

# Configuration Manual

MSc Research Project

Predicting the Bid-Ask Spread of Equity Options: A  
Machine Learning Approach Applied to Amazon and AMD

Pablo José González Pardo

Student ID: x24120511

School of Computing  
National College of Ireland

Supervisor: Brian Byrne

**National College of Ireland**  
**MSc Project Submission Sheet**  
**School of Computing**



**Student Name:** Pablo José González Pardo  
**Student ID:** x24120511  
**Programme:** MSc in FinTech **Year:** 2024/2025  
**Module:** MSc Research Project  
**Lecturer:** Brian Byrne  
**Submission Due Date:** 11<sup>th</sup> August 2025  
**Project Title:** Predicting the Bid-Ask Spread of Equity Options: A Machine Learning Approach Applied to AMZN and AMD  
**Word Count:** 362 **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:** Pablo José González Pardo

**Date:** 10<sup>th</sup> of August 2025

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
<b>Attach a Moodle submission receipt of the online project submission,</b> to each project (including multiple copies).	<input type="checkbox"/>
<b>You must ensure that you retain a HARD COPY of the project,</b> both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

<b>Office Use Only</b>	
Signature:	
Date:	
Penalty Applied (if applicable):	

# Configuration Manual

Pablo José González Pardo  
Student ID: x24120511

## 1 General Information

This configuration manual outlines the technical environment, tools, and libraries used to develop and execute the research project titled: “*Predicting the Bid-Ask Spread of Equity Options: A Machine Learning Approach Applied to Amazon and AMD*”. This manual aims to provide sufficient detail to replicate the experimental setup. The structure of the project follows a typical data science workflow involving data collection, preprocessing, model training, evaluation, and interpretation.

## 2 Operating System

- **OS:** Windows 11 Pro-64-bit
- **Processor:** Intel ® Core™ i7
- **Python version:** 3.10
- **Platform:** Google Colab (hosted Jupyter notebook environment by Google)

## 3 Data

Two datasets were imported from CSV files previously downloaded via yfinance. The first one contains AMZN option data (calls and puts) collected over seven consecutive market days (from June 4, 2025, to June 12, 2025.) and combined into a single dataset after retrieving the data for each individual day.

The second dataset contains AMD option data of one single market day (June 16, 2025).

## 4 Coding

### 4.1 Libraries used

The following Python packages were installed using pip within the Colab notebook:

- **yfinance:** Financial data retrieval from Yahoo Finance API
- **pandas, numpy:** Data loading, filtering and feature engineering
- **matplotlib, seaborn, plotly:** Visualizations, heatmaps and scatter plots
- **scikit-learn:** Data splitting, preprocessing, linear models, evaluation
- **catboost:** Non-linear gradient boosting model
- **shap, lime:** Model explainability and interpretability

## 4.2 Data Collection and Preprocessing

CSV files were manually downloaded via yfinance and uploaded to Colab. Preprocessing steps included:

- Filtering for call options
- Removing missing and extreme values
- Engineering variables such as:
  - 1) **Moneyness** = Underlying Price / Strike Price
  - 2) **Relative Spread** = (Ask – Bid) / Mid
  - 3) **Dummy variable for In-The-Money (ITM): Converted into integer**

The resulting datasets were used to train and evaluate all models.

## 5 Reproducibility

To replicate this Project is necessary to follow the steps:

1. **Open Google Colab:**  
[https://colab.research.google.com/drive/18j2J6L\\_Jco9QzlkG4Qbl06P\\_Z0n8yeWM?usp=sharing](https://colab.research.google.com/drive/18j2J6L_Jco9QzlkG4Qbl06P_Z0n8yeWM?usp=sharing)
  2. **Upload the notebook and CSV files:**  
7days\_AMZN\_options and 1day\_AMD\_option\_data
  3. **Run all cells sequentially after installing the required libraries**
  4. **Set random\_state=42 where applicable to ensure consistent results**
- No external APIs beyond yfinance were used.

## References

Hull, J. (2022). *Machine Learning in Business and Finance*. Wiley. (no date).