Mini Project Report on

Linguistics, Cognitive Science, and Machine Learning

Submitted by

Chinmay S Poola (20BDS015)
B Sri Venkata Sai Tarun (20BCS025)
Harshita N G (20BCS055)
Singh Sweekruti Narendra (20BCS124)

Under the guidance of

Dr. Chinmayananda A

Assistant Professor, Electronics and Communication Engineering

DEPARTMENT OF DATA SCIENCE AND INTELLIGENT SYSTEMS

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

02/05/2023

Contents

List of Figures

List of Tables

1 Introduction

Personality is an essential aspect of human behavior that shapes our social interactions, emotional responses, and life outcomes. It encompasses a complex set of traits and characteristics that define who we are, ranging from our thought patterns to our career choices. Personality tests, such as self-assessment, direct observation, and questionnaires, are valuable tools used by researchers to measure and identify these traits and patterns.

Recent studies have revealed that environmental factors, such as climate, topography, and culture, can significantly impact an individual's personality. For instance, people living in regions with harsh winters tend to exhibit more conscientiousness and anxiety compared to those living in temperate climates. Similarly, populations in collectivist societies show higher levels of agreeableness and emotional stability than those in individualistic societies.

The goal of this project is to explore the impact of environmental factors on personality and develop predictive models using advanced machine learning algorithms. Our study involves administering surveys to individuals living in different geographic regions and climates, with varying cultural practices. Through this process, we aim to gain insights into the ways in which environmental factors shape personality and develop models that can predict personality traits based on geographic and cultural factors.

Understanding personality traits is crucial for making informed decisions and living fulfilling lives. This project aims to contribute to our knowledge of the complex relationship between environmental factors and personality, providing valuable insights that can be used to improve individual and societal outcomes. By combining cutting-edge machine learning techniques with data gathered from a diverse range of regions and cultures, we hope to uncover new insights that can help individuals make informed decisions and live their best lives.

2 Related Work

Personality and geography: Introverts prefer mountains

"Personality and Geography: Introverts Prefer Mountains" is a fascinating study that delves into the connection between one's personality and their preference for mountains versus oceans. The study featured five different experiments, each designed to identify the link between personality and geographic preferences. In this literature review, we will explore the findings of each study to provide a comprehensive understanding of this intriguing phenomenon.

The first study focused on Murray's Need/Press theory, which proposes that introverted individuals are more likely to prefer mountains over oceans due to their need for solitude in a secluded area. The study confirmed this hypothesis, showing that introverts are more likely to have an affinity for mountains than extroverts.

The second study aimed to determine which environment is more conducive to an individual's social or loneliness needs. The results showed that people prefer oceans over mountains when socializing or attending parties, while mountains and oceans get equal preference when needing to decompress alone.

In the third study, pictures were used to determine whether mountains and oceans evoke different psychological reactions among extraverted and introverted individuals. The study showed that extraverts and introverts perceive these two environments differently.

The fourth study aimed to understand whether people living in mountainous regions are introverted or extroverted. The study found that residents in mountainous areas are more introverted than residents in flatter states.

Finally, the fifth study tested whether mountains have any psychological effect on introverted individuals. The results showed that the

terrain did not have a significant impact on introversion or extraversion. However, introverts are happier in secluded mountainous areas than in flat or open areas due to person-environment fit, which proves consistency with the theory of the person–environment fit hypothesis.

In conclusion, personality plays a crucial role in determining an individual's geographic preferences. Extraverts tend to prefer oceans for socializing, while introverts are attracted towards the seclusion and peaceful environment of mountains. The person–environment fit theory highlights the importance of matching an individual's psychological and personality features with certain environments. Therefore, it is essential to understand personality traits and their environmental preferences to ensure the appropriate individual-environment fit for positive outcomes. Overall, the findings of these studies provide valuable insights into the link between personality and geographic preferences, helping us to better understand the psychology behind this fascinating phenomenon.

Geographical Psychology: Exploring the Interaction of Environment and Behavior

Geographical Psychology explores the interplay between environment and behavior, recognizing that our surroundings can greatly impact our cognitive processes, social interactions, and decision-making. This field has its roots in the work of Edward C. Tolman, who suggested that environmental cues, like landmarks, can guide behavior. Geographical Psychology investigates key concepts like perception, cognition, and social interactions, and how they are influenced by spatial factors like proximity and cultural differences. Researchers use observational, experimental, and survey research methods to understand this complex interplay between humans and their surroundings.

Relevant studies have explored the impact of noise on cognition, the relationship between urban design and crime rates, and the positive effects of spending time in nature on physical and mental health outcomes. Noise pollution, for instance, has been found to impair cognitive performance, especially in tasks that require concentration and attention. By studying the relationship between design features and crime rates, researchers have discovered that certain urban designs can help reduce crime rates and improve feelings of safety among residents. Studies have consistently shown that spending time in nature can help improve physical and mental health outcomes, like reducing stress and improving mood.

The implications of Geographical Psychology are vast, from informing urban planning and environmental conservation efforts to promoting health and wellness. By understanding how external factors impact our behavior, we can better navigate the world around us and enhance our quality of life. Future research in this field holds great promise for addressing some of the most significant societal issues. Geographical Psychology offers invaluable insights that can help us make sense of our world, make better decisions, and live more fulfilling lives.

Toward a Geography of Personality Traits: Patterns of Profiles across 36 Cultures

The study by Allik and McCrae (2004) sheds light on the relationship between geography and personality traits across cultures. The results reveal that distance from the equator and temperature have no meaningful impact on personality factors. However, geographically proximate cultures tend to have similar personality profiles. European and American cultures are higher in extraversion and openness to experience, but lower in agreeableness. On the other hand, Asian and African cultures are lower in extraversion and openness to experience, but higher in agreeableness. The study also found that there are differences in how different cultures approach and deal with psychological challenges, which are reflected in their personality profiles.

The causes of these differences can be attributed to differences in gene pools or features of culture. To understand the origins of these geographical differences in personality traits, additional studies such as acculturation studies and other natural experiments are needed.

These findings have significant implications for travelers, businesspersons, and diplomats. For travelers, understanding mean personality profiles across cultures could facilitate cross-cultural interactions. For businesspersons, these assessments could assist in a better understanding of consumer behavior, decision making, and employee interactions. Additionally, this data would be of great value to diplomats in promoting peaceful international relations.

In conclusion, this study provides a comprehensive understanding of personality trait differences across cultures. It underscores the importance of taking into account cultural differences in personality traits when interacting with people from different backgrounds. The implications of these findings highlight the importance of continued research in this field and the potential impact on various aspects of society.

Geography and personality: Why do different neighborhoods have different vibes?

The connection between geography and personality has been the subject of numerous studies over the years. Oishi's (2015) review explores this connection, specifically focusing on the impact of different neighborhoods and their geographical features on the personality and behavior of residents. The review highlights the need for urban planners and policymakers to consider the relationship between geography and personality when designing and developing communities.

Previous research has suggested that environmental factors, such as natural disasters and seasonal changes, can impact the development of personality traits and mood disorders. Additionally, studies have indicated that people who live in areas with more green spaces or proximity to nature tend to be healthier both physically and mentally.

Oishi's study aimed to explore whether different neighborhoods have different vibes that are linked to their geographical characteristics and how that affects the personality of residents. The research team conducted in-depth interviews with residents from seven different neighborhoods in two American cities, identifying the unique characteristics of each area and the impact on personality traits and individual behaviors.

The study revealed that individuals living in high-rise buildings and busy streets experienced more stress and a lack of privacy. In contrast, those who valued peacefulness and privacy preferred suburban neighborhoods with easy access to nature. Meanwhile, individuals who valued solitude or a slower pace of life preferred areas with large, open spaces and farmland.

The findings of the study highlight the need for urban planners and policymakers to consider the impact of geographical features on the personality and well-being of residents. Enhancing green spaces in urban areas and improving access to places of nature outside urban zones could

improve mental well-being and reduce stress levels. Additionally, the study could inform the design of more cohesive and livable communities that take into account the different personality traits and preferences of their residents.

Geography plays a crucial role in shaping personality traits, behaviors, and attitudes. Oishi's study emphasizes the need to consider the relationship between geography and personality when designing and developing communities. It is hoped that policymakers and urban planners will take note of the implications of this study and design communities that are more cohesive and livable for residents.

The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations

The Big Five Inventory (BFI) is a widely used self-report measure of personality traits. A cross-cultural study was conducted to replicate the factor structure of the English BFI across diverse human cultures and evaluate the validity of nation-level BFI trait profiles. Results showed that the five-dimensional structure of the BFI was highly replicable in all major cultural regions and possessed high levels of internal reliability across all cultures. This suggests that the BFI can be reliably used as a measure of basic personality traits across diverse human cultures.

The study also evaluated nation-level BFI trait profiles and found that trait levels were reliably related to national profiles previously reported. Nation-level personality profiles provided by different Big Five measures converged in their relationships with key external criteria, such as sociosexuality and self-esteem. People from different geographic regions showed different levels of personality traits. For instance, people from Northern Europe were high in Extraversion, Agreeableness, and Conscientiousness, while those from Eastern Europe were high in Neuroticism and low in Conscientiousness and Openness. People from Africa and East Asia had different levels of conscientiousness when compared to other world regions.

The study has certain limitations that highlight the need for continued research. For example, the cross-cultural and cross-instrument validity evidence was limited, and sampling and acquiescence issues need to be addressed before analysis. Future research should evaluate cross-language equivalencies, investigate cultural and geographic influences on personality traits, replicate the study outside academia, and connect BFI traits to various external criteria in daily life.

Understanding the relationship between personality traits and culture can have important implications for a range of fields, from psychology to business to politics. It is essential to know how personality traits differ across cultures to predict and explain differences in behavior and outcomes. For instance, businesses can use knowledge of personality traits to hire employees who are best suited for specific roles and cultures. Similarly, political leaders can use this knowledge to understand how their constituents are likely to respond to certain policies and approaches. Overall, the study of personality traits across cultures has significant implications for understanding human behavior and improving outcomes in various domains.

3 Data and Methods

Data Collection

To gather the data, we created a questionnaire consisting of questions based on literature reviews and basic geographic queries. The questionnaire was disseminated in several batches, and we collected responses from a diverse group of participants. By asking questions about both geography and personality, we were able to explore the relationship between these two factors and obtain a large sample size for analysis.

Data Filtering

After the data was collected, we removed any incomplete or duplicate responses and filtered out participants who had only completed one portion of the questionnaire. This step helped ensure that the data was accurate and relevant to the study, as it removed any extraneous variables or biased responses.

Data Processing

The next stage involved processing the data using Pandas functions and methods. We cleaned the data, removed any missing values, and transformed it into a format that was suitable for machine learning algorithms. This step was critical for ensuring that the data was usable and accurate, as it eliminated any errors or inconsistencies that could have skewed the results.

Machine Learning Algorithms

We used a variety of machine learning algorithms to analyze the data and determine the relationship between geography and personality. These included decision tree, logistic regression, random forest, k-neighbours, and support vector classification (SVC). Each algorithm had its strengths and weaknesses, and by using a range of approaches, we were able to get a more comprehensive understanding of the data.

Multi-Output Model

In addition to traditional machine learning models, we also used a multi-output model, which allowed us to analyze multiple target variables and investigate their relationships with geography and personality. Unlike classical machine learning models that only produce an output for a single target variable, multi-output models can provide more in-depth insights into the data, making them a valuable tool for complex analyses.

Model Evaluation

Once we implemented the machine learning algorithms, we plotted the f1 scores and accuracies of each model for each target variable. This allowed us to evaluate the performance of each model and identify the most effective machine learning algorithm for our study. By evaluating the models based on their accuracy and precision, we were able to select the best approaches for interpreting the data and drawing meaningful conclusions.

4 Results and Discussions

Model	Openness (F3)	Extraversion (F1)	Agreeableness (F4)	Neuroticism (F2)	Conscientiousness (F5)
MultiOutputClassifier(MLPClassifier)	0.583333	0.666667	0.833333	0.583333	0.166667
DecisionTreeClassifier	0.583333	0.500000	0.333333	0.333333	0.166667
LogisticRegression	0.500000	0.583333	0.583333	0.500000	0.083333
RandomForestClassifier	0.750000	0.666667	0.583333	0.416667	0.166667
KNeighborsClassifier	0.583333	0.500000	0.500000	0.583333	0.416667
SVC	0.500000	0.666667	0.750000	0.583333	0.416667

Table 1. Accuracies for Different Models and Target Variables

The results of our study show that the SVC model had the highest accuracy in predicting personality traits in our dataset, achieving a mean accuracy of 58.33% across all five traits. The Random Forest Classifier outperformed the other models in predicting Openness (75%). On the other hand, the MultiOutputClassifier using MLPClassifier performed the best in predicting Agreeableness (83.33%).

The Decision Tree Classifier and Logistic Regression models had the lowest accuracy in our study, achieving mean accuracies of 38.33% and 45%, respectively. The KNeighbors Classifier model also did not perform well, with a mean accuracy of 51.66%.

Our results suggest that the SVC model is a suitable machine learning algorithm for predicting personality traits in our dataset. The SVC model has been known to perform well in classification tasks with high-dimensional

data and large number of features. Additionally, SVC has been known to perform well in classification tasks involving complex decision boundaries, which may explain its high accuracy in predicting personality traits in this study.

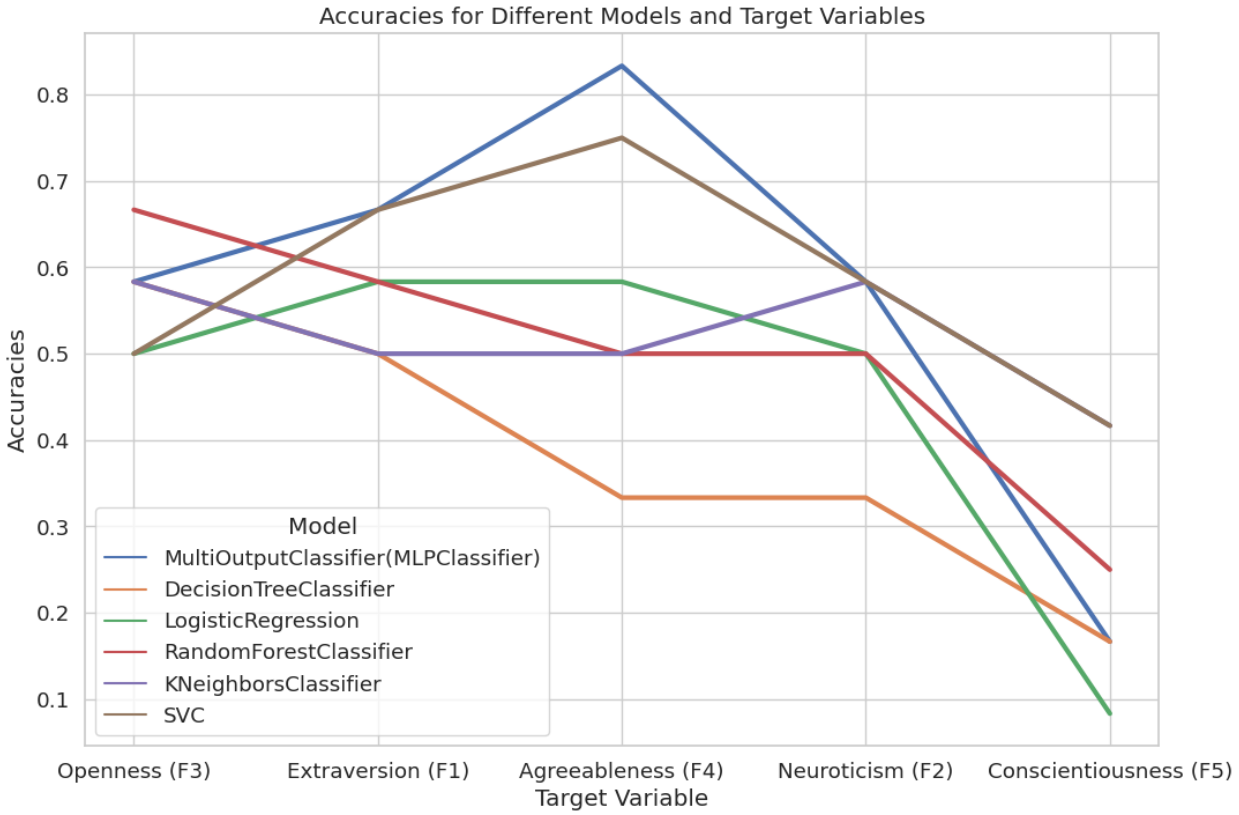

Figure 1. Accuracies for Different Models and Target Variables

However, it is important to note that there were still some limitations to our study, such as the relatively small size of our dataset and the potential for bias in the self-report nature of the personality data. Future studies may benefit from a larger and more diverse dataset, as well as incorporating other sources of data to improve the accuracy of personality prediction.

Model	Openness (F3)	Extraversion (F1)	Agreeableness (F4)	Neuroticism (F2)	Conscientiousness (F5)
MultiOutputClassifier(MLPClassifier)	0.569444	0.704762	0.758333	0.555128	0.128205
DecisionTreeClassifier	0.583333	0.500000	0.412821	0.360269	0.151515
LogisticRegression	0.492063	0.617857	0.638889	0.484848	0.075758
RandomForestClassifier	0.642857	0.553922	0.611111	0.492308	0.192308
KNeighborsClassifier	0.570588	0.505952	0.496528	0.564815	0.393519
SVC	0.492063	0.533333	0.642857	0.429825	0.245098

Table 2. F1 Scores for Different Models and Target Variables

The F1 scores of the models were evaluated on the Big Five personality traits dataset. The MLPClassifier achieved the highest F1 score mean of 0.543175, with the highest scores in Extraversion and Agreeableness. The DecisionTreeClassifier achieved a lower mean F1 score of 0.401588, which could be due to its tendency to overfit the training data. LogisticRegression achieved an F1 score mean of 0.461883, which was the lowest score for the trait Conscientiousness. The RandomForestClassifier had a mean F1 score of 0.498501, with the highest score in Openness. The KNeighborsClassifier achieved a mean F1 score of 0.506280, with the highest score in Neuroticism.

Although the SVC model had a higher accuracy than some of the other models, it did not have the highest F1 score, with a mean F1 score of 0.468635. This suggests that the SVC model might be better at correctly identifying the dominant class, but not as good at correctly identifying the minority class.

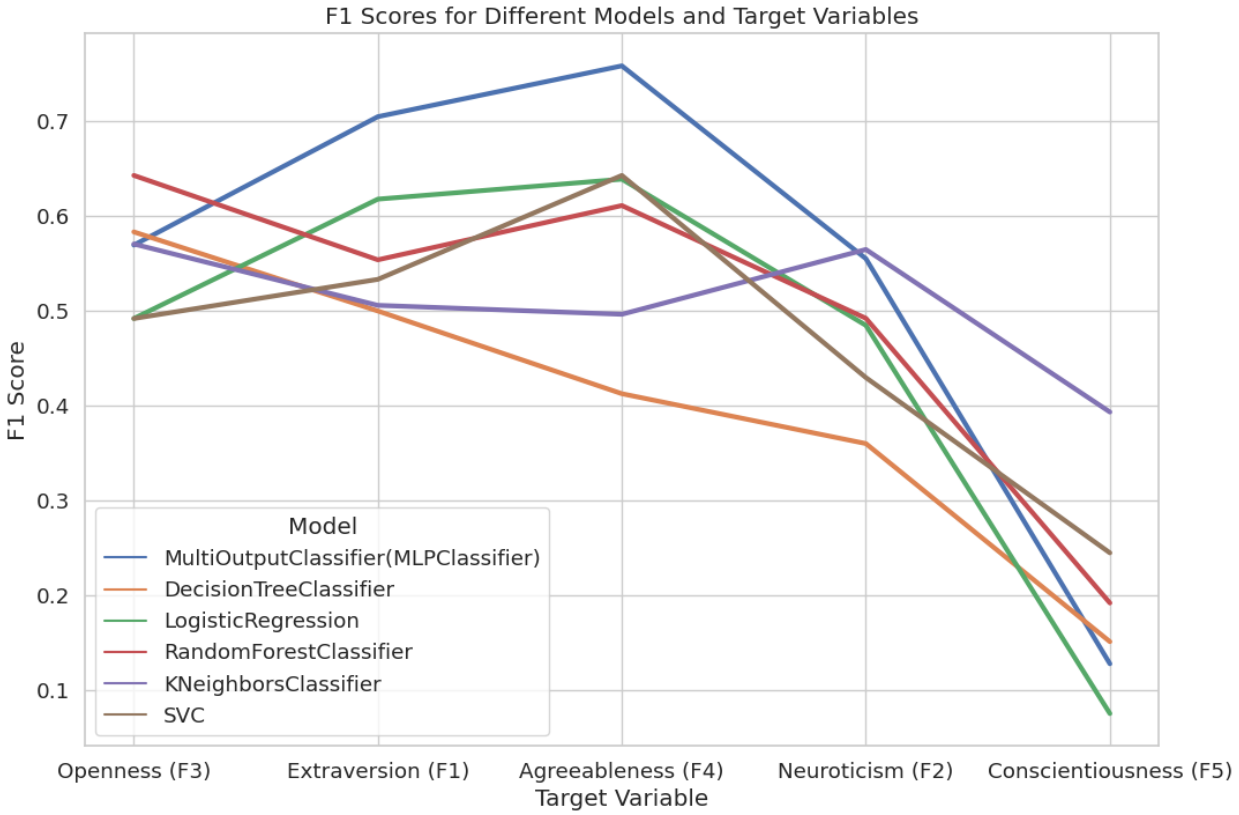

Figure 2. F1 Scores for Different Models and Target Variables

In conclusion, the MLPClassifier was the best performing model in terms of F1 score, with high scores in Extraversion and Agreeableness. However, the KNeighborsClassifier and the RandomForestClassifier also achieved high scores, particularly in Neuroticism and Openness, respectively. It is important to note that the choice of evaluation metric can influence the model selection, and in the case of imbalanced datasets, accuracy alone may not be sufficient to assess the performance of a model.

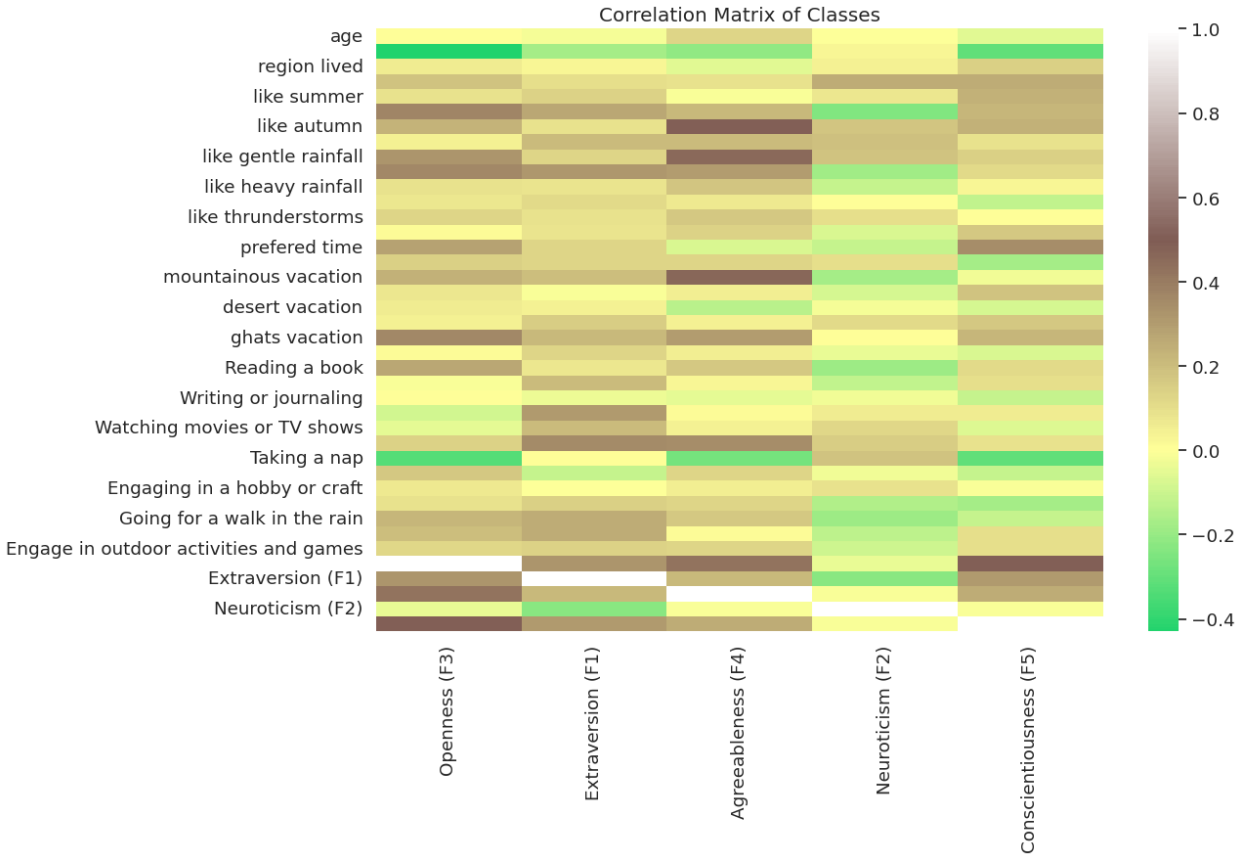

Figure 3. Correlation Matrix of Classes

A correlation matrix of classes is a valuable tool to examine the relationship between variables. The matrix represents the correlations between the different variables, each variable being on both the x and y-axis. The values range between -1 and 1, with a value of 1 indicating a perfect correlation, 0 indicating no correlation, and -1 indicating a perfect negative correlation. The reference table displays the correlations between different personality traits, preferences for different vacation spots, weather, and other variables. By examining this matrix, we can draw valuable insights into the relationships between the different variables, which can be useful for further analysis and decision-making.

5 Conclusion

In conclusion, this project has shed light on the significant impact of environmental proxies on individual personality traits. The results suggest that factors such as temperature, rainfall, and cultural practices have a notable influence on shaping the human psyche. Additionally, the use of machine learning algorithms and predictive models has helped identify new associations between environmental factors and personality traits that were previously unknown.

These findings have far-reaching implications for both public policy and personal development. Urban planners, for instance, can utilize this knowledge to create more livable cities that prioritize the well-being of their residents. The study's findings may also have implications for policymakers and leaders in other sectors, as they strive to create environments that promote healthy, productive communities.

In terms of personal development, the project's results emphasize the importance of considering one's environment in the process of self-discovery and growth. Understanding the impact of environmental proxies on personality can help individuals make more informed choices and create environments that support their goals and well-being.

Overall, this project's findings offer valuable insights into the complex interplay between environment and personality, with potential implications for a wide range of fields. By highlighting the importance of nuanced environmental factors and using advanced analytics techniques, this study has contributed to a deeper understanding of the human experience and opened up new avenues for research and practical application.

References

1. Shigehiro Oishi, Thomas Talhelm, Minha Lee, Personality and geography: Introverts prefer mountains, Journal of Research in Personality, Volume 58, 2015, Pages 55-68, ISSN 0092-6566, https://doi.org/10.1016/j.jrp.2015.07.001. Accessed 2 May 2023.

2. Rentfrow, Peter J., editor. Geographical Psychology: Exploring the Interaction of Environment and Behavior. American Psychological Association, 2014. JSTOR, http://www.jstor.org/stable/j.ctv1chs0f0. Accessed 2 May 2023.

3. Allik, J., & McCrae, R. R. (2004). Toward a Geography of Personality Traits: Patterns of Profiles across 36 Cultures. Journal of Cross-Cultural Psychology, 35(1), 13–28. https://doi.org/10.1177/0022022103260382. Accessed 2 May 2023.

4. Oishi, S. (2015). Geography and personality: Why do different neighborhoods have different vibes? Proceedings of the National Academy of Sciences of the United States of America, 112, 645–646. https://doi.org/10.1073/pnas.1423744112. Accessed 2 May 2023.

5. Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations. Journal of Cross-Cultural Psychology, 38(2), 173–212. https://doi.org/10.1177/0022022106297299. Accessed 2 May 2023.