For Report 2

Below is an in‐depth guide to structure your Final Group Project’s Second Report for FT5005. This report builds on your first report but requires additional analysis and detailed work in several new sections. The overall strategy is to revise and update Sections 1–3 for consistency and then expand with new sections addressing features engineering, methods, results, and conclusions.

### **1. Revising and Updating Sections 1–3**

Although you already submitted the First Report, it is important that the final report is fully consistent. You may choose to update these sections if you have made improvements or changes to your industry selection, research objective, or data preprocessing. Revisions could include:

* **Section 1: Overview** • Clearly state your target industry and restate your research objective, including a brief discussion of your additional dataset and the rationale behind your features engineering decisions.  
   • Provide updated supporting evidence (using academic papers, government statistics, and reputable analysts’ reports) to justify your chosen additional data or the approach you plan to take.
* **Section 2: Existing Literature** • Include at least one paper that supports the idea that your engineered feature(s) causally affect your target variable(s).  
   • Create or update the table listing similar forecasting studies, including details such as the best method used, performance metrics, and important features. Adding additional studies or results can help strengthen the literature review and show awareness of the domain.
* **Section 3: Data** • Describe the additional dataset you collected (or will collect) to engineer new features.  
   • Report summary statistics (mean, median, max, min) for numerical variables and give an overview for categorical variables after your preprocessing (before scaling).  
   • Optionally, include exploratory data analysis (charts, graphs) in an appendix to help illustrate your data characteristics and any potential issues.

*(See citeturn1file0 for detailed assignment instructions.)*

### **2. New Sections for the Second Report**

The final report has a page limit of 20 pages (excluding references) and is expected to contain more detailed analysis. The new sections are the core of your Second Report:

#### **Section 4: Features Engineering**

* **Objective:** Develop at least three new features that will boost your model’s predictive performance.
* **Requirements:** • **At least one feature based on literature review:** Use insights from your initial research to justify this feature. For example, if a paper suggests that sentiment extracted from earnings call transcripts predicts revenue, then use that to create a relevant variable.  
   • **At least one “advanced method” feature:** This could involve sophisticated text mining (such as topic modeling, custom-built sentiment dictionaries, or supervised learning approaches for feature extraction) or a complex financial ratio that isn’t just a simple calculation.  
   • Explain the domain knowledge behind each engineered feature, detailing why it is relevant and how you expect it to improve prediction.
* **Documentation:** Include explanations, formulas, or pseudocode as needed. Visualizations (e.g., scatter plots, correlation heatmaps) that demonstrate the relationship between your new features and the target variable(s) are highly encouraged.

#### **Section 5: Explain Your Methods Clearly**

* **Modeling and Techniques:** • Describe the **prediction methods** you have explored (e.g., OLS, ARIMA, or machine learning models such as random forests, gradient boosting, etc.).  
   • Discuss your rationale for selecting the final model.  
   • **Stacking:** You must implement at least one case of stacking. Explain how stacking is applied in your analysis and discuss whether it improved performance relative to individual models.
* **Reporting:** Include enough technical details so that someone reviewing your report can understand and, if necessary, replicate your analytical process. Highlight any preprocessing steps (scaling, feature selection, etc.) and how they integrate with your models.

#### **Section 6: Results**

* **Performance Evaluation:** • Present your main performance metrics clearly. For continuous outcomes (such as revenue or EBITA), R-square is the standard metric; if you predict a binary outcome (like direction of abnormal returns), ROC-AUC may be used.  
   • Start by establishing simple benchmarks:  
  + **Baseline 1:** Use lagged values (Y(t-1)) to predict Y(t) without any machine learning model.
  + **Benchmarking Models:** Compare your final best model with those from your first report (and any additional simple models, such as OLS).
* **Comparative Analysis:** • Compare performance with and without your newly engineered features.  
   • Compare cases with and without stacking. Optionally, investigate additional cases such as models with feature selection or dimension reduction to further support your performance claims.  
   • Use tables and figures to summarize your findings.

#### **Section 7: Discussions and Conclusion**

* **Interpretation and Insights:** • Summarize your main findings and discuss the impact of the new features and methods on the prediction performance.  
   • Reflect on what the comparative results say about your approach. For example, point out which features added the most value and whether stacking significantly improved your predictions.
* **Limitations and Future Work:** Discuss potential improvements, limitations in your data or feature engineering, and suggest directions for future enhancements.

### **3. Additional Tips for a Strong Final Report**

* **Integration and Consistency:** Ensure that all sections, including revised Sections 1–3 and the new Sections 4–7, form a coherent story. The narrative should smoothly transition from discussing the industry and data to presenting feature engineering, modeling efforts, results, and concluding remarks.
* **Visualizations and Tables:** Your report is encouraged to include informative visualizations and tables to illustrate: • The performance of different models and features. • Summary statistics of your dataset. • Comparative performance analyses.  
    
   These visual aids not only help in demonstrating your results but also make your report more engaging.
* **Code and Data Submission:** Remember that you must include: • A zipped file of the final version of your Python code and data files (in CSV format) for the prediction modeling section.  
   • A separate zipped file that contains your raw data and the code that leads to your final preprocessed input data.
* **Presentation Components:** Aside from the report, you need to prepare presentation slides and a recorded presentation. The presentation should: • Briefly introduce the analytics problem. • Outline your data source and the key dependent variable. • Summarize baseline and best-case model performance. • Demonstrate your improved prediction performance thanks to advanced features and methods.

*(For more details on specific deadlines, page limits, and submission formats, see citeturn1file0.)*

### **4. Summary of Submission Components**

* **Final Report Document:** • Updated Sections 1–3 and newly added Sections 4–7 (maximum 20 pages excluding references).
* **Code Submissions:** • Final version of Python code for prediction modeling.  
   • Data files in CSV format.  
   • Raw data and preprocessing code in an additional zip file.
* **Presentation:** • Slides (PDF or PPT).  
   • Recorded video of the presentation.

By following these guidelines and clearly articulating the purpose and methodology behind your feature engineering, your final report will demonstrate both technical rigor and domain insight. This will showcase your ability to replicate and enhance financial forecasting using machine learning techniques while aligning with the project’s learning objectives.

If your team needs clarification on any specific aspect—whether it’s the structure of a particular section, the type of visualizations to include, or details on implementing advanced feature engineering techniques—please feel free to ask for further guidance!