MScFE 600 FINANCIAL DATA

Group Work Project # 2

See grading rubric here.

Below you will find two separate parts and each one includes a set of questions that you are required to answer in words, using code, or both words and code. The 'Submission requirements and format' section on Page 4 provides a detailed explanation on how to complete and submit the assignment through the online platform.

Groups of 3 members must complete both Parts 1 and 2.

Groups of 2 members or 1 member will complete Part 1 but NOT Part 2.

Part 1. Assessing Models with Alternative Data

Read the paper: "*An Intelligent Approach for Predicting Stock Market Movements in Emerging Markets Using Optimized Technical Indicators and Neural Networks*" by Sagaceta Mejia et al.:

https://www.degruyter.com/document/doi/10.1515/econ-2022-0073/html?srsltid=AfmBOorXs_wOzum9iij5b6EpD-oo_3Dd66uktp0de9a6WA5k8002IveF

Below are five questions from the paper designed to test students' understanding of the data, methodology, financial problem, application, and evaluation in the paper. Answer the 5 questions as detailed in the lists below:

Q1. Data Understanding:

- What types of data are used in the paper to predict stock market movements, and how are technical indicators derived from this data?
- Discuss the importance of using such indicators in forecasting stock price trends.

Q2. Security Understanding

- Pick one of the 3 funds (ECH, EQZ, or IVV). Write a 1-page (strict limit!) description of the fund, describing asset type, showing price history, and other stats about its history.

- Why do the authors decide to run a classification problem rather than a regression problem? Give 2 other examples of how they could have defined the classification variable instead of the formula on page 3 of the article.

Q3. Methodology Understanding

- Separate the 2nd section (2 Materials and Methods) by writing a new section 2 called Data. What are the subcategories of this section? (For example, Data Processing should be one. What are the others?)
- Call Section 3 Methodology. What are the subcategories of this section? Hint: One should be LASSO. What are the others?
- How would you divide descriptive statistics from models? (Hint, think about Pearson correlation versus LASSO).
- Outline the new section 3 with subcategories. Explain the optimization process of technical indicators used in the paper. How do the authors improve the predictive power of these indicators, and why is it important to optimize them for the neural network model?

Q4. Feature Understanding

- What does the paper consider a feature?
- How do you distinguish a feature from a method? From a model?
- What are the categories of features that you have learned?
- optimization process of technical indicators used in the paper. How do the authors improve the predictive power of these indicators, and why is it important to optimize them for the neural network model?

Q5. Optimization Understanding

- What is cross-validation in words?
- What is k-fold cross validation in words?
- What is the Jaccard distance>
- Compare the Jaccard distance to 2 of the distance metrics discussed in the lessons.
- How do the authors define an optimal solution?

Now that you answered the 5 questions, complete the following steps:

Step 1. Financial Problem:

- Section 3 of the paper discussed results and practical implications. What is the financial problem the authors aim to solve with their model?
- How does predicting stock market movements in emerging markets differ from predicting in developed markets, and why is this distinction significant for the model's design?

Step 2. Application:

- Describe the main takeaways of the results.
- What specific features seemed useful from the study?

Step 3. Replication

- Pick one of the funds.
- Download its data.
- Pick 1 of the easier metrics (e.g. Correlation, or Dispersion ratio are easier than LASSO).
- Implement the cross-fold validation or k-cross-fold validation.
- Reproduce the table
- Reproduce the graphs.

Part 2. Evaluating One Particular Type of Alternative Data

In the paper *"Alternative data in finance and business: emerging applications and theory analysis (review)"* by Sun at al.:

(https://jfin-swufe.springeropen.com/articles/10.1186/s40854-024-00652-0), there are 10 subcategories of alternative data listed (See Table 1 Alternative data types, features, users, and providers on page 7).

Pick 1 subcategory, and write a 5 page user guide that addresses

1) sources of data
2) types of data
3) quality of data
4) ethical issues
5) Python code to import and structure into useful data structures
6) exploratory data analysis of sample data
7) short literature search that links to papers citing research

Note: You do not have to read the entire paper, but reading general portions (e.g. first 7 pages), and portions specific to your selected subcategory will help.

Every Group Work Project Submission is composed by a report in PDF format and coding (see details in the 'Submissions requirements and format' below.

In the report, be sure to include a **list of references** and use them properly. This is mandatory for every GWP assignment you will submit in this Program.

Just compiling a list of references is not enough. **In your paper you also must show where each reference was used specifically (in-text citation)**.

Use the In-Text Citations and References Guide to learn how to add in-text citations and references.

In the Student Resource Center located in the dropdown menu under your name you find several other resources to complete your Group Work Projects properly:

- Academic Writing Guide,
- Anti-Plagiarism Guide,
- How to use LIRN, the online free library

Important note on the use of AI: Carefully read the WQU **Academic Policy on the use of AI** explaining how the use of AI tools is restricted and regulated. Severe penalties apply for excessive and improper use of AI.

Submission requirements and format

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

1. **1 PDF document*** with written answers from all the questions and steps as detailed above.
    a. Use the available Report Template and fill out the required information on group members on the first page
2. A **zipped folder** including:
    a. .ibynb executable Jupyter notebook**
    b. 1 PDF document **with the output** from the Jupyter notebook. To include the output, RUN the code before downloading the PDF.

** **Use Google Docs to collaborate**. Start by uploading the Report Template provided in the Course Overview. Once your report is completed, click File → Download → PDF Document (.pdf) to obtain the copy for your submission.*

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

*The PDF file with your report 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 using the following rubric:

Quantitative Analysis (open-ended questions) 60 Points	Technical and Non-technical Reports 40 Points	Writing and Formatting 20 Points
The group is able to apply results, formulas, and their knowledge of theory to real-life finance scenarios by doing the following: ● 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	Technical Reports contain 3 parts: 1) summary of key results; 2) interpretation of results; and 3) the recommended course of action that can reasonably follow from those results and interpretations. **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 the names of Python code that was used.	A submission that looks professional should include: ● The axes labels and scales in graphs. ● No significant grammar errors or typos. ● Organized, clear structure, and easy to read document. ● Proper citations and bibliography using MLA format.
	Non-technical Reports contain 3 parts: 1) clear explanation of results; 2) the recommended course of action that follows; and 3) the identification of factors that impact each portfolio. **Note**: AVOID all references to model names, algorithms, unnecessary details, and focus on the investment decision.	

Revised: January 21, 2025