Digital Manifest
Sex-Specific and Regional Analysis of Heart Disease Prediction Using Machine Learning Algorithms: Insights from the UCI Irvine Public Heart Disease Datasets (Cleveland and Long Beach)

Jonathan Asanjarani
City University of New York Graduate Center
DATA 79000: Capstone Project and Thesis
Advisor: Johanna Devaney
November 25th, 2024

Project Components

1. Capstone Report (Print and Digital)

- **File Name:** Project_Write_up12.30.24.docx.pdf
- **File Type:** PDF
- **Description:** Full written report detailing research objectives, methodology, results, and discussions.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/Project_Write_up12.30.24.docx.pdf

2. Exploratory Data Analysis (EDA) Notebook

- **File Name:** EDA_4_binary_classification.ipynb
- **File Type:** Google Collab Notebook (.ipynb)
- **Description:** Python notebook detailing data cleaning, univariate, bivariate, and multivariate analyses, including visualization and statistical tests.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/EDA_4_binary_classification.ipynb

3. Machine Learning Model Implementation for Cleveland

- **File Name:** ML_Algo_4_binary_classification.ipynb
- **File Type:** Google Collab Notebook (.ipynb)
- **Description:** Google Collab Notebook containing code for implementing and evaluating machine learning models (Random Forest, XGBoost, and ensemble methods) using the Cleveland dataset.

- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/ML_Algo_4_binary_classification.ipynb

4. Machine Learning Model Implementation for VA Long Beach

- **File Name:** ML_Algo_4_bin_classification_va_longbeach.ipynb
- **File Type:** Google Collab Notebook (.ipynb)
- **Description:** Google Collab Notebook containing code for implementing and evaluating machine learning models (Random Forest, XGBoost, and ensemble methods) using the VA Long Beach dataset.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/ML_Algo_4_bin_classification_va_longbeach.ipynb

5. Cleveland Processed Dataset

- **File Name:** processed.cleveland.data
- **File Type:** ZIP archive (contains .data files)
- **Description:** Includes cleaned and transformed versions of the Cleveland dataset used in the study.
- **URL:** https://github.com/Jdasanja/masters_thesis/blob/main/processed.cleveland.data

6. VA Long Beach Processed Datasets

- **File Name:** processed.va.data
- **File Type:** ZIP archive (contains .data files)
- **Description:** Includes cleaned and transformed versions of the VA Long Beach dataset used in the study.
- **URL:** https://github.com/Jdasanja/masters_thesis/blob/main/processed.va.data

7. Data Transformation Script Cleveland

- **File Name:** ML_Algo_4_binary_classification.ipynb
- **File Type:** Google Collab Notebook (.ipynb)
- **Description:** Custom Python scripts for data preprocessing and feature engineering, including transformations applied to Cleveland dataset.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/ML_Algo_4_binary_classification.ipynb

8. Data Transformation Script VA Long beach

- **File Name:** ML_Algo_4_bin_classification_va_longbeach.ipynb
- **File Type:** Google Collab Notebook (.ipynb)
- **Description:** Custom Python scripts for data preprocessing and feature engineering, including transformations applied to VA Long Beach dataset.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/ML_Algo_4_bin_classification_va_longbeach.ipynb

9. ASCVD Risk Score Implementation Cleveland

- **File Name:** ACSVD_calculation_of_Cleveland.ipynb
- **File Type:** Jupyter Notebook (.ipynb)
- **Description:** Python notebook implementing the ASCVD Risk Calculator for the Cleveland dataset.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/ACSVD_calculation_of_Cleveland.ipynb

10. ASCVD Risk Score Implementation VA Long Beach

- **File Name:** ACSDV_Calculation_4_va_longbeach.ipynb
- **File Type:** Jupyter Notebook (.ipynb)
- **Description:** Python notebook implementing the ASCVD Risk Calculator for the Cleveland dataset.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/ACSDV_Calculation_4_va_longbeach.ipynb

11. A Note on Technical Specifications

- **File Name:** A Note on Technical Specifications.pdf
- **File Type:** PDF
- **Description:** PDF that provides an overview of the project's development environment, data sources, processing methods, file formats, version control, and external tools used to ensure reproducibility and transparency.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/A%20Note%20on%20Technical%20Specifications.pdf

12. Data Dictionary

- **File Name:** Data Dictionary.pdf
- **File Type:** PDF
- **Description:** PDF that outlines key variables, transformations, critical functions, and classifiers used in the project, providing detailed descriptions to ensure clarity and reproducibility.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/Data%20Dictionary.pdf

13. Digital References

- **File Name:** Digital References.pdf
- **File Type:** PDF
- **Description:** PDF that provides detailed citations for all software, tools, datasets, and external resources used in the project, ensuring transparency and enabling reproducibility.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/Digital%20References.pdf

14. Data Management Plan

- **File Name:** Data Management Plan Overview.pdf
- **File Type:** PDF
- **Description:** Comprehensive plan outlining data handling, storage, and ethical considerations.
- **URL:** https://github.com/Jdasanja/masters_thesis_final/blob/main/Data%20Management%20Plan%20Overview.pdf