

MScFE 642 DEEP LEARNING FOR FINANCE

Group Work Project # 3

[See grading rubric here.](#)

Tasks

Step 1

- a. Gather information on the time series of the prices of any security of your choice: equity, cryptocurrencies, options, bonds, volatilities... Be sure to include a graphical and textual description of the data used.
- b. You will build a predictive model to forecast the returns (or volatilities or yield changes, etc.) of the security you have chosen. Build the labels of the model so that it is apparent that there exists some leakage of information between any choice of the training and test samples in the predictive model.
- c. Use three DL models to predict the time series, using a single train/test split. The three models should be:
 - a. 1) an MLP,
 - b. 2) an LSTM, and
 - c. 3) a CNN based on GAF.
- d. Provide information of the results from backtests of the trading strategies that arise from each of the three models.

Note: The time series should not exceed **2,000 observations**. You can set the model as regression or classification problems (or both).

Groups of 2 members: In Step 1.c, choose 2 models as follows:

- a. 1) either MLP or LSTM, and
- b. 2) a CNN based on GAF.

Step 2

Now you will backtest the model using a walk forward method.

- a. Train your models using a (non-anchored) walk forward method that exploits a train/test split with 500 observations in each set.
- b. Train your models using a (non-anchored) walk forward method that exploits a train/test split with 500 observations in each training set and 100 observations in the test sets.
- c. Discuss how the results of the backtests in parts a and b change with respect to those in Step 1.
- d. Discuss how the results of the backtests compare between parts a and b. Can backtest overfitting due to leakage explain the results?

Groups of 2 members: Answer the same questions for the 2 models you chose.

Step 3

Now you are going to repeat the same process as in Step 2 but alleviating the leakage of information between training and test sets.

- a. Set up and describe a method to reduce the extent of leakage between training and test samples.
- b. Using your method to reduce leakage, train your models using a (non-anchored) walk forward method that exploits a train/test split with 500 observations in each set.
- c. Using your method to reduce leakage, train your models using a (non-anchored) walk forward method that exploits a train/test split with 500 observations in each training set and 100 observations in the test sets.
- d. Discuss how the results of the backtests compare between parts b and c. Has overfitting apparently disappeared, as compared to Step 3?

Groups of 2 members: Answer the same questions for the 2 models you chose.

Step 4

As a team, the group will organize both the Python file for all the steps as well as work together on a report that contains the answers to all the questions above (except for the code) in a clear and organized manner (think about how you would present this to your boss). Be sure that the question number appears in the Python notebook before your response. For example, "Step 3a. We describe a method ..."

Submission Requirements and Format

One team member submits the following on behalf of the entire group:

- 1 PDF document containing ONLY the answers to the questions, EXCLUDING code.
- 1 zip file that contains:
 - Jupyter notebook that is executable
 - html of both the Jupyter notebook containing all code, results, and graphs

****Use Google Colab or GitHub to collaborate** in completing the executable Python program.*

NOTE: *The PDF must be uploaded **separately** from the zipped folder that includes any other types of files. This allows Turnitin to generate a similarity report.*

Rubric

Your instructor will evaluate your group submission for GWP1 using the following rubric:

Quantitative Analysis (Open-Ended Questions)	Technical and Non-Technical Reports	Writing and Formatting
40 Points	30 Points	20 Points
<p>The group is able to apply results, formulas, and their knowledge of theory to real-life finance scenarios by doing the following:</p> <ul style="list-style-type: none"> • Providing all the necessary information to support their arguments. • Presenting arguments that reflect group discussion and research. • Using authoritative references to support a position and provide updated information. • Concluding with practical takeaways for more insightful financial decision-making. 	<p>Technical reports contain 3 parts:</p> <ol style="list-style-type: none"> 1) code for each question (be sure to explicitly state the question number), 2) the corresponding output of that code, and 3) <u>interpretations and/or recommended courses of action that reasonably follow from those results.</u> <p>Note: Technical reports will include the technicalities of models, such as names, methods of estimation, parameter values, etc., and exclude generalities about the work done. It should NOT include names of Python code that were used.</p>	<p>A submission that looks professional should:</p> <ul style="list-style-type: none"> • Include the axes, labels, and scales in graphs. • Be free of significant grammatical errors or typos. • Be an organized, well-structured, and easy-to-read document. • Include proper citations and a bibliography in MLA format.
	<p>Non-technical reports contain 3 parts:</p> <ol style="list-style-type: none"> 1) clear explanation of results; 2) the recommended course of action that follows; and 3) the identification of factors that impact each portfolio. <p>Note: AVOID all references to model names, algorithms, and unnecessary details. Instead, focus on the investment decision.</p>	