National College of Ireland

Configuration Manual

MSc Research Project
MSc Data Analytics: MSCDA_JANOL24

Jair Rodrigues da Silva Junior
Student ID: x23272112

School of Computing
National College of Ireland

Supervisor: Mohammed Hasanuzzaman

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Jair Rodrigues da Silva Junior
Student ID:	X23272112
Programme:	MSC Data Analytics **Year:** 2
Module:	Research Project
Lecturer:	Mohammed Hasanuzzaman
Submission Due Date:	28/01/2026
Project Title:	**Configuration manual** – Evaluating the Generalisability of Deep Learning Models for Global Fertility Rate Forecasting: A Region-Aware Comparative Analysis
Word Count:	**620 Page Count: 4**

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: Jair Rodrigues da Silva Junior

Date: 28/01/2026

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

Jair Rodrigues
Student ID: x23272112

1 Verification of the system

All components of the system have been executed end-to-end on a local Jupyter Notebook environment ensuring reproducibility of all results, figures, tables, and models outputs. The following items were individually tested and verified:

1. Data acquisition from the World Bank API [1] via the *wbapi* Python library
2. Data preprocessing pipeline, including:
    a. Cleaning and imputation
    b. Min-max scaling
    c. Correlation matrix generation
3. Model training workflows, including ARIMA, Dense, LSTM variants, BiLSTM, and CNN architectures using TensorFlow/Keras.
4. Spain pilot experiment, including hyperparameter grid evaluation.
5. Cross-regional generalisation, including strict and per-country scaling workflows.
6. Evaluation metrics (MAE, RMSE, MAPE) computation.
7. Automatic generation of all results, including:
    a. Top Spain configurations
    b. Region-level weighted scores
    c. Diagnostic plots
8. Exports, including temporary CSVs and visual outputs (PNG).

2 Description of sub-systems

The project system is composed of the following subsystems:

2.1 Data Acquisition

- Retrieves Total Fertility Rate (TFR) and socioeconomic indicators from the World Bank WDI API [1] using the *wbgapi* Python package.
- Handles multiple-country, multi-indicator queries.
- Outputs raw dataframes for cleaning and modelling.

2.2 Data Preprocessing

- Performs merging of indicators across countries and years.
- Handles missing values using forward fill interpolation.
- Applies Min-Max scaling (either global Spain-scaler and per-country scaler).
- Generates correlation heatmaps and filtered sets.

2.3 Model Construction

Implements eight predictive models:

1. ARIMA (Statsmodels)
2. Dense Neural Network
3. LSTM
4. Auto-LSTM
5. A-LSTM (Attention)
6. GRU
7. BiLSTM
8. 1D CNN

Built using TensorFlow/Keras, with a shared hyperparameter grid.

2.4 Model Training

- Uses chronological 80/20 train-test split.
- Includes early stopping and validation tracking.
- Trains each architecture-configuration combination for Spain.

2.5 Model Evaluation & Metrics

- Computes MAE, RMSE, and MAPE.
- Produces sorted model rankings.
- Generates diagnostic plots.

2.6 Cross-Regional Generalisation

- Applies selected recurrent model (GRU) to global regions.
- Two configurations supported:
  - Strict generalisation
  - Per-country scaling
- Produces weighted regional performance tables.

2.7 Visualisation

- Generates:
  - Correlation heatmaps
  - Time-series plots
  - Boxplots / region-wise visualisations
- Saves images to disk.

3 Installation Guide

This section provides installation instructions for setting up the runtime environment.

3.1 Required Software

- Python 3.10
- Jupyter Notebook (Anconda or standalone)
- Internet connection for World Bank API

3.2 Versions considered

Library	Version

Python	3.10
TensorFlow	2.15+
Keras	Included with TF
Wbgapi	1.x
Pandas	2.1+
Numpy	1.26+
Statsmodels	0.14+
Scikit-learn	1.3+
Matplotlib	3.8+
Seaborn	0.13+

3.3 Obtaining API Data

The World Bank API [1] is openly available. No authentication key is required.

4 System configuration

4.1 Notebook configuration

No external configuration files are required. All configuration is handled within the notebook:

- Global random seeds
- Hyperparameter grids
- Region lists
- Indicator definitions
- API query setups
- Output directory creation

4.2 External dependencies

- World Bank API [1] assessed through *wbgapi*
- No keys, token, or private credentials required.
- All accessed endpoints are public and open license.

5 Implementation details

5.1 Execution Environment

The final build was executed on:

- Windows (local machine)
- Jupyter Notebook
- CPU-only environment (GPU optional but not required)

5.2 Model configuration settings

- Sequence lengths tested (time steps): 3, 5, 7, 10.
- Hidden units tested: 64, 128.
- Dropout rate: 0.2
- Optimiser: Adam
- Loss Function: MSE

5.3 Training settings

- Batch size: 16
- Epochs: up to 60 with early stopping (patience = 8)
- Train-test:80/20
- Scaler fitting strategy: global vs per-country

5.4 Cross-Regional generalisation workflow

- Load Spain-trained GRU weights
- Apply either:
    a. Spain scaler (strict)
    b. Country-specific scaler
- Produce regional weighted MAE tables.

6 User Guide

1. Open the notebook "FRP Model_v12.ipynb"
2. Run all cells from top to bottom
3. The notebook automatically:
    a. Downloads data
    b. Preprocesses
    c. Trains all models
    d. Reproduces Spain ranking
    e. Runs Cross-regional generalisation
    f. Saves plots and tables

References

[1] W. B. Group, “World Bank Open Data.” [Online]. Available: https://data.worldbank.org/