

Digital References

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

Software and Tools Used

1. Google Colab

- Description: Cloud-based Python environment with GPU access for accelerated computation.
- URL: <https://colab.research.google.com>
- Accessed: November 2024

2. Python

- Version: 3.8
- Description: High-level programming language used for data analysis, modeling, and visualization.
- URL: <https://www.python.org>
- Accessed: November 2024

3. Scikit-learn

- Version: 1.2.0
- Description: Library for machine learning algorithms, preprocessing, and evaluation.
- URL: <https://scikit-learn.org/stable/>
- Accessed: November 2024

4. XGBoost

- Version: 1.6.0
- Description: Gradient boosting library optimized for supervised learning tasks.
- URL: <https://xgboost.ai>

- Accessed: November 2024

5. Pandas

- Version: 1.4.3
- Description: Data manipulation and analysis library for structured data.
- URL: <https://pandas.pydata.org>
- Accessed: November 2024

6. NumPy

- Version: 1.23.0
- Description: Library for numerical computations and array processing.
- URL: <https://numpy.org>
- Accessed: November 2024

7. Matplotlib

- Version: 3.6.0
- Description: Visualization library for static and interactive graphics.
- URL: <https://matplotlib.org>
- Accessed: November 2024

8. Seaborn

- Version: 0.12.2
- Description: Statistical data visualization library built on Matplotlib.
- URL: <https://seaborn.pydata.org>
- Accessed: November 2024

9. ASCVD Risk Calculator

- Version: GitHub Repository
- Description: Python implementation of the ASCVD Risk Calculator for cardiovascular risk prediction.
- URL: <https://github.com/brandones/ascvd/tree/master>
- Accessed: November 2024

Datasets

1. Cleveland Heart Disease Dataset

- Source: UCI Machine Learning Repository
- Description: Dataset used for binary classification of heart disease presence.
- URL: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- Accessed: November 2024

2. VA Long Beach Heart Disease Dataset

- Source: UCI Machine Learning Repository
- Description: Dataset used for regional generalization of machine learning models.
- URL: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- Accessed: November 2024

Guidelines and Methodological References

1. Mueller, Andreas C., & Guido, Sarah

- Title: *Introduction to Machine Learning with Python*
- Publisher: O'Reilly Media
- Publication Date: 2016
- URL: [https://github.com/dlsucomet/MLResources/blob/master/books/\[ML\]%20Introduction%20to%20Machine%20Learning%20with%20Python%20\(2017\).pdf](https://github.com/dlsucomet/MLResources/blob/master/books/[ML]%20Introduction%20to%20Machine%20Learning%20with%20Python%20(2017).pdf)

2. Software Sustainability Institute

- Title: *How to Cite and Describe Software*
- URL: <https://www.software.ac.uk/how-cite-and-describe-software>
- Accessed: November 2024

Additional Resources for Citing Software and Data

1. Digital Curation Centre

- Title: *How to Cite Datasets and Link to Publications*
- Authors: Ball, A., & Duke, M.
- Publisher: Digital Curation Centre
- Publication Date: 2011
- URL: <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>
- Accessed: November 2024

2. DataCite

- Title: *Why Cite Data?*
- URL: <https://www.datacite.org/>
- Accessed: November 2024