National College of Ireland

Configuration Manual

MSc Research Project
MSc in Data Analytics

Ripu Murdhan Chandramohan
Student ID: x20186673

School of Computing
National College of Ireland

Supervisor: Qurrat Ul Ain

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Ripu Murdhan Chandramohan
Student ID:	X20186673
Programme:	MSc in Data Analytics **Year:** 2022
Module:	Research Project
Lecturer:	Qurrat Ul Ain
Submission Due Date:	16-09-2022
Project Title:	Stroke Detection and Prediction Using Deep Learning Techniques and Machine Learning Algorithms
Word Count:	1090 **Page Count:** 7

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: Ripu Murdhan Chandramohan

Date: 16/09/2022

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

Ripu Murdhan Chandramohan
Student ID: x20186673

1 Introduction

In this Stroke Prediction and Detection Research, Python was used to develop Deep Learning methods and machine learning algorithms. The specifications for the system as well as its application are detailed in this handbook.

2 Technical Requirements

2.1 Hardware Required

System Software: Windows 10, 64 bits.

RAM: 12 GB

As far as hardware goes, these are all the requirements that have been satisfied.

2.2 Software Required

To execute the Python code, you will need to have Anaconda Jupyter notebook installed. To go to the start-up folder, type cd /some folder name at the command prompt. To open the Jupyter Notebook app, type jupyter notebook in the search bar. The user interface for the notebook will load in a new tab or window in your browser.

Computation Syntax: Python

The latest version of Anaconda, version 4.3, together with the Jupyter Notebook extensions it provides are required, as they enable users to specify environment-specific kernels directly from the Jupyter Notebook interface.

3 Implementation

Collecting dataset:
Dataset is available at https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset and may be downloaded there. There will be a file with the name Stroke Prediction Dataset available there and all it takes to download it is one click**.**

Jupyter Notebook data loading: At first, the dataset has to be stored on any local disk that is available in the system.

Figure 1 illustrates that the Dataset was kept on the D disk of my own computer, which was positioned within the stroke dataset folder.

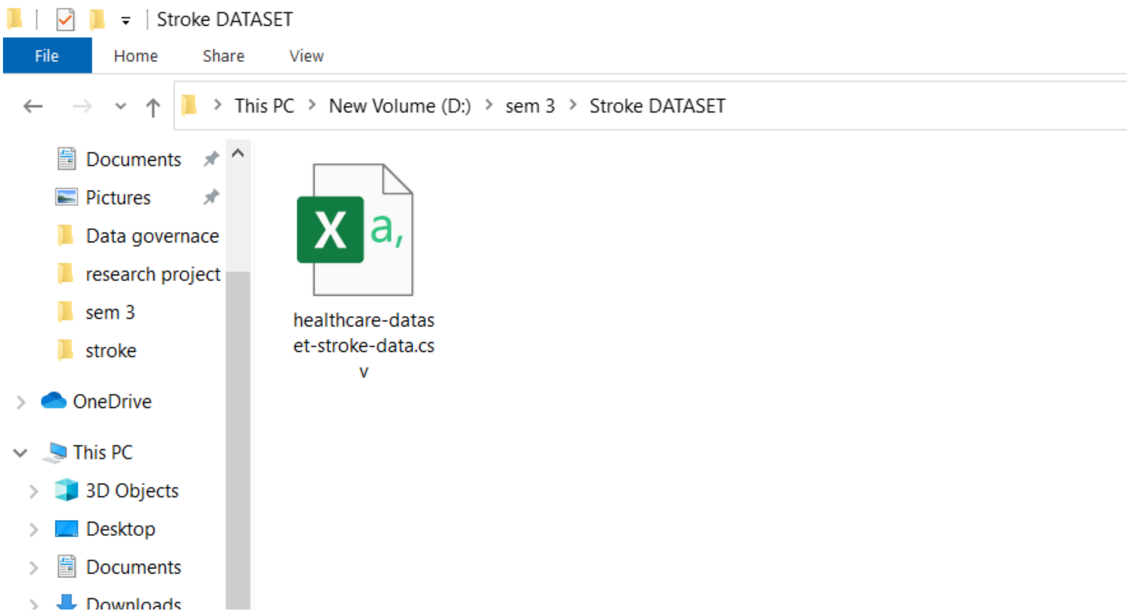

Figure 1: My PC's D drive held the dataset.

Figure 2 demonstrates that the jupyter notebook version 6.3.0 may be launched from the ANACONDA navigator. This version of the notebook uses the python programming language for its implementation.

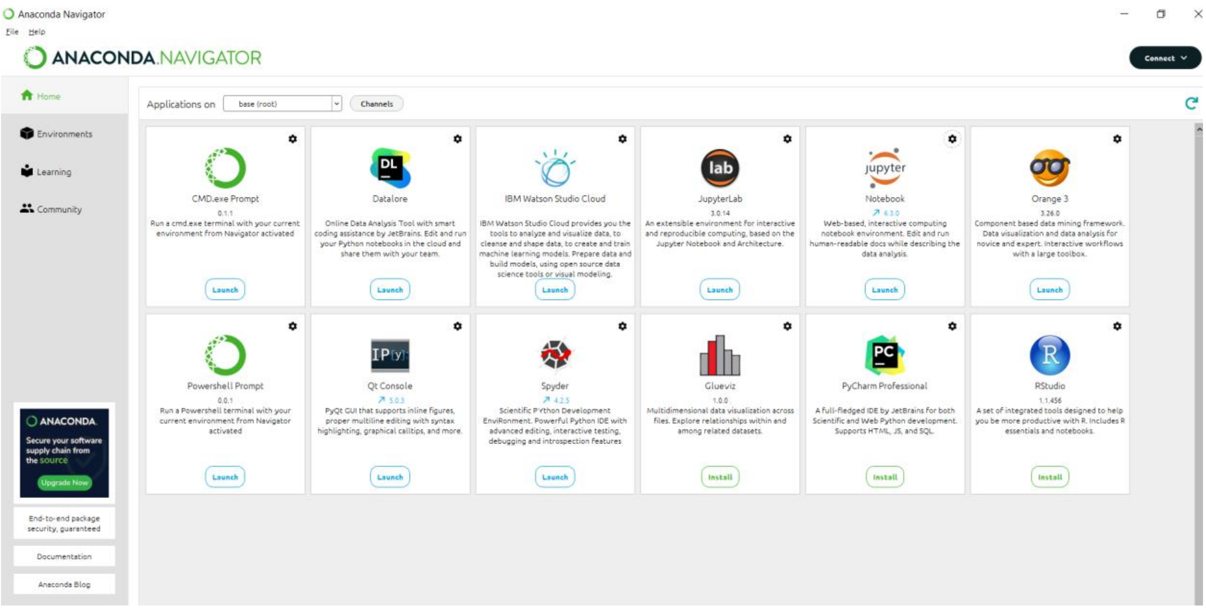

Figure 2: Anaconda Navigator

The essential libraries are depicted in Figure 3 and before the data can be loaded, they need to be imported.

```python
import pandas as pd
import numpy as np
import sklearn.metrics
import plotly.graph_objs as go
import plotly.express as px
import matplotlib.pyplot as plt
import tensorflow as tf
import seaborn as sns #data viz
import warnings
warnings.filterwarnings("ignore")
```

Figure 3: Import essential libraries

Necessary libraries: Importing the required libraries is required in order to use any of Python's built-in functions.

```python
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import  confusion_matrix, roc_auc_score, ConfusionMatrixDisplay,RocCurveDisplay, classification_report, roc_
from sklearn.model_selection import cross_val_score
from imblearn.over_sampling import SMOTE
from imblearn.over_sampling import ADASYN
from sklearn.ensemble import AdaBoostClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics, model_selection

from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.impute import KNNImputer
from sklearn.model_selection import train_test_split
from tensorflow.keras import layers, models
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.wrappers.scikit_learn import KerasClassifier
from tensorflow import keras
from sklearn.metrics import accuracy_score, recall_score ,precision_score, f1_score
from matplotlib.gridspec import GridSpec
from pywaffle import Waffle
```

Figure 4: Library requirements for experimental research and assessment techniques

These are the required libraries that need to be imported into the Jupyter notebook in order to develop models for Stroke Detection and Prediction Using Deep Learning Techniques and

Machine Learning Algorithms. As shown in figure 4, these are the libraries that need to be imported.

Figure 5 illustrates the process of installing the inflection, TensorFlow, keras, pywaffle, and plotly packages. Installing this package is required in order to make a prediction about stroke detection prior to putting into practice Deep Learning Techniques and Machine Learning Algorithms.

```
!pip install inflection
!pip install tensorflow
!pip install keras
!pip install pywaffle
!pip install plotly
```

Figure 5: package installation

It is necessary to make a copy of the location of the dataset taken from the system and paste it into the Jupyter notebook. After that, as seen in figure 6, the path that was copied has to be put inside of the brackets.

```
#Read In Dataset
stroke_data = pd.read_csv('D:/sem 3/healthcare-dataset-stroke-data.csv', index_col='id')
```

Figure 6: Set dataset path

When you press Shift+Enter from here on out, the cell you're working on will begin executing, and the dataset will load without a problem.
Following that, the phases of the implementation can be proceeded in the sequence that was specified in the Python code that was supplied.

Rest of the process:
After importing data into a Jupyter notebook, as shown above, running the supplied Python code is as easy as clicking the cell button and selecting Run All from the notebook's main menu.

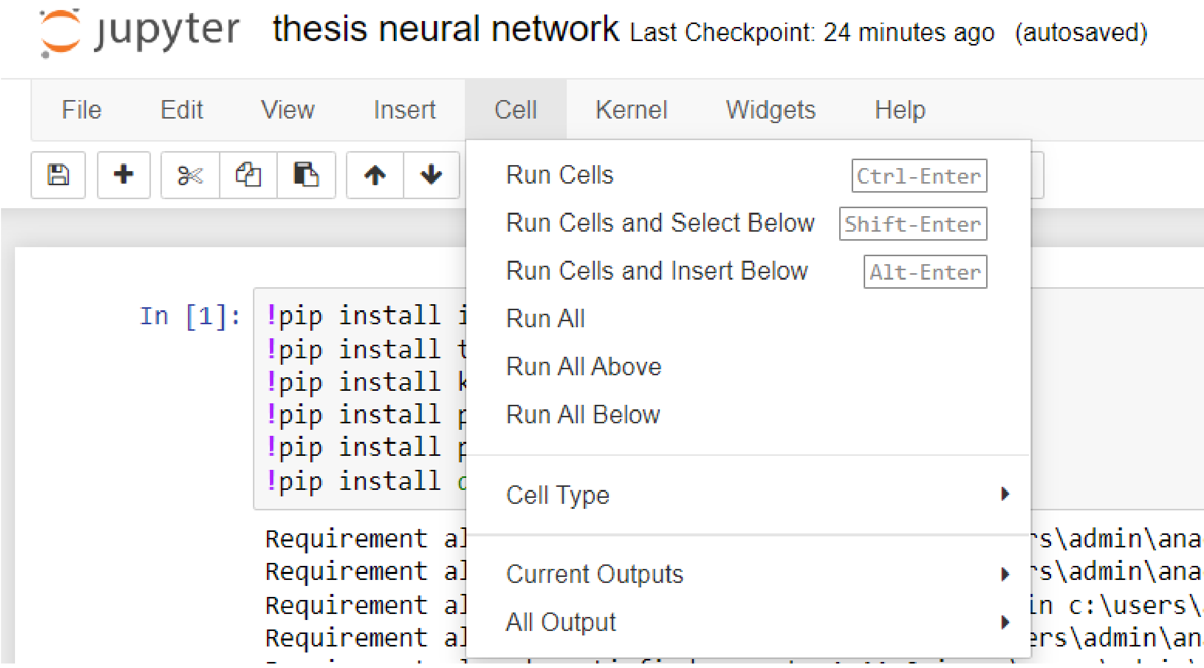

Figure 7: Main menu of Jupyter notebook

4 Conclusion

In order to correctly execute the entire code in Jupyter, it is possible to follow the procedures that have been explained above. The machine has 12 gigabytes of random-access memory, which ensures that the code will execute more quickly and smoothly.

5 Exploratory Data Analysis Visualization

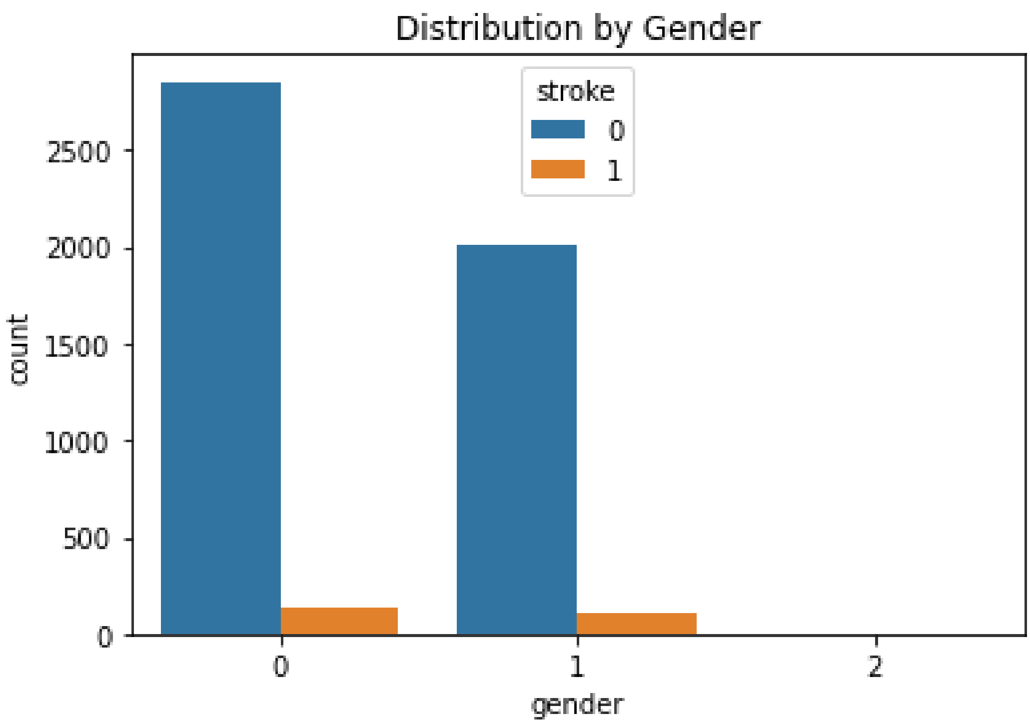

Figure 8: Distribution by Gender

Figure 8 illustrates the Distribution by Gender, which reveals that women have a higher count for no stroke than males do. This finding may be deduced from the fact that women outnumber men.

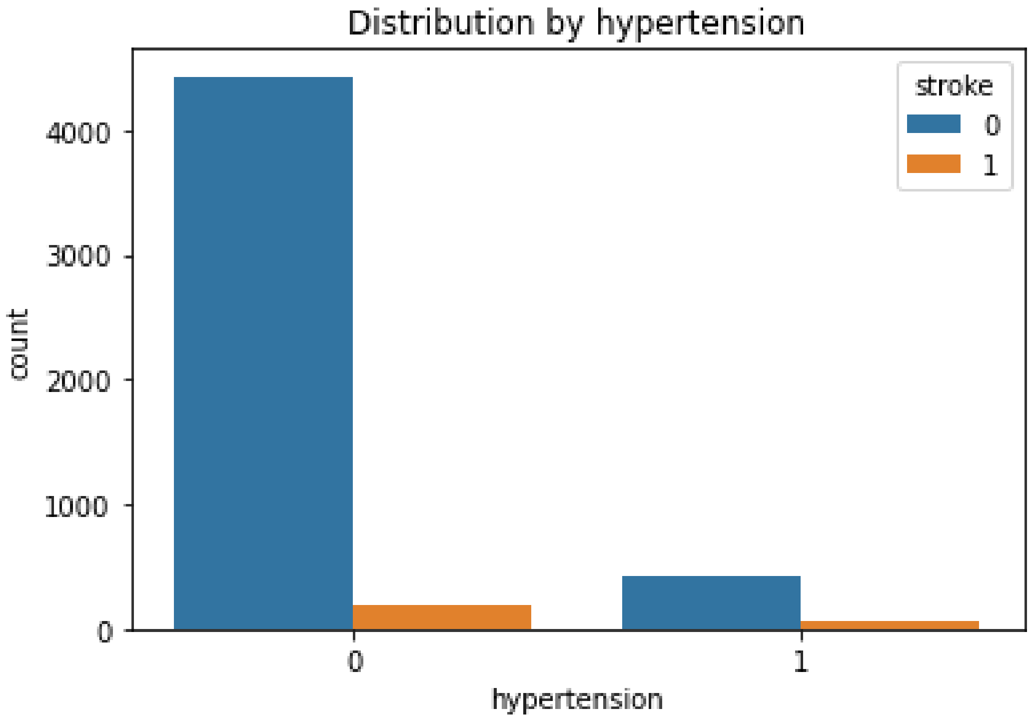

Figure 9: Distribution by Hypertension

Figure 9 depicts the Distribution by Hypertension, which demonstrates that women are influenced more by count than men are. This is shown by the fact that women are more likely to have hypertension. This conclusion may be inferred from the fact that there were more women than males who had their blood pressure examined for hypertension distribution.

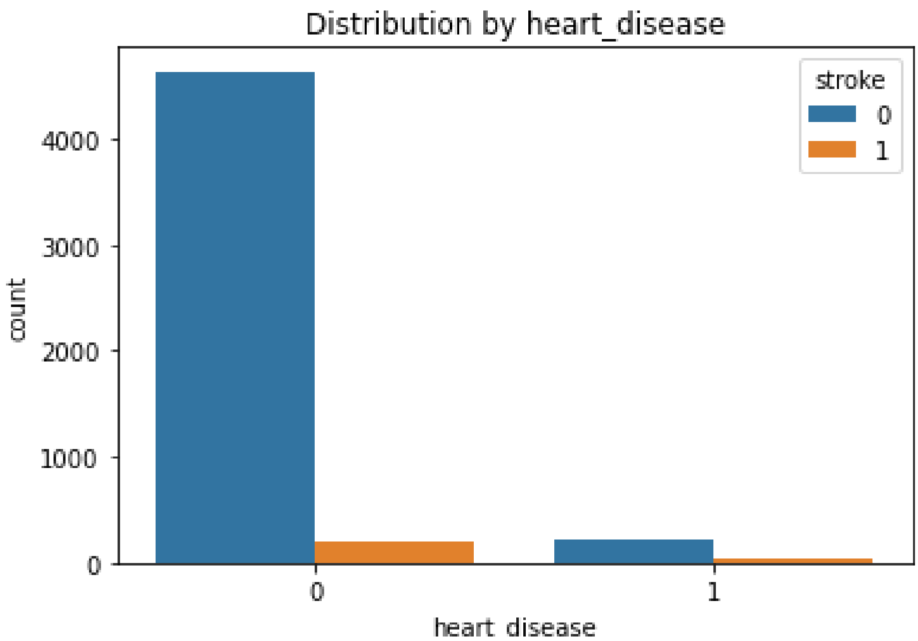

Figure 10: Distribution of Heart Disease

Figure 10 illustrates the Distribution by Heart Disease, which demonstrates that women are impacted more by count than men are. This is demonstrated by the fact that women are depicted more frequently. This is demonstrated by the fact that women have a higher risk of developing heart disease than males do. It is possible to draw this inference from the observation that there were more females than males whose blood pressure was measured in order to determine the prevalence of heart disease.