Alzheimer’s Disease Prediction

TESSA CLARY

Description

- This will be an analysis of the “Alzheimer’s Disease and Healthy Aging Data” dataset from Data.gov. This dataset is a .csv first published by the CDC in July 2023.
- Alzheimer’s is a disease that we are still learning to predict successfully in its earlier stages. Early detection and diagnosis of the disease can help us understand how to prevent and treat it within the population so there is a lot of potential value in analyzing this dataset. Studying various health factors and demographics of those likely to be impacted for correlations and model anomalies can help identify those who are most at risk.
- https://catalog.data.gov/dataset/alzheimers-disease-and-healthy-aging-data

Questions

- Demographic Analysis
  - Age
  - Race
  - Gender
  - Location

- Correlation among Health Factors
  - Arthritis
  - Mental Health
  - Sleep Prevalence
  - Depression
  - Smoking / Drinking
  - Etc.

Prior Work

- “Population measures of subjective cognitive decline: A means of advancing public health policy to address cognitive health”:

- https://alz-journals.onlinelibrary.wiley.com/doi/full/10.1002/trc2.12142

- “A smart Alzheimer’s patient monitoring system with IoT-assisted technology through enhanced deep learning approach”:

- https://link.springer.com/article/10.1007/s10115-023-01890-x

Dataset

- “**Alzheimer's Disease and Healthy Aging Data**”

- https://catalog.data.gov/dataset/alzheimers-disease-and-healthy-aging-data

- Has been downloaded onto local machine

Proposed Work

- Data Cleaning
  - Columns are not consistently filled out
  - Check for missing or incorrect values
  - Check for outliers

- Data Preprocessing
  - Feature Engineering
  - Normalization
  - Data Visualization- check for anomalies or patterns

- Data Integration
  - Data Transformation
  - Potentially integrate additional external datasets

Tools

- Python
- Numpy
- Pandas
- SQL
- Matplotlib
- Github

Evaluation

- Data Visualization will help identify anomalies and other interesting trends in the dataset.
- A regression analysis can be performed- high R values would infer a strong relationship between the given variables.
  - More than one may be performed on a subset of data based on the types of relationships being modeled.
- Validation can be done towards the end to ensure that the results I have created matches reasonably well with other publicly available work.