Progress Project 3

CSIS 4495-050 Applied Research Project

300372531 Hyunhee Kim

1. Work Logs

Date	Number of Hours	Description of work done
2025-10- 26	2	Research personality test types and reference
2025-10-30	2	Planning personality modeling frame
2025-10-31	3	Search personality dataset and start to cleaning dataset
2025-11-01	2	Start to preprocessing personality (Big Five personality) dataset and working
2025-11-02	4	Implement KMeans clustering to group personality trait at kim branch, Misc/ Data Modeling file name “personality_K_means.ipynb”
2025-11-03	2	Issues identified with implementing KMeans clustering approach -> finding appropriate K and too long executive time.
2025-11-04	4	Restart personality modeling and fix earlier issues, finished KMeans clustering part push at Kim branch, Misc /Data Modeling/“Personality_Kmeans2.ipynb”
2025-11-05	3	Implementing Random Forest model based on Kmeans clustering result, push at Kim branch, Misc /Data Modeling/ “Personality_rdforest.ipynb”
2025-11-06	2	Applying the test code, it was used to verify that it works correctly. Pysh at Kim branch, Misc/ Data Modeling / “Personality_final.ipynb”

2. Description of Work Done

During this period, I focused on developing and refining the personality modeling process. I began by researching different personality test types and references, which helped me design the overall framework for the personality modeling approach. After that, I searched for suitable personality datasets and performed data cleaning to ensure consistent data quality.

Once the data was prepared, I started preprocessing the Big Five personality dataset to structure it for clustering. On November 2, I implemented the K-Means clustering algorithm to group personality traits and pushed the initial version (personality_K_means.ipynb) to the Kim branch. The following day, I

encountered issues related to selecting an appropriate value of K and long execution times.

To address these issues, I restarted the personality modeling process on November 4 and completed an improved version of the K-Means clustering, which was finalized and pushed (Personality_Kmeans2.ipynb) after working overnight from November 4 to early November 5. Then, I implemented a Random Forest model (Personality_rdforest.ipynb) based on the K-Means clustering results, working from the night of November 5 through after midnight on November 6 before pushing the final version.

Finally, on November 6, I applied the test code to verify that the integrated model worked correctly, confirming the successful operation of the personality prediction pipeline (Personality_final.ipynb).

3. **Repo Check-ins (Completed)**

Misc/Data Modeling/personality_K_means.ipynb: Implemented K-Means clustering to group personality traits and explored initial clustering results. (check-ins were pushed to the Kim branch.)

Misc/Data Modeling/Personality_Kmeans2.ipynb: Resolved issues from the initial clustering process by determining an appropriate value of K and optimizing execution time.
Worked overnight on November 4–5, and the final check-in was completed on November 5. (check-ins were pushed to the Kim branch.)

Misc/Data Modeling/Personality_rdforest.ipynb: Implemented a Random Forest model based on the K-Means clustering output to classify personality types more accurately.
Worked from the night of November 5 through after midnight on November 6, with the final check-in completed on November 6. (check-ins were pushed to the Kim branch.)

Misc/Data Modeling/Personality_final.ipynb: Applied test code to verify model functionality and confirmed correct operation of the final integrated pipeline. (check-ins were pushed to the Kim branch.)