National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Ajinkya Gajananrao Chintawar
Student ID: 23113561

School of Computing
National College of Ireland

Supervisor: Dr William Clifford

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Ajinkya Gajananrao Chintawar
Student ID:	23113561
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr William Clifford
Submission Due Date:	29/01/2025
Project Title:	Configuration Manual
Word Count:	797
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:	Ajinkya Gajananrao Chintawar
Date:	28th January 2025

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

Ajinkya Gajananrao Chintawar
23113561

1 Introduction

This configuration manual describes the system setup, tools, and procedures for implementing and deploying the research project *"The Fusion of Mind and Motion: Classifying Activity Levels with Biometric and Psychological Insights"*. This ensures reproducibility and is helpful to any future researchers or users interested in continuing this work.

2 System Configuration

2.1 Hardware Configuration

Component	**Specification**
Processor	Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz
RAM	16.0 GB (15.9 GB usable)
Storage	Minimum 500 GB SSD
GPU (Optional)	NVIDIA GTX 1060 or higher for model training acceleration

Table 1: Hardware Configuration

2.2 Libraries and Software Versions

This section highlights the libraries and software versions used for this project, which would ensure the reproducibility and consistency of the results. These listed tools were indispensable in data preprocessing, visualization, modeling, and deployment.

2.2.1 Libraries

The following are the Python libraries used during different stages of the project:

- **pandas:** Version 1.4.2
- **plotly:** Version 5.6.0
- **seaborn:** Version 0.11.2
- **matplotlib:** Version 3.5.1
- **scikit-learn:** Version 1.0.2

- **imbalanced-learn:** Version 0.12.3
- **joblib:** Version 1.4.2

2.2.2 Software Versions

The implementation of the project used the following software:

- **Operating System:** Windows 11
- **Python Environment:** Python 3.9.18
- **Jupyter Lab:** JupyterLab 4.0.11
- **Tableau:** Version 2023.3

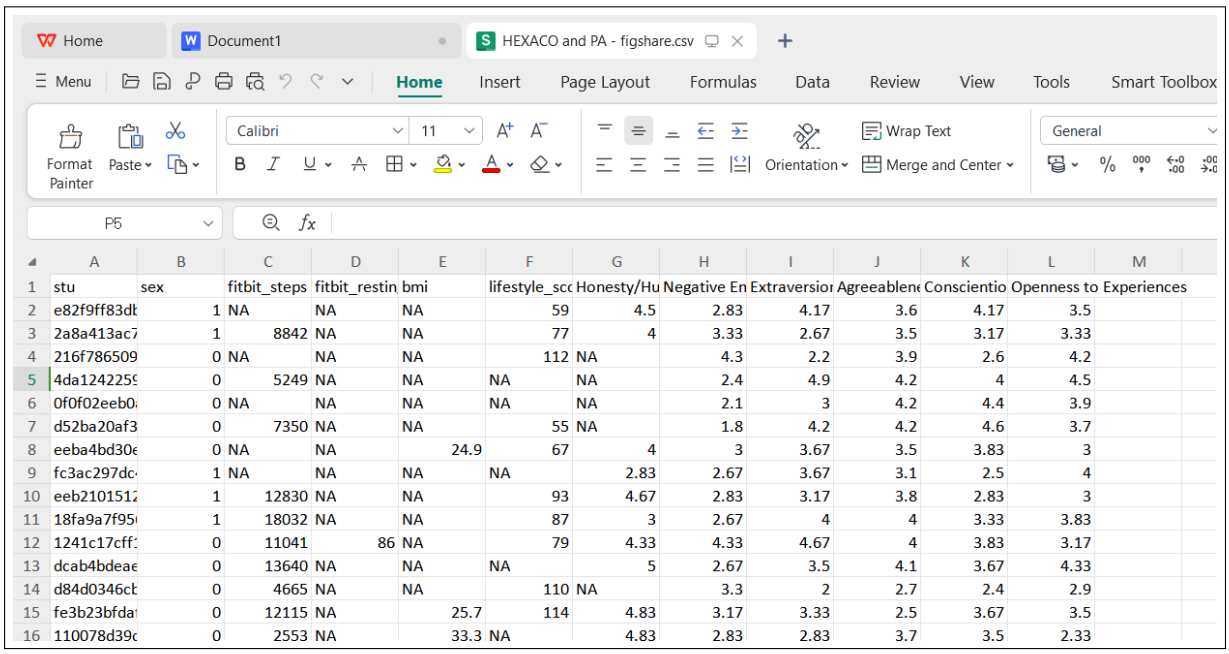
These libraries and software versions above ensure compatibility and reproducibility of the code running in various environments.

3 Dataset Preparation

3.1 Training and Testing Data

The training and testing data are obtained from the publicly available dataset found in Figshare, containing all the biometric and psychological metrics of importance for the study. The dataset is under CCO Licence and can be accessed at the following link: Figshare Dataset

The dataset's key features are visually represented in Figure 1.



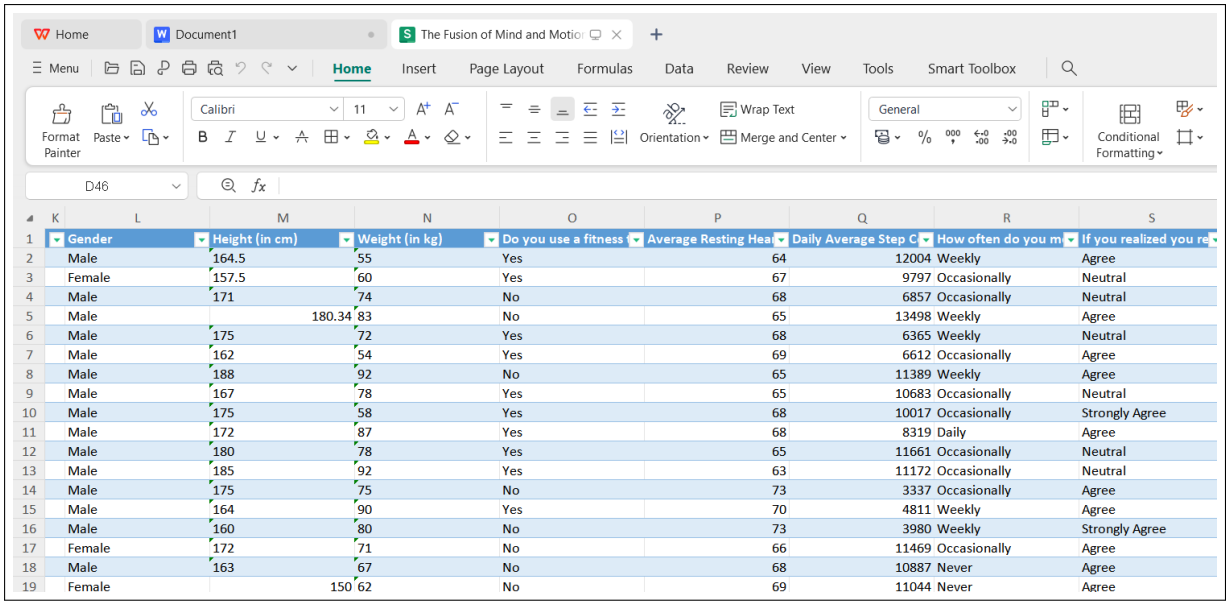
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	stu	sex	fitbit_steps	fitbit_restin	bmi	lifestyle_sc	Honesty/Hu	Negative En	Extraversior	Agreeablen	Conscientio	Openness to	Experiences
2	e82f9ff83db	1	NA	NA	NA	59	4.5	2.83	4.17	3.6	4.17	3.5	
3	2a8a413ac7	1	8842	NA	NA	77	4	3.33	2.67	3.5	3.17	3.33	
4	216f786509	0	NA	NA	NA	112	NA	4.3	2.2	3.9	2.6	4.2	
5	4da1242259	0	5249	NA	NA	NA	NA	2.4	4.9	4.2	4	4.5	
6	0f0f02eeb0	0	NA	NA	NA	NA	NA	2.1	3	4.2	4.4	3.9	
7	d52ba20af3	0	7350	NA	NA	55	NA	1.8	4.2	4.2	4.6	3.7	
8	eeba4bd30e	0	NA	NA	24.9	67	4	3	3.67	3.5	3.83	3	
9	fc3ac297dc	1	NA	NA	NA	NA	2.83	2.67	3.67	3.1	2.5	4	
10	eeb2101512	1	12830	NA	NA	93	4.67	2.83	3.17	3.8	2.83	3	
11	18fa9a7f95	1	18032	NA	NA	87	3	2.67	4	4	3.33	3.83	
12	1241c17cff	0	11041	86	NA	79	4.33	4.33	4.67	4	3.83	3.17	
13	dcab4bdeae	0	13640	NA	NA	NA	5	2.67	3.5	4.1	3.67	4.33	
14	d84d0346cb	0	4665	NA	NA	110	NA	3.3	2	2.7	2.4	2.9	
15	fe3b23bfda	0	12115	NA	25.7	114	4.83	3.17	3.33	2.5	3.67	3.5	
16	110078d39c	0	2553	NA	33.3	NA	4.83	2.83	2.83	3.7	3.5	2.33	

Figure 1: Snippet of the Figshare Dataset

3.2 Deployment Data

The deployment dataset is survey data from Microsoft Forms, targeting people who are 18 and above in age. This dataset underwent preprocessing to make them match the training data.

Some key features of the Participants dataset are visualized in Figure 2.



	Gender	Height (in cm)	Weight (in kg)	Do you use a fitness	Average Resting Hea	Daily Average Step C	How often do you m	If you realized you re
2	Male	164.5	55	Yes	64	12004	Weekly	Agree
3	Female	157.5	60	Yes	67	9797	Occasionally	Neutral
4	Male	171	74	No	68	6857	Occasionally	Neutral
5	Male	180.34	83	No	65	13498	Weekly	Agree
6	Male	175	72	Yes	68	6365	Weekly	Neutral
7	Male	162	54	Yes	69	6612	Occasionally	Agree
8	Male	188	92	No	65	11389	Weekly	Agree
9	Male	167	78	Yes	65	10683	Occasionally	Neutral
10	Male	175	58	Yes	68	10017	Occasionally	Strongly Agree
11	Male	172	87	Yes	68	8319	Daily	Agree
12	Male	180	78	Yes	65	11661	Occasionally	Neutral
13	Male	185	92	Yes	63	11172	Occasionally	Neutral
14	Male	175	75	No	73	3337	Occasionally	Agree
15	Male	164	90	Yes	70	4811	Weekly	Agree
16	Male	160	80	No	73	3980	Weekly	Strongly Agree
17	Female	172	71	No	66	11469	Occasionally	Agree
18	Male	163	67	No	68	10887	Never	Agree
19	Female	150	62	No	69	11044	Never	Agree

Figure 2: Snippet of the Participants Dataset

4 Execution and Setup

This section outlines the steps required to set up the environment and execute the project code. The instructions ensure smooth replication of the results and provide guidance for using the project notebooks effectively.

4.1 Setup Instructions

To execute the project code, follow these straightforward steps:

1. Open the Anaconda Command Prompt and navigate to the folder where the project files are located. For example:

   ```
   cd /path/to/project/
   ```

2. Launch JupyterLab by typing the following command:

   ```
   jupyter lab
   ```

3. In JupyterLab, open the `Classification_ActivityLevel.ipynb` notebook and execute it:

- Select the "Restart Kernel and Run All Cells" option to ensure all code cells execute in sequence without errors.

4. Afterward, open the `Deployment.ipynb` notebook and execute it similarly:
    - Again, choose the "Restart Kernel and Run All Cells" option for consistency.

Figure 3 below shows an example of navigating to the project directory and launching JupyterLab using the Anaconda Command Prompt.

Figure 3: Example: Navigating to the Project Directory and Launching JupyterLab.

4.2 Running the Code

To execute the project, follow these simple steps in JupyterLab:

1. **Classification Notebook:**
    - Open the file `Classification_ActivityLevel.ipynb` in JupyterLab.
    - Once opened, go to the "Kernel" menu and select *Restart Kernel and Run All Cells* to ensure all cells execute properly in sequence.
    - This notebook handles data preprocessing, exploratory data analysis, and model training, giving you insights and training results.
2. **Deployment Notebook:**
    - Open the file `Deployment.ipynb` in JupyterLab.
    - Similarly, select *Restart Kernel and Run All Cells* from the "Kernel" menu to execute the entire notebook.
    - This notebook uses the trained model to predict activity levels for the survey dataset and prepares data for visualization.

To better understand this process, Figure 4 shows an example of how to restart the kernel and run all cells in JupyterLab.

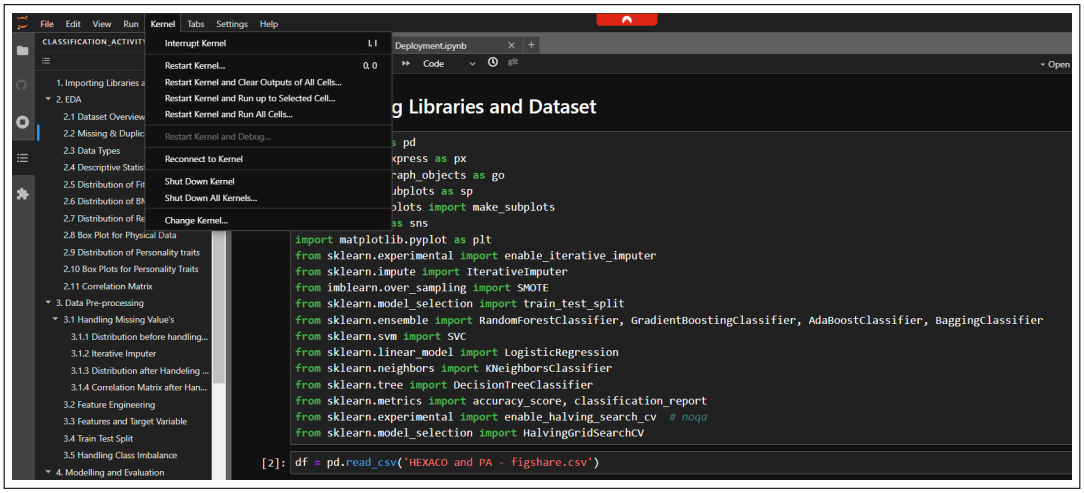


Figure 4: Example: Restart Kernel and Run All Cells in JupyterLab.

4.3 Visualization

For exploring results interactively:

- Access the Tableau Dashboard developed for visualizing relationships between physical activity levels, personality traits, and biometric data.
- Link: Activity Level Dashboard on Tableau

5 Conclusion

This configuration manual serves as an accessible and practical guide to setting up and running the project, ensuring that the process is both easy to follow and adaptable for future research. By seamlessly integrating biometric data and personality traits, the project demonstrates how machine learning can effectively classify physical activity levels while providing valuable, actionable insights. Careful attention was given to ethical considerations, including maintaining participant anonymity and adhering to strict data privacy standards. This approach not only lays the groundwork for future advancements in personalized health monitoring but also highlights the potential of machine learning to make a meaningful impact on well-being.