conscientiousness showed positive associations with happiness and lifestyle satisfaction, while neuroticism exhibited negative correlations.

Table 2: Correlation matrix between personality traits and happiness indicators

	extraversion	agreeableness	conscientiousness	neuroticism	openness	happy
extraversion	1	0.37	0.47	-0.31	0.45	-0.11
agreeableness	0.37	1	0.61	-0.37	0.68	-0.08
conscientiousness	0.47	0.61	1	-0.51	0.46	-0.33
neuroticism	-0.31	-0.37	-0.51	1	-0.12	0.45
openness	0.45	0.68	0.46	-0.12	1	0.05

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

NIET Journal of Management (NIETJOM)
Volume 14, Issue 2 Winter / Summer 2023 ISSN: 0975-7643 (Print)

happy	-0.11	-0.08	-0.33	0.45	0.05	1

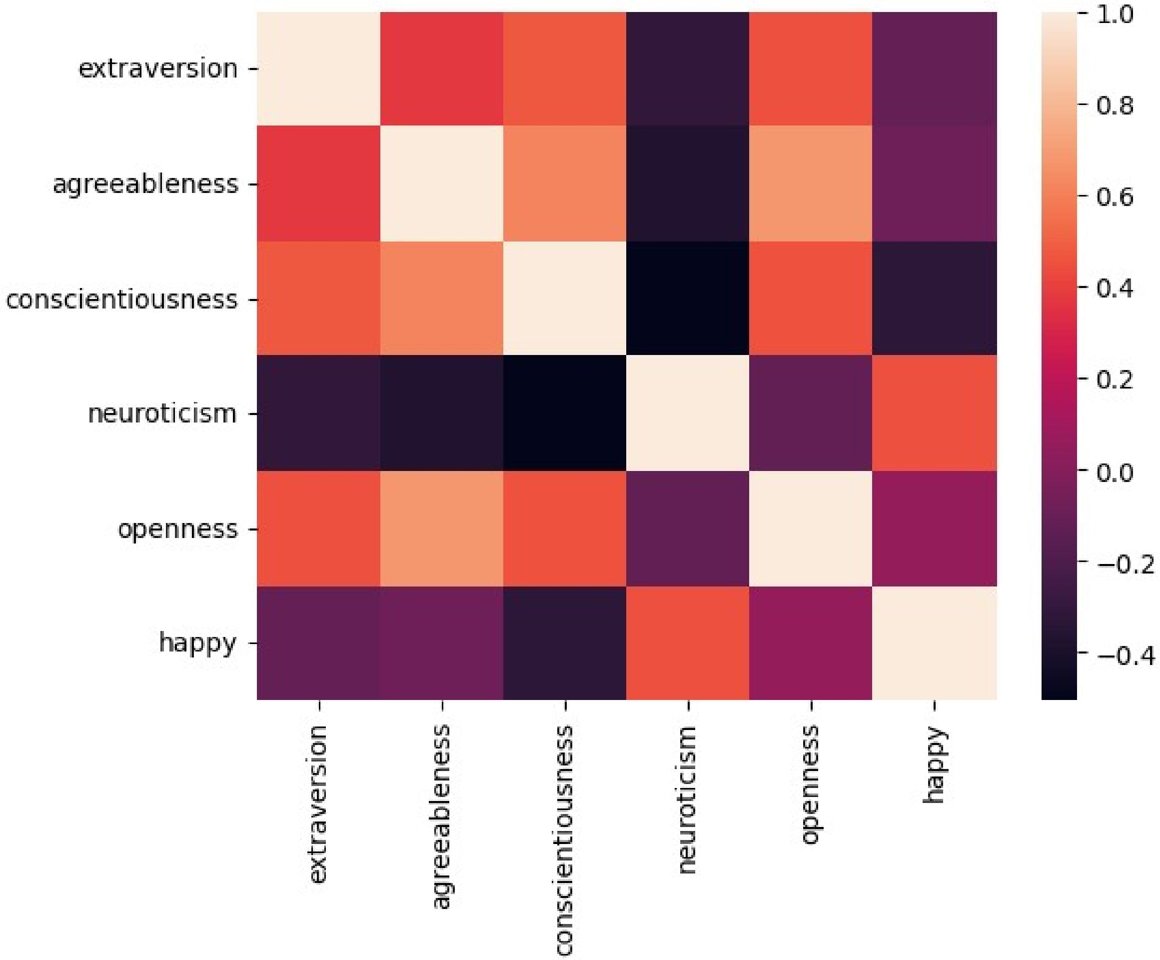

Fig. 3: Correlation heatmap of personality traits and happiness

4.4 Emotion Analysis

Emotion classification results indicated that self-descriptions containing optimistic and socially expressive language received higher joy scores. Individuals with higher extraversion scores frequently demonstrated emotionally positive textual patterns.

Table 5: Distribution of Participants Across Behavioral Clusters Identified Using K-Means Clustering

joy_score
87

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

0.17
0.25
0
0.01
0.05
0.21
0.96

The table presents the number of participants assigned to each semantic behavioral cluster generated through K-Means clustering on textual embedding representations. The clusters represent distinct personality-expression patterns identified from self-descriptive textual responses.

4.5 Value Classification Results

Zero-shot classification identified achievement and self-direction as dominant workplace-related values among respondents. Responses emphasizing autonomy, learning, and recognition were frequently associated with self-direction.

Table 4: Frequency distribution of dominant workplace values

Dominant Value	Frequency
benevolence	41
achievement	18
security	14
power	12
stimulation	1
self-direction	1

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

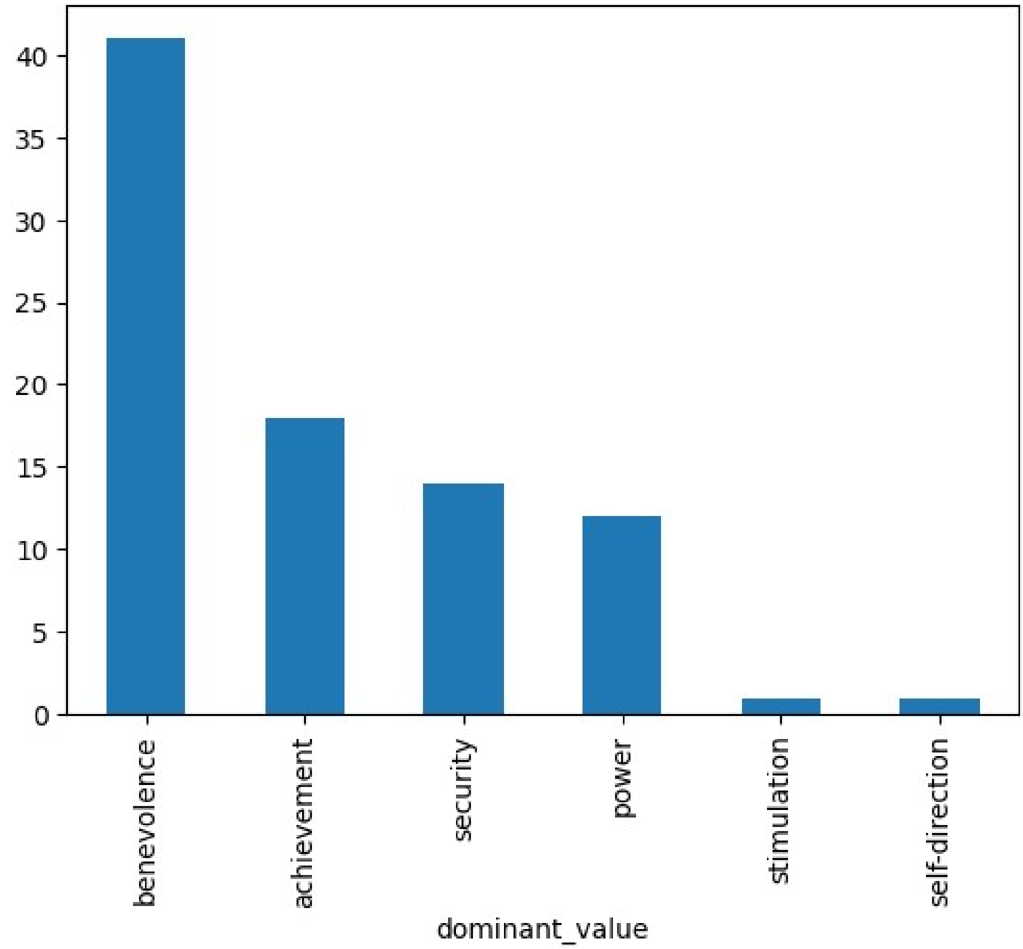

Fig. 4: Distribution of dominant workplace values identified through zero-shot classification

4.6 Behavioral Clustering

K-Means clustering identified distinct behavioral identity groups within semantic space. UMAP visualization demonstrated clear separation between clusters, suggesting meaningful latent behavioral categories.

The clusters represented varying combinations of:

- Social expressiveness
- Emotional positivity
- Career motivation
- Self-reflective tendencies

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

- Personal growth orientation

Table 5: Distribution of Participants Across Behavioral Clusters Identified Using K-Means Clustering

Cluster	Number of Participants
0	12
1	17
2	10
3	20
4	28

The table summarizes the statistical distribution of joy-related emotional scores extracted from participant self-descriptions using transformer-based emotion classification models. Higher joy scores indicate stronger positive emotional expression in textual responses.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

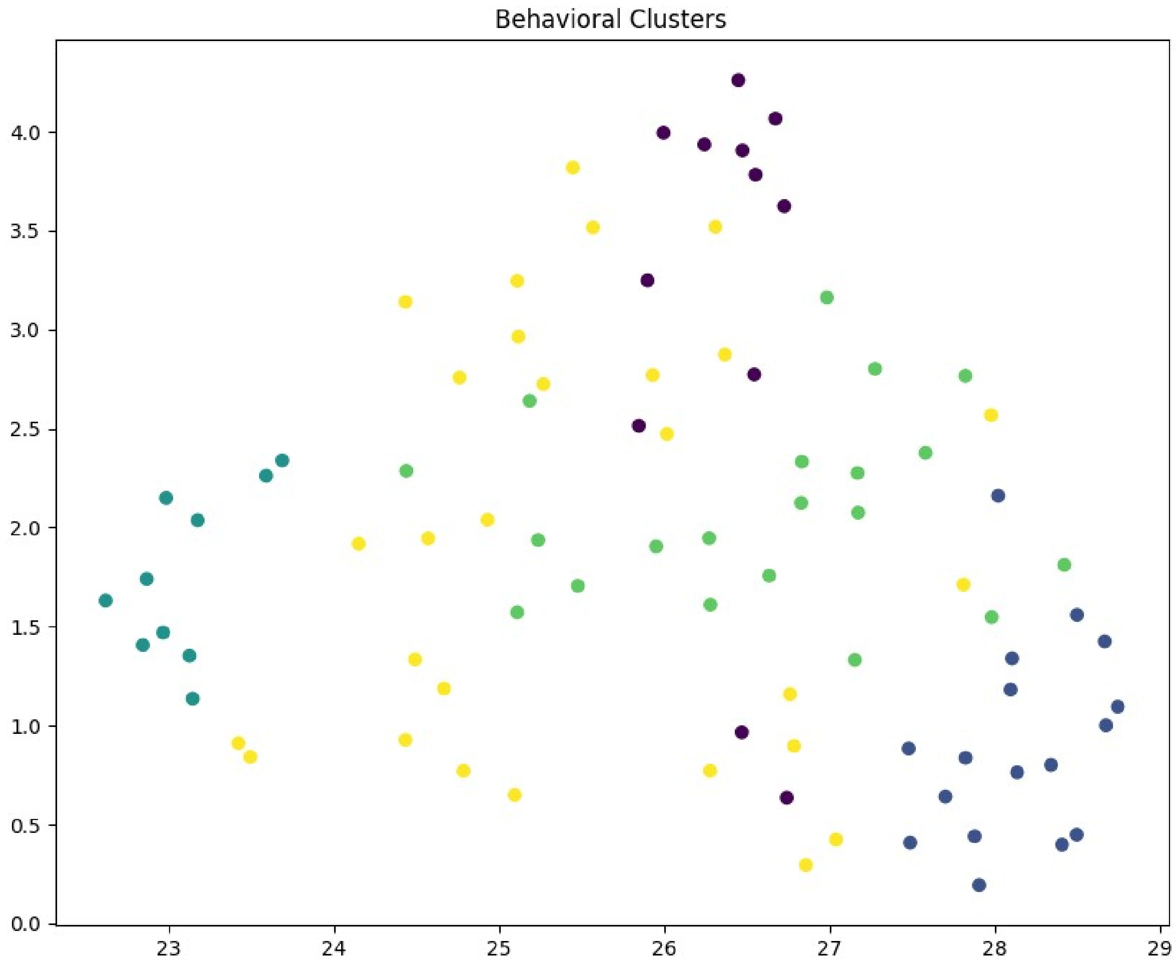

Fig. 1: UMAP visualization of semantic personality clusters generated using K-Means clustering

Each point represents one participant. Spatial proximity indicates semantic similarity. Colors represent behavioral clusters.

4.7 Regression Findings

Regression analysis indicated that personality traits significantly predict happiness outcomes. Extraversion and conscientiousness positively influenced happiness levels, while neuroticism negatively affected emotional well-being.

Table 3: Multiple regression analysis predicting happiness

Variable	Coefficient	Std Error	t-value	p-value
const	2.0173	0.8637	2.3357	0.022
extraversion	0.0586	0.1711	0.3427	0.7327

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

agreeableness	0.2078	0.1924	1.0796	0.2835
conscientiousness	-0.4101	0.1858	-2.2072	0.0301
neuroticism	0.5416	0.1622	3.3396	0.0013
openness	0.1749	0.2249	0.7774	0.4392

The inclusion of control variables improved explanatory power, suggesting that professional satisfaction and lifestyle satisfaction also contribute substantially to happiness.

5. Discussion

The findings demonstrate that computational methods can effectively integrate psychometric and textual analysis for behavioral interpretation. The study validates the use of NLP-based semantic embeddings and transformer models in personality research.

The results support previous psychological research indicating strong relationships between personality and subjective well-being. Additionally, textual responses revealed nuanced emotional and motivational patterns that traditional psychometric measures alone may not fully capture.

The clustering results suggest that semantic identity expression contains latent behavioral structures that can be computationally identified. This has implications for behavioral analytics, organizational psychology, and personalized recommendation systems.

The study also demonstrates the practical value of combining structured survey data with unstructured textual data for richer behavioral profiling.

6. Conclusion

This research presented a data-driven framework for analyzing personality, emotions, and textual self-description using machine learning and NLP techniques. By integrating Big Five psychometric analysis with semantic embeddings, emotion detection, value classification, clustering, and regression modeling, the study successfully identified meaningful behavioral patterns.

The findings confirm that personality traits significantly influence happiness and emotional expression. The use of textual analytics further enhances behavioral understanding by revealing latent semantic and emotional dimensions.

The study contributes to computational psychology and behavioral analytics by demonstrating how modern AI-driven approaches can complement traditional psychometric methods.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

Future research may expand the framework using larger datasets, multilingual analysis, deep learning architectures, and longitudinal behavioral tracking.

7. Acknowledgements

The authors express gratitude to all survey participants whose responses contributed to the research. Appreciation is also extended to the open-source machine learning and NLP communities for providing accessible computational tools and transformer models used in the study

References

[1] L.R. Goldberg, An alternative description of personality: The Big-Five factor structure, *Journal of Personality and Social Psychology*, 59(6), 1990, 1216-1229.

[2] P.T. Costa and R.R. McCrae, The Revised NEO Personality Inventory (NEO-PI-R), *Sage Publications*, 1992.

[3] J. Devlin, M.W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *Proceedings of NAACL-HLT*, 2019, 4171-4186.

[4] N. Reimers and I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, *Proceedings of EMNLP*, 2019, 3982-3992.

[5] T. Wolf, L. Debut, V. Sanh et al., Transformers: State-of-the-art natural language processing, *Proceedings of EMNLP*, 2020, 38-45.

[6] L. McInnes, J. Healy, and J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, *arXiv preprint arXiv:1802.03426*, 2018.

[7] J. Hartmann, J. Heitmann, C. Siebert, and C. Schamp, Emotion English DistilRoBERTa-base, *Hugging Face Transformers Repository*, 2022.

[8] Y. Liu, M. Ott, N. Goyal et al., RoBERTa: A robustly optimized BERT pretraining approach, *arXiv preprint arXiv:1907.11692*, 2019.

[9] T. Pedregosa et al., Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, 12, 2011, 2825-2830.

[10] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, 2009.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.