

Configuration Manual

Research Practicum Part 2
MSc Data Analytics

Athul Ajayghosh .
Student ID: 23416068

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Athul Ajayghosh .
Student ID: 23416068
Programme: MSc Data Analytics **Year:**2025.....
Module: Research Practicum Part 2
Lecturer: Vikas Sahn
Submission Due Date: Thursday 11th December 2025
Project Title: Critical study of Machine Learning Models for Retail Sales Predictions
Word Count:941..... **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:

A handwritten signature in black ink, appearing to read "Athul Ajayghosh", written over a dotted line.

Date:

.....24-11-2025.....

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

Athul Ajayghosh .
Student ID: 23416068

1 Introduction

The following configuration manual contains the technical instructions to install, configure as well as run the forecasting system created in this research project. The system uses seventeen types of forecasting systems, such as classical time-series systems, machine learning systems, deep learning networks, and hybrid systems.

The manual aims at allowing every researcher, lecturer, or external examiner to re-create the entire modelling pipeline, re-run the experiments, and confirm the findings provided in the dissertation. This document will include the software requirements, directory structure, setup procedures and the instructions to run and retraining models.

2 System Requirements

This project was deployed and run in its entirety in a Python-based data analytics environment. The whole modelling, experimentation and evaluation procedures were based on available open-source tools to make the process replicable. The experiments were conducted on a local machine with an AMD Ryzen 7 (7000 series) processor, NVIDIA RTX 4060 graphics card, and 16GB of RAM and Windows 11 Home Single Language

- Version : 25H2
- OS build : 26200.7171
- Experience : Windows Feature Experience Pack 1000.26100.265.0.

Jupyter Notebook was used as the primary development environment owing for transparency, the capability of running sequential pipelines, and the presentation attributes of both code and findings. It was implemented in Python 3.12 and a variety of scientific libraries such as:

data cleaning, transformation and aggregation: **panda and NumPy**.
scikit-learn, XGBoost, CatBoost, and LightGBM machine learning models.
TensorFlow in the deep learning models of LSTM, GRU, and CNN-LSTM.
statsmodels and **pmdarima** to classically statistics modelling.
Prophet for Structural time-series prediction.
Visualizations using matplotlib and seaborn

To avoid dependency conflicts and to make sure that the libraries could be reproducible, all libraries were installed in a virtual Python environment with the help of pip.

2.1 Software used

- Python 3.12.7

- Jupyter Notebook
- Required Python libraries:
 - Pandas: 2.3.1
 - NumPy: 1.26.4
 - Matplotlib: 3.10.5
 - Seaborn: 0.13.2
 - Scikit-learn: 1.7.2
 - XGBoost: 3.0.0
 - CatBoost: 1.2.8
 - LightGBM: 4.6.0
 - TensorFlow: 2.20.0
 - Statsmodels: 0.14.5
 - Pmdarima: 2.0.4
 - Prophet: 1.2.1

3 Dataset

The forecasting system is based on the UCI Online Retail ¹ dataset, which was downloaded directly on the UCI Machine Learning Repository to make the data integrity. The raw data was given in.xlsx format and the data imported using pandas.

4 Directory Structure

```
Retail-Forecasting/
├── OnlineRetail.xlsx
└── Retail_Forecasting.ipynb
```

5 Code Execution

1. Start Jupyter Notebook

Open Jupyter notebook in your python environment. This brings up a browser window into which all notebooks would be open.

2. Open the project file

Navigate to the project directory and open Retail_Forecasting.ipynb. The entire forecasting pipeline is present in this notebook.

3. Run Cells Sequentially

Process all the cells one at a time. The layout of the notebook is a modular workflow, and therefore the order in which cells are run is important to guarantee that all the variables, datasets, and models are loaded properly. Give Time to Model Implementation. The run time of classical and machine learning models is low and the run time of deep learning and hybrid models can be several seconds to train based on hardware.

4. Review Console Outputs

Both models display accuracy, training time and prediction time in the notebook. These are an indication that the cell has been successfully performed.

¹ <https://archive.ics.uci.edu/dataset/352/online%2Bretail>

5. Test Generated Forecasts.

Each model has its forecast arrays that are stored in the notebook and plotted in the notebook- including actual vs. predicted comparisons.

6. Review Evaluation Part.

The last part summarizes all the results automatically in the form of a resultsdf and creates comparison charts and a full dashboard figure.

7. Locate Exported Outputs

Upon completion of the execution, the following files will be automatically generated to the project directory (modelcomparison.txt and combinedashboard.png)

8. Repeat or Modify as Needed

To make changes to the model parameters or even re-run the experiment, just edit the result corresponding cell and run it again. The automatic update of results will occur.