National College of Ireland

Configuration Manual

MSc Research Project
MSc in Data Analytics

Sarang Sanjay Khandare
Student ID: 23173360

School of Computing
National College of Ireland

Supervisor: Prof. Harshani Nagahamulla

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sarang Sanjay Khandare
Student ID:	23173360
Programme:	MSc in Data Analytics **Year:** 2024
Module:	MSc Research Project
Lecturer:	Prof. Harshani Nagahamulla
Submission Due Date:	12/12/2024
Project Title:	Hypertension Risk Prediction using Machine Learning Models and ensemble techniques
Word Count:	830 **Page Count: 4**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sarang Sanjay Khandare

Date: 10/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Sarang Sanjay Khandare
Student ID: 23173360

1 Introduction

This document will provide instructions on how to best replicate the work undertaken as part of the associated research project. This research project was built, taking into consideration that is replicable. Thus, the setup and pre-requisites are relatively straightforward and easy to follow. This document will provide details about the necessary hardware and software requirements, as well as the file directory set up to run the code used in the research project.

2 Hardware requirements

The data used in this research project was moderately sized, yet it is recommended that the research be carried out on a machine with sufficient CPU and processing power to handle some computationally intensive models. Figure 1 describes the hardware details of the device the research was carried out on.

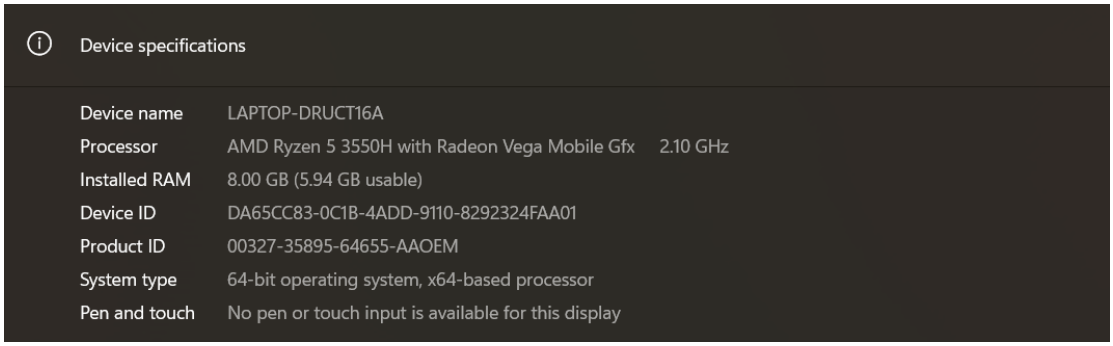

Figure1: Hardware specifications

3 Software requirements

The software requirements section highlights the libraries and software versions used in this research project. This will ensure the replicability and consistency of the results. Central to this project, the use of *Jupyter notebook IDE* was done. *Jupyter Notebook IDE* can be accessed through *Anaconda Navigator* or *Anaconda prompt.*

3.1 Software Environment:

The implementation of the project was done on these software requirements:

- Operating System: *Windows 11 (Version 23H2)*
- Environment Management: *Anaconda Navigator (Version 23.3.1)*
- Integrated Development Environment: *JupyterLab (Version 4.1.6)* and *Jupyter Notebook (Version 7.0.6).*

3.2 Python and Libraries:

This section highlights the python libraries, and their versions used in this research project. The Language used in the project was *Python 3.10.13* within the Jupyter Notebook These libraries are necessary to replicate data preprocessing, Exploratory data analysis and modelling.
The python libraries used in the project are mentioned in Figure 2.

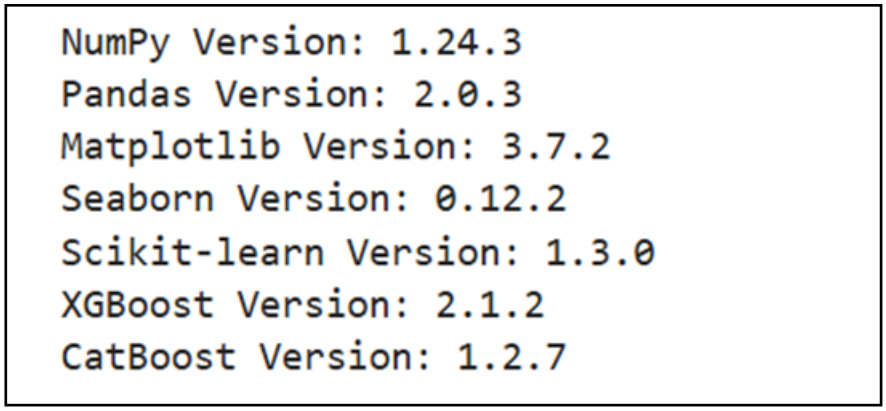

Figure 2: Library Versions

3.3 Anaconda Navigator

This research project used *Anaconda Navigator*, a desktop graphical user interface (GUI). It allows many different environments to be leveraged to perform data analysis and data related projects. Firstly, install the *Anaconda* or *Miniconda* on the device. The *anaconda navigator 2.5.0 version* was used in this project. Once downloaded and launched the navigator will look like Figure 3.

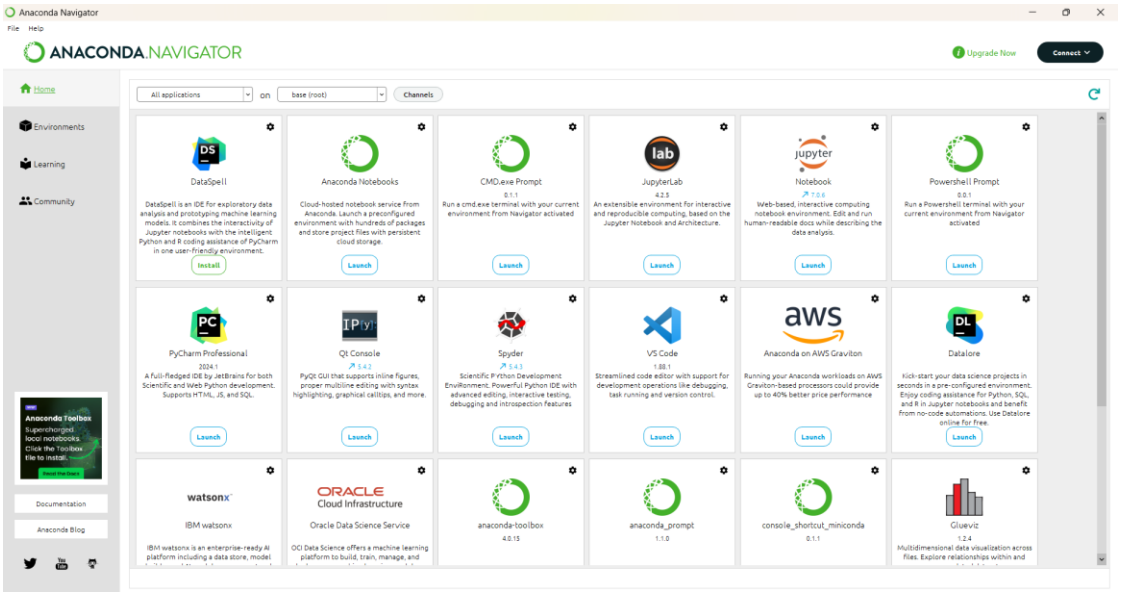

Figure 3: Anaconda navigator interface

3.4 Directory Setup

The output of this research project is a single *Jupyter Notebook* file which includes all the phases of the research in order. The ipynb file must be accessed through *Jupyter IDE*. The dependency needed to run the code file are located in the same folder as the data where the *Jupyter Notebook* file is saved. The files needed for the replication of this projects can be seen in the figure 4. These files can directly be cloned from *GitHub*: https://github.com/SARANGK999/THESIS.git.

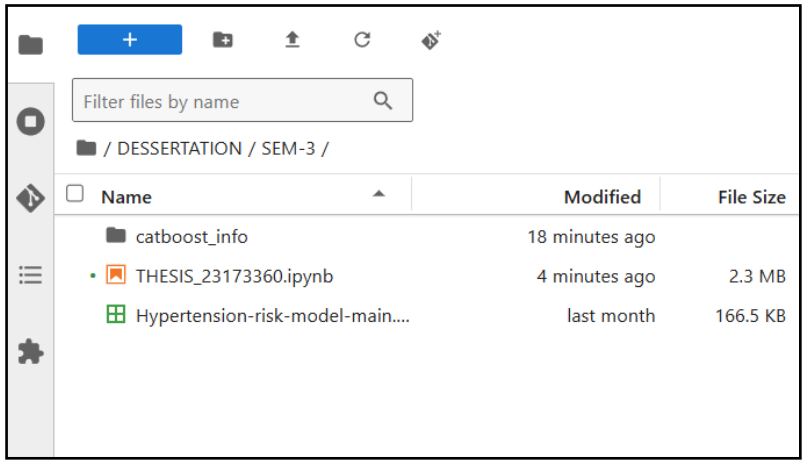

Figure 4: Directory setup.

4 Data:

The data used in this research project is publicly available on the *Kaggle Repository*. The data can be found on: https://www.kaggle.com/datasets/khan1803115/hypertension-risk-model-main. The data can directly be downloaded and used as it is by keeping it in the same directory as the python file. Or the entire setup can be cloned from GitHub. This data can be used freely under the licence: https://www.mit.edu/~amini/LICENSE.md.

5 Workflow:

After setting up *the Jupyter IDE* and dependencies in the same file path, the code file can be run directly to get the required outputs. The workflow of the python file is described step by step in figure 5.

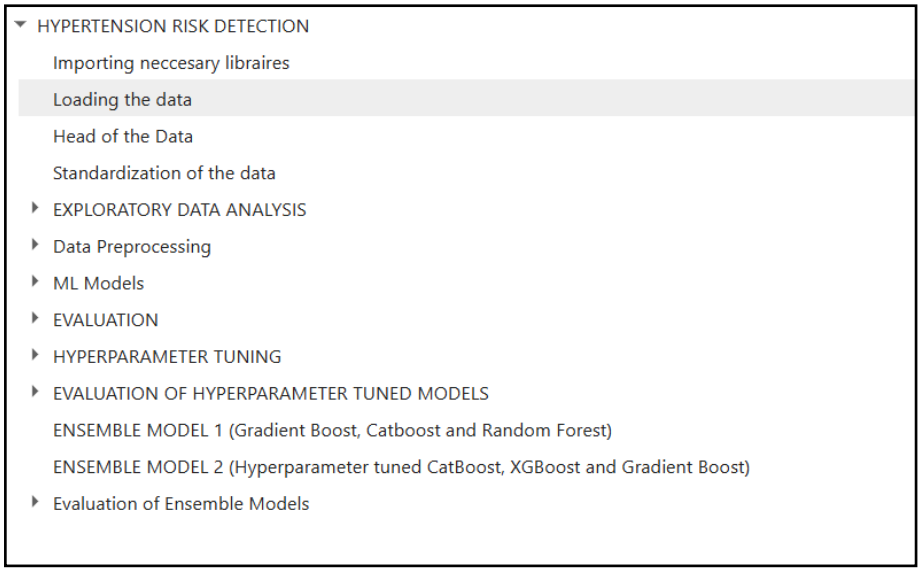

Figure 5: Workflow

6 Conclusion

The Configuration Manual describes the steps needed to replicate the analysis done in the research project. This document lists the necessary hardware, software and directory setup to replicate this project.