Project Journal

Mohammed Sameer Ahmed
X23337257
Team E
MSc AI

Code Artefact

1. Created **.env** file for securing the MongoDB credentials. Sometimes the code editor has to be restarted for environment variables to be updated. It took about **an hour**.
2. Merged code of other teammates in one file. Errors occurred after merging which was solved by mutual discussion and online resources. It took **1 day**.
3. Rewrote the code in a **modular** and **readable** way. Initially the code had poor structure, no comments and improper variable names which was resolved by having a meeting. It took **2 days** for this task.
4. Made the **GitHub repository** for project and pushed the code. Sharing the files was not easy before the GitHub repo. It took about **an hour**.

Related to Alzheimer's dataset:

1. Imported the dataset (`import_json`) and stored the semi structured data before processing in MongoDB (`store_data_in_mongodb`). The dataset was too large and unstructured, so it was rejected by GitHub and MongoDB. Removed some unrelated rows before and tweaked the structure a bit (`remove_non_smoking_data`, `get_columns_and_data`) to push it to GitHub and upload to MongoDB. It took **1 day**.
2. Created pandas dataframe (`create_alzheimers_dataframe`) and performed data preprocessing (`clean_and_convert_alzheimers_data`, `convert_tobacco_data_for_alzheimers_data`) and merged the required fields of tobacco dataset into azheimer's dataset `alzheimers_merged_df` and uploaded the structured dataset to database. Joins on dataframes were not clear so learnt it from online resources. It took **1 day**.
3. Split the `alzheimers_merged_df` into two `merged_df_gender` and `merged_df_race` based on race and gender. Plotted lineplot (`show_line_plot`) and geoplot (`show_geo_plot`). Concept of double line plotting and geoplot was not clear so took help from online resources. It took **half day**.
4. Created the machine learning model `machine_learning_model`. Used online resources to find out which model is suitable for best results. It took **half day**.

Report

1. Gave written matter about my work, to be included in Abstract, Introduction part, Methodology part and Conclusion part. It took **1 day**.
2. Read the research paper for Related Work part and contributed critical written matter about ‘Construction of a risk prediction model for Alzheimer’s disease in the elderly population’. It took **1 day**.
3. Generated IEEE style reference for the literary work, model, technologies used and the dataset for References part. It took **1 Hour**.
4. Assited in the fine tuning, formatting. Rephrasing the sentences to reduce number of words was challenging. It took **2 hours**.