National College of Ireland

Configuration Manual

MSc Research Project
Master of Science in Data Analytics Information

Cristian Andres Gajardo Montecino
Student ID: 24122246

School of Computing
National College of Ireland

Supervisor: Sallar Khan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Cristian Andrés Gajardo Montecino ……

Student ID: 24122246 ……

Programme: Master of Science in Data Analytics **Year:** 2024…….. Information

Module: MSc Research Project……

Lecturer: Sallar Khan……

Submission Due Date: 15/09/2025.....

Project Title: Evaluating The Predictive Performance Of LSTM Models On Chilean Pension Funds: A Comparative Study Using The S&P 500 Index

Word Count: 460…… **Page Count: 3**……

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: Cristian Gajardo Montecino ……

Date: 09/08/2025……

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

Cristian Andrés Gajardo Montecino
Student ID: 24122246

1 Section 1: Development Environment and Configuration Hardware

The operating system installed on this system runs Windows 11.

- Processor: Intel Core i7
- RAM: 16 GB
- Storage: ≥ 50 GB free space

Software

- The programming language Python 3.10 serves as the primary tool for this project.
- Development Environment: Jupyter Notebook

Libraries Used

- The financial market data downloading library is named yfinance.
- pandas – data manipulation and transformation
- requests – HTTP requests for API and web data retrieval
- time – timing and delays in requests
- glob – file path matching and batch file handling
- matplotlib.pyplot – plotting and visualizing results
- sklearn.preprocessing.MinMaxScaler – normalization of features
- tensorflow – deep learning framework
- tensorflow.keras.models.Model – model definition
- tensorflow.keras.layers → Input, LSTM, Dense, Layer, concatenate – neural network layers
- tensorflow.keras.callbacks.EarlyStopping – stopping training to prevent overfitting
- The evaluation metrics used in the project are from sklearn.metrics and include mean_absolute_error, mean_squared_error, and r2_score.

Key Configurations

- All models trained with a 60-day lookback window
- The train and test data allocation follows an 80:20 ratio.
- Batch size: 32, epochs: 100 with early stopping
- Loss function: Mean Squared Error (MSE)
- Optimizer: Adam

2 Section 2: Data Acquisition and Preprocessing

1. Data Acquisition
   - The data exists in CSV format through Cuota value & SP500.csv.
   - Fund A AFP Habitat data source: https://www.spensiones.cl/apps/valoresCuotaFondo/vcfAFP.php
   - S&P 500 Index: Downloaded using yfinance

2. Preprocessing Steps

   - Temporal Alignment: Synchronization of dates between both time series
   - The forward fill method was used to handle missing values in the dataset.
   - Normalization: Min-Max scaling to range [0, 1]
   - Date Conversion: Index converted to datetime format

LSTM Input Structure:
   - Multivariate LSTM → sequence with 2 features (Fondo A and S&P 500)
   - Attention-Based LSTM → same as multivariate but with attention mechanism
   - Multi-Input LSTM → two independent branches, each processing 1 feature sequence separately

3 Section 3: Model Training, Execution, and Evaluation

1. Implementation
   - Execute Jupyter Notebooks in the following order:
   - Data loading and preprocessing
   - Model architecture definition (Multivariate, Attention-Based, Multi-Input)
   - Model training

2. Training Configuration

   - 80/20 train-test split
   - Early stopping enabled to prevent overfitting
   - The trained models are stored in ".h5" or ".keras" format for later use.

3. Evaluation

Metrics Used:
   - Root Mean Squared Error (RMSE)
   - Mean Absolute Error (MAE)
   - R^2 (Coefficient of Determination)

Results Obtained:

   - Multivariate LSTM → RMSE: 945.02, MAE: 753.05, R^2: 0.9805
   - Attention-Based LSTM → RMSE: 1409.25, MAE: 1056.44, R^2: 0.9567

- Multi-Input LSTM → RMSE: 909.62, MAE: 684.22, R^2: 0.9820 Your third section. Change the header and label to something appropriate.

Source Code Repository:

- All source code for this project is available at: https://github.com/CrisGM87/LSTM-Models

Link Video Presentation

- https://www.youtube.com/watch?v=2KvWF_-CZCA

4 Section 4: Interpretation

The Multi-Input LSTM reached its best results through the independent processing of local and global inputs before their combination. The Attention-Based LSTM provided better interpretability yet failed to achieve superior results than the Multi-Input LSTM for accuracy.