

Final Project Description

Abstract

This document provides guidelines for the final project of Financial Data Mining (Spring 2025). Any project-related correspondence should be directed to zijguo@stat.rutgers.edu (Instructor) and zw425@stat.rutgers.edu, mz666@stat.rutgers.edu (TAs), ensuring all team members are included in the CC field.

1 Project Overview

The goal of this course is to integrate heuristics, algorithms, and theoretical learning methods for various applications. The project allows you to apply learned concepts, engage with real-world applications, and potentially develop novel ideas, methods, and algorithms.

Key project details:

- **Group Size:** Projects must be completed in teams of two students. If you have difficulty finding a partner, please contact us.
- **New Projects:** Projects for FSRM 588 must be **distinct** from projects previously completed in other courses. Recycling projects from other FSRM classes or statistics courses is not allowed.

2 Important Dates

- **Group Formation:** Due on February 21 at noon.
- **First Proposal:** Due on March 14 at noon. Upload the proposal to Canvas as an assignment with the filename FSRM588: Group-(your group number)-Proposal-1.pdf". Include the full names of all team members. Only one submission per team is required.
- **Second Proposal:** Due on April 11 at noon. Upload the proposal with the filename FSRM588: Group-(your group number)-Proposal-2.pdf". Include all team members' names and your R code. Only one submission per team is required.

- **Presentation:** Scheduled for May 9 or May 12, 9:00 AM–2:30 PM.
- **Final Write-up:** Due on May 20 at noon. Upload to Canvas with the filename “FSRM588: Group-(your group number)-Final.pdf”. Include all team members’ names and your R code. Only one submission per team is required.

3 Project Submission Guidelines

- **First Proposal:** Maximum 3 pages, outlining your plan and preliminary results.
- **Second Proposal:** Should be about 90% complete. Maximum 4 pages including figures, tables, and references. Describe your research problem, methodology, progress, and future plans. An abstract summarizing your research and findings should be included. This should resemble a draft of your final report.
- **Presentation:** Teams will present their work in class. Details on presentation format and duration will be provided later. Reserve time for questions and feedback.
- **Final Write-up:** Should be a polished, improved version based on presentation feedback. Maximum 5 pages including figures, tables, and references. It should include:
 - An abstract understandable to a general audience.
 - Clearly stated models and tuning parameter selection.
 - A description of the estimation methods/algorithms used.

Reference research papers for guidance on report formatting.

4 Project Evaluation Criteria

Evaluation will consider the following aspects:

- **Research Problem:** The novelty and relevance of the research question.
- **Data Preparation:** Selection, cleaning, and processing of the dataset.
- **Methodology:** Appropriateness, justification, and implementation of chosen methods.
- **Application:** Preference for real-world applications with meaningful statistical insights.

- **Presentation:** Clarity, organization, and communication skills.
- **Write-up:** Quality of writing, organization, and presentation of results.

5 Project Topics

1. **Basic Tasks:** Identify an interesting tabular dataset and address a significant question through comprehensive analysis and comparison.
2. **Advanced Tasks:** Consider more complex challenges, including:
 - Processing *image* and *text* data, utilizing modern tools such as diffusion models and transformers.
 - Conducting an extensive pre-processing task for a real scientific application (e.g., protein structure analysis).
 - Implementing and testing advanced methodologies such as:
 - Domain Adaptation
 - Distributional Regression
 - Others (to be discussed with the instructor)