Configuration Manual

MSc Research Project
Artificial Intelligence for Business

Karla Priscila Vale de Sousa
Student ID: 23152516

School of Computing
National College of Ireland

Supervisor: Victor Del Rosal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Karla Priscila Vale de Sousa
Student ID:	X2315254
Programme:	MSc Artificial Intelligence for Business **Year:** 2023
Module:	MSc Research Project
Supervisor:	Victor Del Rosal
Submission Due Date:	16/09/2024
Project Title:	Optimising Direct Marketing Through Data-Driven Analytics and Predictive Models
Word Count:	384 **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: Karla Priscila Vale de Sousa

Date: 16th September 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

Karla Priscila Vale de Sousa
23152516

1 Introduction

This document gives a detailed guidance for configuring the necessary hardware and software environments essential for executing the research project. It includes instructions on setting up the system, specifying the required software libraries, hardware specifications, and other essential components.

2 System Specification

This section provides a detailed overview of the hardware and software components used in this research project.

2.1 Hardware Specification

This subsection outlines the hardware specifications of the device used in this research.

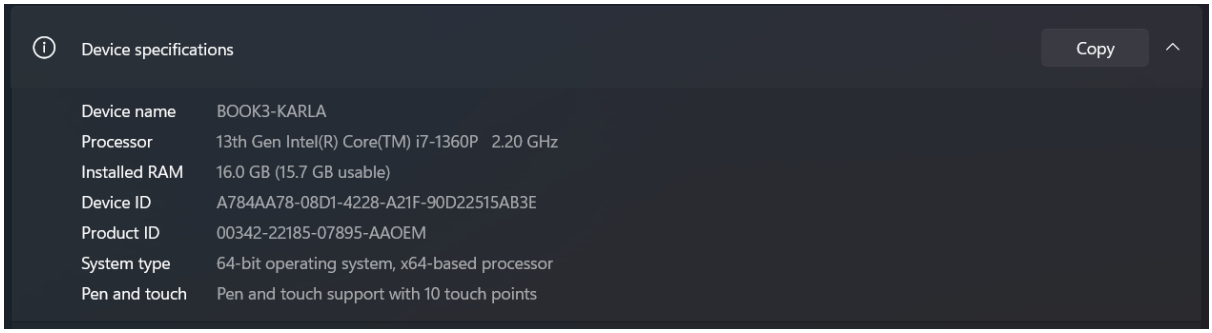

Figure 1: Device Hardware Specification

2.2 Software Specification

This subsection details the software environment configured on the device to conduct the research. It includes the operating system specifications, as well as versions of crucial software tools such as Anaconda Navigator, Jupyter Notebook, and Jupyter Lab. Anaconda was employed as the primary platform to manage the Python environment and associated packages required for this research. Jupyter Notebook and Jupyter Lab were launched through Anaconda to open and work within the Python 3 environment, facilitating code execution, data analysis, and documentation.

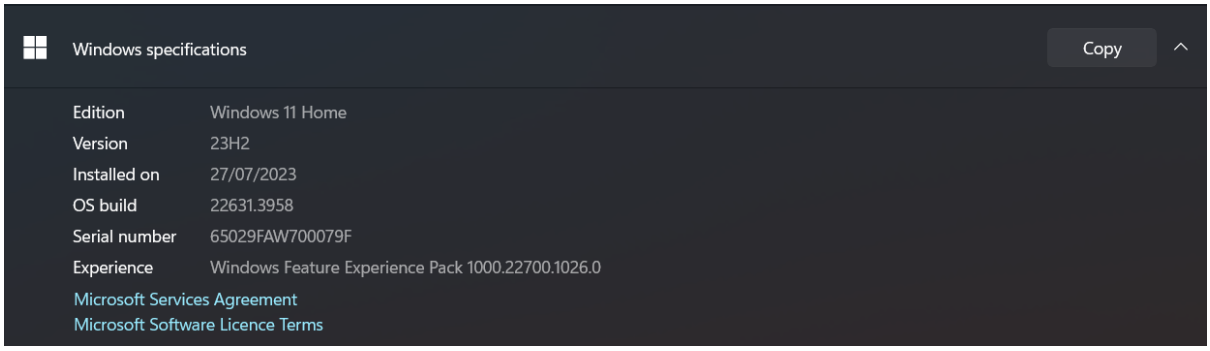

Figure 2: Device Software Specification

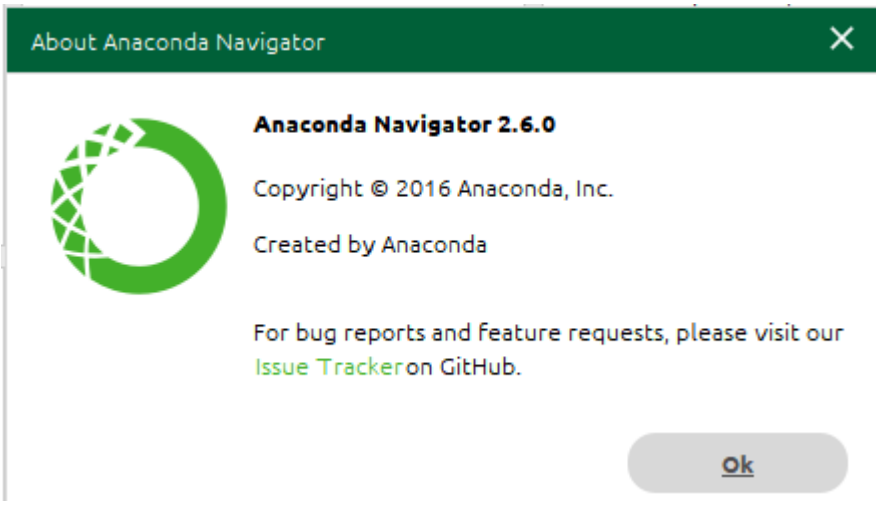

Figure 3: Anaconda Navigator Version

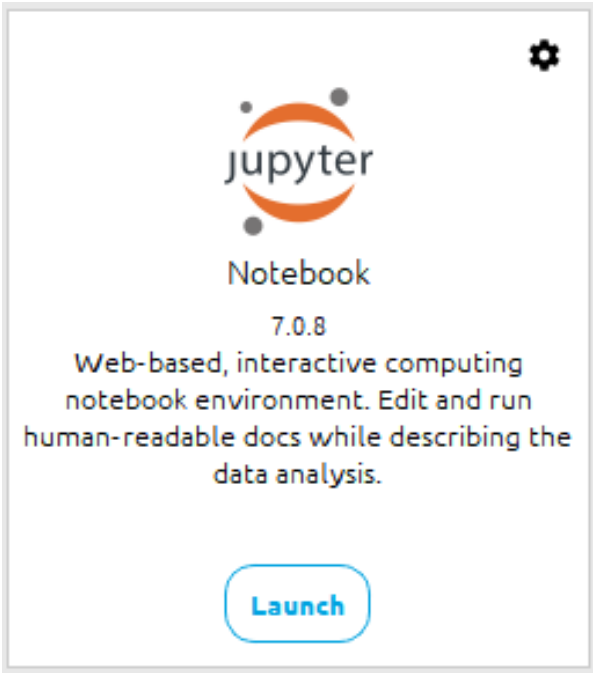

Figure 4: Jupyter Notebook Version

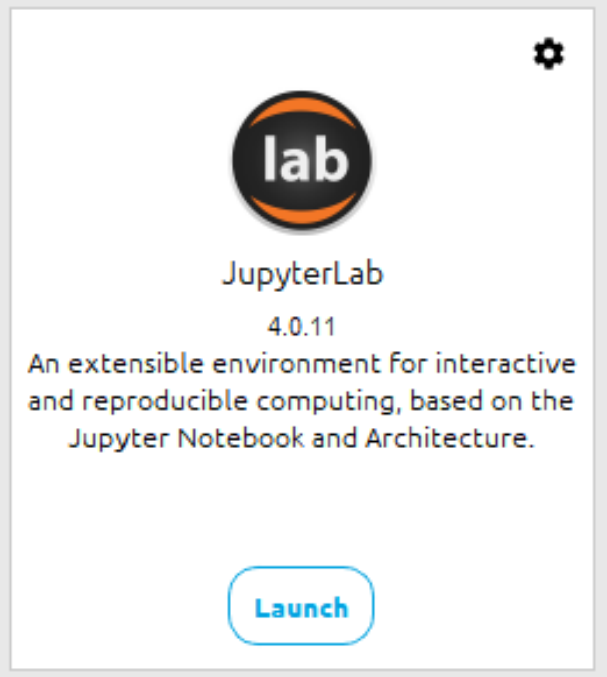

Figure 5: Jupyter Lab Version

```
!python --version
```

```
Python 3.12.4
```

Figure 6: Python Version

3 Implementation

3.1 Directory Structure

The directory structure of the "ifood-da-case" project is organized into four main folders: case, data, notebooks, and report. The case folder contains the case description. The data folder contains csv raw and processed data files, while the notebooks folder includes Jupyter notebooks that cover different stages of the project, such as exploratory data analysis, customer segmentation, and predictive modelling, along with a file listing the library versions used. The report folder contains the eda_prof file, a detailed report generated by ydata-profiling that was generated for initial exploratory data analysis.

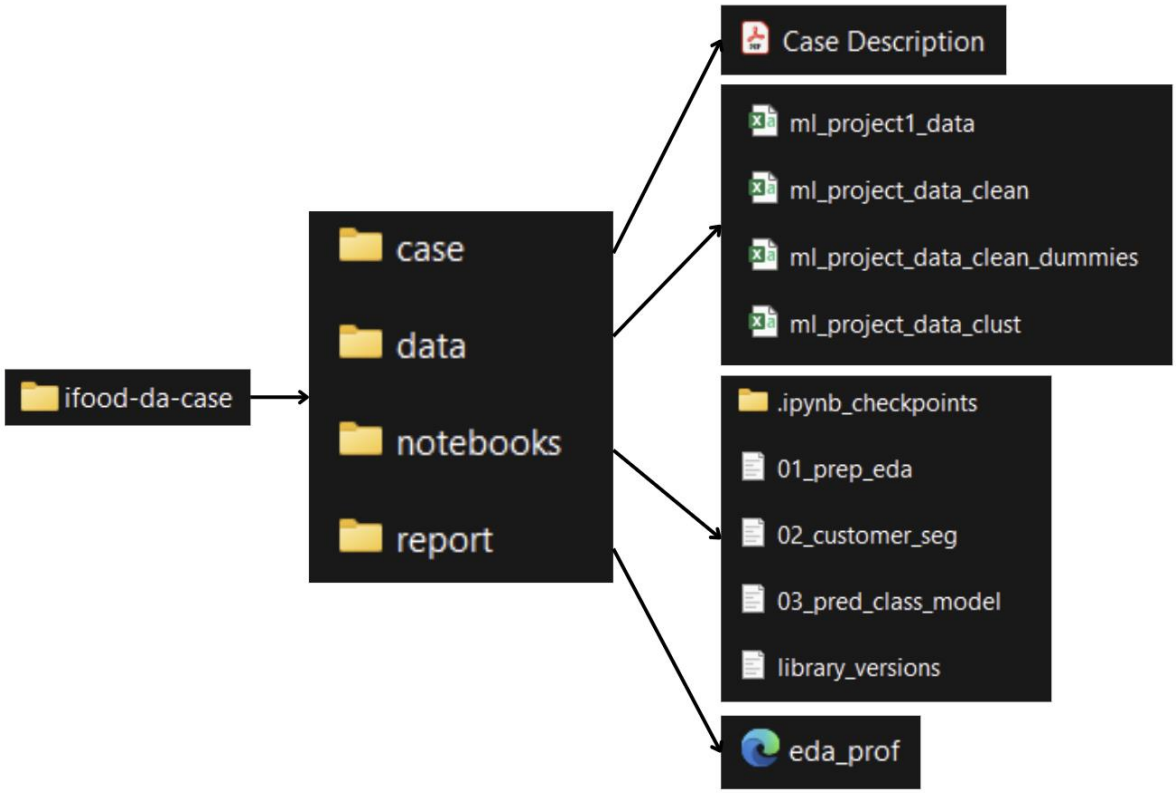

Figure 7: Project Directory Structure

3.2 Package Requirements

Various Python libraries were employed to preprocess data, build, evaluate, and visualize a machine learning model. NumPy (1.26.4) and Pandas (2.2.2) handled numerical operations and data management, Scikit-Learn (1.4.2) provided tools for building and evaluating the machine learning model, Imbalanced-Learn (0.12.3) managed class imbalances, and Matplotlib (3.8.4) and Seaborn (0.13.2) created visualizations to interpret model performance. All were integrated in a Python (3.12.4) environment.

```
Versions of the packages:

---------------------------------
Package              |    Version
---------------------------------
Imbalanced-Learn     |     0.12.3
Matplotlib           |      3.8.4
NumPy                |     1.26.4
Pandas               |      2.2.2
Scikit-Learn         |      1.4.2
Seaborn              |     0.13.2

---------------------------------
Python               |     3.12.4
```

Figure 8: Python Libraies Versions

```
import pandas as pd
import matplotlib
import seaborn as sns
import numpy as np
import sklearn
import matplotlib.pyplot as plt
import seaborn as sns
```

```
from ydata_profiling import ProfileReport
```

Figure 9: Python Libraries 01_prep_eda

```
import os
os.environ["OMP_NUM_THREADS"] = "9"
```

```
import pandas as pd
import matplotlib
import seaborn as sns
import numpy as np
import sklearn
import matplotlib.pyplot as plt
```

```
from matplotlib.colors import ListedColormap
from matplotlib.ticker import PercentFormatter
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.preprocessing import OneHotEncoder, StandardScaler, MinMaxScaler, PowerTransformer
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
```

Figure 10: Python Libraries 02_customer_seg

```
import pandas as pd
import matplotlib
import seaborn as sns
import numpy as np
import sklearn
import imblearn
import matplotlib.pyplot as plt
```

```
from sklearn.dummy import DummyClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import StratifiedKFold
from sklearn.preprocessing import OneHotEncoder, StandardScaler, MinMaxScaler, PowerTransformer
from sklearn.compose import ColumnTransformer
from imblearn.pipeline import Pipeline
from sklearn.feature_selection import SelectKBest, f_classif
from imblearn.under_sampling import RandomUnderSampler
from sklearn.model_selection import cross_validate
from sklearn.metrics import RocCurveDisplay
from sklearn.model_selection import cross_val_predict
from sklearn.metrics import PrecisionRecallDisplay
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, ConfusionMatrixDisplay
```

Figure 11: Python Libraries 03_pred_class_model