A Note on Technical Specifications

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

Development Environment:

- Platform: Google Colab, a cloud-based Python environment with access to GPUs for accelerated computation.
- Programming Language: Python (version 3.8).
- Libraries and Frameworks: Scikit-learn, XGBoost, Pandas, NumPy, Matplotlib, Seaborn, and others detailed in the References section.

Data Storage and Processing:

- Dataset Sources: Cleveland and VA Long Beach datasets from the UCI Machine Learning Repository.
- Data Cleaning and Preprocessing: Conducted within Google Collab Notebooks using Python-based libraries.
- File Formats: CSV files for datasets; Python files (.py) and Jupyter Notebooks (.ipynb) for code.

Version Control and Repository:

- The project's source code, processed datasets, and supplementary materials are hosted on a GitHub repository:
    - Repository URL: https://github.com/Jdasanja/masters_thesis_final
    - Latest Version: Updated regularly with detailed commit history for reproducibility.

External Tools and Models:

- ASCVD Risk Calculator implemented via an open-source Python package from GitHub:
    - Repository URL: https://github.com/brandones/ascvd/tree/master.