National College of Ireland

Configuration Manual

MSc Research Project
MSc in Data Analytics

Rory Gibney
Student ID: 20167482

School of Computing
National College of Ireland

Supervisor: Mohammed Hasanuzzam

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Rory Gibney
Student ID:	20167482
Programme:	MSc in Data Analytics **Year:** 2022
Module:	MSc Research Project
Lecturer:	Mohammed Hasanuzzaman
Submission Due Date:	15/12/2022
Project Title:	Using supervised learning techniques to predict kicking outcomes in the NFL
Word Count:	**Page Count: 5**

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: Rory Gibney

Date: 15/12/2022

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

Rory Gibney
Student ID: 20167482

1 Introduction

The sole purpose of this document is to provide instruction as to how best replicate the work undertaken as part of the associated research project. This research project was built with ease of replicability at its essence; therefore, the set up and pre-requisites are relatively straightforward to follow. This document will detail the necessary hardware and software requirements, as well as the file directory set up to run the code used in this research project.

2 Pre-Requisites

The data that was analysed through this analysis varies in terms of size. Due to the fact that some of the datasets are quite large, it is recommended this research be carried out on a machine with sufficient CPU and processing power to handle downloading and connecting to this data. Below is a figure detailing the hardware details of the machine used to carry out this research:

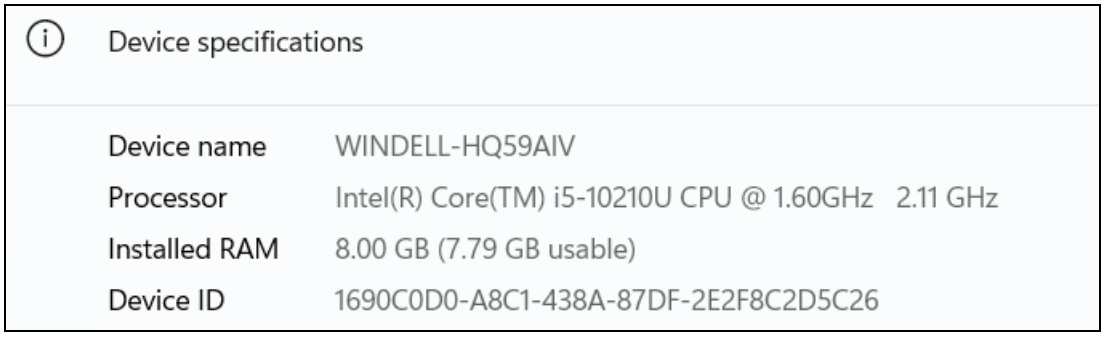


Figure 1- Hardware details of machine that conducted this research

In order to replicate this analsys effectively, a similar spec machien is needed. On average to run one of the scripts associated with this project takes this machine 5-10 minutes.

3 Software requirements

3.1 Python Base Library Download

Central to this project was the use of the Jupyer notebook IDE, or integrated development environment. But before discussing it, the Python programming standard library must be installed on the machine looking to replicate. Figure 2 below shows the version that was used in this analysis:

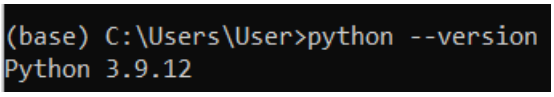


Figure 2- Python version used to run analysis

At the time of writing this manual, there is a more recent version of Python[1] available for download. It is advised the latest version of Python is installed to ensure the most up-to-date standard library classes can be used.

3.2 Anaconda Navigator

This research used Anaconda Navigator, which is a desktop graphical user interface, or GUI. It allows several different environments to be leveraged to conduct data analysis and data science projects. Instructions on how to install this can be found here[2]. The version of Anaconda used in this analysis can be seen below:

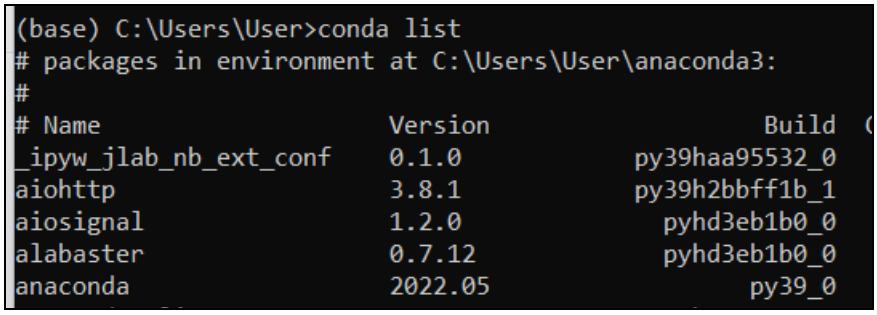


Figure 3- Anaconda version used to conduct analysis

Once downloaded, an interface like Figure 4 below will appear once the Anaconda Navigator application is opened.

[1] https://www.python.org/downloads/release/python-3111/
[2] https://docs.anaconda.com/navigator/install/

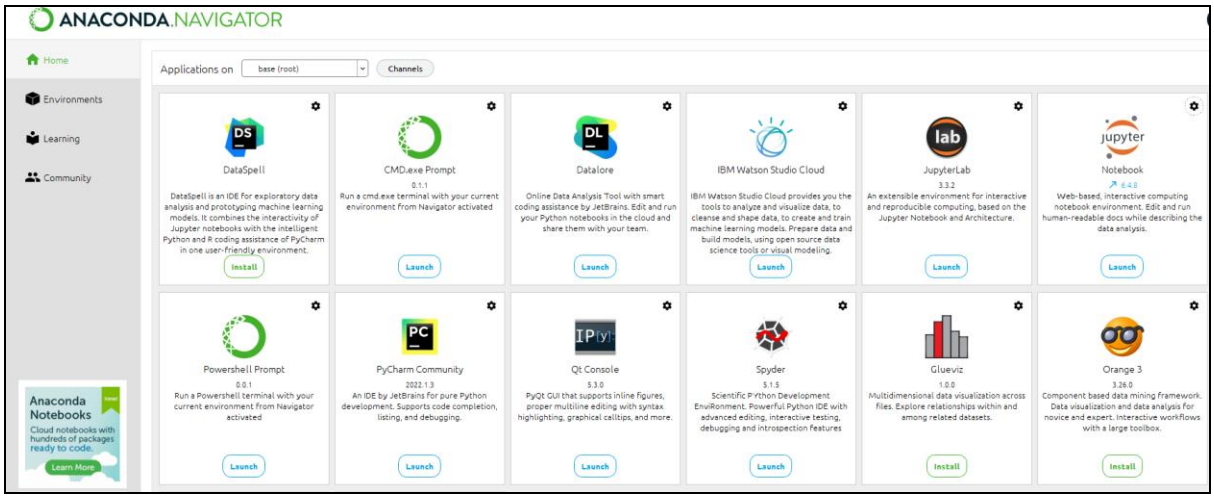


Figure 4- Anaconda Navigator GUI

Howver, before being able to replicate the analysis conducted in this research project, there are several complemantary Python packahges that will need to be installed in the Anaconda environemnet. The Ananconda Prompt[3] can be used to install the neccesary packages; a list of these packages can be found below:

- Pandas
- Numpy
- Glob
- Matplotlib
- Os
- Math
- Sklearn
- Imblearn
- Seaborn

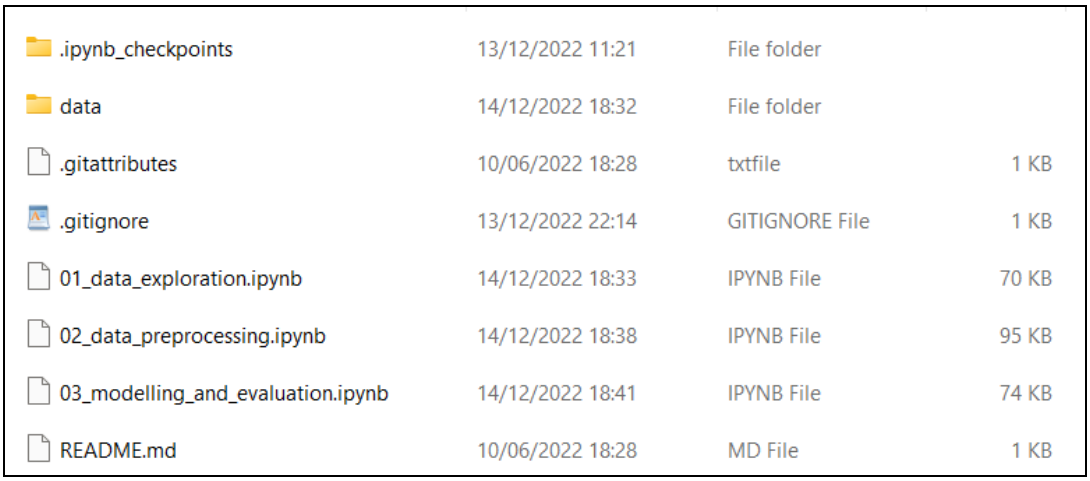
4 Jupyter Notebook

The outputs for this project are Jupyter notebook [4] files; there are 3 in total, depicting the phases this research went through. They include the following files:

- 01_data_exploration.ipynb
- 02_data_preprocessing.ipynb
- 03_modelling_and_evaluation.ipynb

These must be access using the Jupyter IDE. The final dependency needed to be fulfilled to run these files is a data folder located in the same repository as where the Jupyter files are saved. A screenshot of the required directory set up can be seen below in Figure:

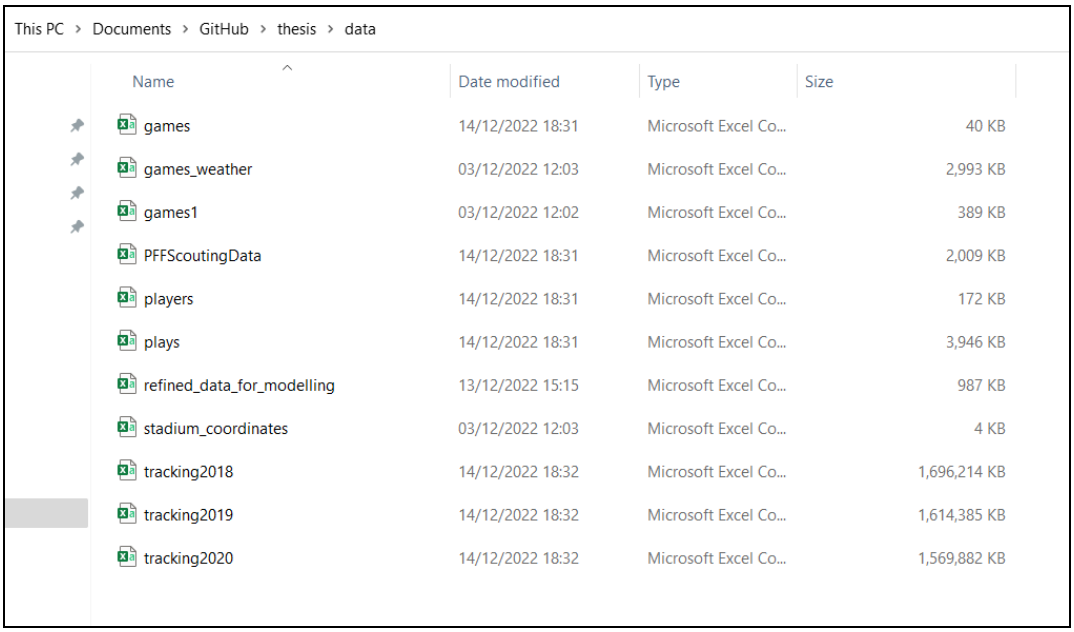
[3] https://anaconda.org/conda-forge/prompt
[4] https://jupyterbook.org/en/stable/file-types/notebooks.html



.ipynb_checkpoints	13/12/2022 11:21	File folder	
data	14/12/2022 18:32	File folder	
.gitattributes	10/06/2022 18:28	txtfile	1 KB
.gitignore	13/12/2022 22:14	GITIGNORE File	1 KB
01_data_exploration.ipynb	14/12/2022 18:33	IPYNB File	70 KB
02_data_preprocessing.ipynb	14/12/2022 18:38	IPYNB File	95 KB
03_modelling_and_evaluation.ipynb	14/12/2022 18:41	IPYNB File	74 KB
README.md	10/06/2022 18:28	MD File	1 KB

Figure 5- Directory set up to replicate analysis

Finally, these are the files present in the data directory for the Jupyter notebooks to access, import and analyse:



This PC › Documents › GitHub › thesis › data

Name	Date modified	Type	Size
games	14/12/2022 18:31	Microsoft Excel Co...	40 KB
games_weather	03/12/2022 12:03	Microsoft Excel Co...	2,993 KB
games1	03/12/2022 12:02	Microsoft Excel Co...	389 KB
PFFScoutingData	14/12/2022 18:31	Microsoft Excel Co...	2,009 KB
players	14/12/2022 18:31	Microsoft Excel Co...	172 KB
plays	14/12/2022 18:31	Microsoft Excel Co...	3,946 KB
refined_data_for_modelling	13/12/2022 15:15	Microsoft Excel Co...	987 KB
stadium_coordinates	03/12/2022 12:03	Microsoft Excel Co...	4 KB
tracking2018	14/12/2022 18:32	Microsoft Excel Co...	1,696,214 KB
tracking2019	14/12/2022 18:32	Microsoft Excel Co...	1,614,385 KB
tracking2020	14/12/2022 18:32	Microsoft Excel Co...	1,569,882 KB

Figure 6- Datasets used in the analysis

To access the Jupyter notebook files, Github[5] can be used to take the latest version. Kaggle can be used to get both all the datasets needed also.

[5] https://github.com/rorygibney96/thesis

5 Conclusion

This configuration manual comprehensively describes the steps needed to replicate the analysis. It depicts the software and directory set up to re-run the results of this research project. The video presentation can also be found at the link in the footnote[6]

[6] https://studentncirl-my.sharepoint.com/personal/x20167482_student_ncirl_ie/_layouts/15/stream.aspx?id=%2Fpersonal%2Fx20167482%5Fstudent%5Fncirl%5Fie%2FDocuments%2FRecordings%2FMeeting%20with%20Rory%20Gibney%2D20221214%5F155815%2DMeeting%20Recording%2Emp4&ga=1