

Machine Learning – Assignment II

All about supervised learning, too

Overview

This assignment is based on the Kaggle competition, [Jane Street - Real-Time Market Data Forecasting](#). It aims to help you apply machine learning concepts to real-world financial prediction problems. You may choose one of the following two options:

1. **Participate in the competition and submit results.**
 2. **Conduct open-ended research using the competition data.**
-

Option 1: Participate in the Kaggle Competition

Objective

By participating in the Kaggle competition, you will tackle a real-world financial forecasting challenge, applying learned machine learning methods to develop effective predictive models. Additionally, you will gain experience in competitive modeling and iterative problem-solving.

Requirements

1. **Valid Submission**
 - You must complete **at least one valid submission** to the competition, which should earn a position on the leaderboard.
 - **Directly using public discussion notebooks is prohibited.** If you reference public solutions, you must explicitly document the differences and improvements in your approach.
2. **Model Development and Improvement**
 - Record your model development process, including the various methods, features, and parameter adjustments you tried.
 - Analyze each submission's results and use these insights to refine and improve your model.
3. **Report**
 - **Model Design and Evolution:** Document how your model evolved from the initial version to the final submission, highlighting challenges faced and solutions implemented.
 - **Results Analysis:** Compare the performance of different model iterations and explain the reasons for performance variations.
 - **Innovations:** Emphasize unique features or methods in your model and explain how they improve upon existing approaches.
 - **References:** Cite all sources and clarify your contributions when referencing other work.

Evaluation Criteria

- **Effectiveness and Originality (50%):** Your submissions must demonstrate effectiveness and originality, avoiding direct replication of others' work.
- **Model Development (30%):** Your process of model development and improvement should showcase deep thinking and consistent effort.

- **Report Quality (20%):** Reports should be well-structured, clear, and adhere to academic standards.
-

Option 2: Conduct Open-Ended Research

Objective

This track is designed to foster your research skills and creative thinking. Using the Kaggle competition dataset, you will define and explore a research problem of academic value and practical significance, proposing innovative solutions and conducting an in-depth analysis.

Selecting and Designing the Research Problem

1. **Choose a Challenging Problem**
 - **Reasoning and Purpose:** Your research problem should address real challenges in financial market forecasting, and you must explain why the problem is meaningful.
 - **Design Considerations:** Your problem design should take into account data characteristics, market-specific traits, and the applicability of machine learning methods.
 - **Expected Outcomes:** Clearly state your goals, such as improving prediction accuracy, identifying new features, or proposing a novel model.
2. **Integration and Expected Impact**
 - **Method Integration:** Explain how you will combine different data sources, features, or models, and the anticipated improvements from this integration.
 - **Results Expectations:** Specify what results you aim to achieve and how they will contribute to advancing financial forecasting.

Requirements

1. **Thorough Literature Review**
 - Investigate relevant academic papers and industry reports to understand existing methods and research gaps.
 - Cite relevant literature in your report and explain how your research addresses these gaps or improves upon current methods.
2. **Data Exploration and Processing**
 - Conduct a comprehensive exploratory data analysis (EDA) to understand data distribution, correlations, and patterns.
 - Handle issues such as missing values, anomalies, and noise to ensure data quality.
3. **Innovative Methods and Implementation**
 - **Model Design and Innovation:** Build on classroom knowledge, proposing improvements to existing models or introducing new algorithms.
 - **Feature Engineering:** Develop novel feature extraction or selection techniques to enhance predictive performance.
 - **Limitations: Avoid using deep learning methods.** Focus on innovations within traditional machine learning.
4. **Experimentation and Analysis**
 - Design rigorous experiments to ensure reliable and reproducible results.
 - Use appropriate evaluation metrics to objectively measure performance.
 - Analyze results deeply, explaining observed patterns and discussing model strengths and weaknesses.
5. **Report**
 - **Introduction:** Outline the research background, significance of the problem, and objectives.
 - **Methodology:** Detail your research approach, model design, and implementation specifics.
 - **Results:** Present experimental findings, supported by visualizations.

- **Discussion:** Analyze results, assess the model's effectiveness, and suggest areas for improvement.
- **Conclusion:** Summarize contributions and propose directions for future work.
- **References:** Include all citations following academic standards.

Evaluation Criteria

- **Problem Selection and Significance (20%):** How challenging and meaningful the research problem is.
 - **Methodological Innovation (30%):** The originality of the approach and its effectiveness in addressing the problem.
 - **Experimentation and Analysis (30%):** The rigor of the experiments, reliability of results, and depth of analysis.
 - **Report Quality (20%):** The clarity, structure, and adherence to academic standards in your report.
-

Tips for Success

- **For competition participants:**
 - **Originality is key:** Ensure your submissions reflect your own work. If referencing public resources, clearly explain your contributions and improvements.
 - **Process documentation:** Keep a detailed log of your experiments, adjustments, and insights.
 - **Leverage discussion forums:** Gain inspiration but translate ideas into your own implementations.
 - **For open-ended research participants:**
 - **Define a meaningful problem:** Choose a challenging problem that aligns with your interests and showcases your skills.
 - **Be creative:** Experiment with new features, algorithms, or data integrations.
 - **Set realistic expectations:** Think critically about the potential outcomes of your research.
-

Academic Integrity Statement

By submitting this assignment, you affirm that the work is your own and complies with the institution's academic integrity policies. Any form of plagiarism or misconduct will result in severe penalties.

Submission Requirements

- **Report:** Submit a PDF file named "StudentID_Name_Report.pdf."
- **Code:** Submit a zipped folder of your code files named "StudentID_Name_Code.zip."
- **Submission Method:** Upload all materials to the MOODLE system.