Technical Test: Data Scientist - Stock Forecasting

# Objective

This technical test evaluates a candidate's ability to understand, summarize, and apply research in the domain of time series forecasting, specifically stock price prediction. It assesses their data science skills, including literature review, data understanding, model implementation, and communication.

# Part 1: Paper Review and Presentation (50% of Total Score)

## Task

Candidates will be provided with a research paper (e.g., a recent publication from a reputable journal or conference) on stock forecasting. Their task is to:

1. **Read and Understand:** Thoroughly read and comprehend the provided research paper.
2. **Summarize Key Concepts:** Identify the core problem addressed, the proposed methodology, key findings, and limitations of the paper.
3. **Prepare a Presentation:** Create a short presentation (e.g., 10-15 slides, 15-20 minute presentation time) summarizing the paper. The presentation should be geared towards a non-expert but technically curious audience.

**Paper reference:** <https://drive.google.com/file/d/1Jylxu6xEUuN5Rda10rfhoWfmkHMYvUrp/view?usp=sharing>

## Evaluation Criteria

* **Clarity and Conciseness:** How well the candidate distills complex information into an easily understandable format.
* **Accuracy of Summary:** Correct interpretation of the paper's content, methodology, and results.
* **Identification of Key Insights:** Ability to highlight the most important contributions and implications of the research.
* **Structure and Flow:** Logical organization of the presentation.
* **Communication Skills:** Effective verbal and visual communication during the presentation.

# Part 2: Code Implementation (50% of Total Score)

## Task

Based on the methodology described in the provided paper, candidates are required to implement a simplified version of the proposed stock forecasting model.

1. **Data:** Use publicly available stock data (e.g., from Yahoo Finance, Kaggle datasets) for a chosen stock (e.g., a well-known index like S&P 500 components, or a specific company).
2. **Model Implementation:** Implement the core logic of the model described in the paper. If the paper describes a very complex model, candidates should focus on implementing the most crucial components or a simplified variation that captures the essence of the approach. Clearly state any simplifications made.
3. **Experimentation:**
   * Train the model on historical data.
   * Evaluate its performance using appropriate time series metrics (e.g., MAPE, RMSE).
   * Compare the implemented model's performance against a simple baseline model (e.g., a Naive forecast, ARIMA).
4. **Code Quality and Documentation:**
   * Write clean, well-commented, and readable code.
   * Include a `README.md` file explaining:
     + Any dependencies.
     + Assumptions made.
     + A brief summary of the results and insights from the implementation.

## Expected Deliverables

* Jupyter Notebook (`.ipynb`) or Python script (`.py`) containing the code.
* `README.md` file.
* Any necessary data files (if not publicly accessible via code).

## Evaluation Criteria

* **Correctness of Implementation:** Accuracy in translating the paper's methodology into code.
* **Code Quality:** Readability, modularity, error handling, and adherence to best practices.
* **Data Handling:** Effective loading, preprocessing, and manipulation of time series data.
* **Model Evaluation:** Appropriate use of metrics and a clear understanding of results.
* **Problem Solving:** How the candidate addresses challenges or simplifications.
* **Documentation:** Clarity and completeness of the `README.md` file.

# Instructions for Candidates

* **Time Limit:** You will be given 7 days to complete both parts of the test.
* **Resources:** You may use any libraries, frameworks, and online resources. Please cite any significant external code snippets or ideas.
* **Submission:** Submit all deliverables (presentation file, code files, `README.md`) in a single compressed folder to [rheco.kusuma@mandirisekuritas.co.id](mailto:rheco.kusuma@mandirisekuritas.co.id) and CC to [handik.yuwono@mandirisekuritas.co.id](mailto:handik.yuwono@mandirisekuritas.co.id) and [michael.manurung@mandirisekuritas.co.id](mailto:michael.manurung@mandirisekuritas.co.id)
* **Additional Information:**
  + You can utilize <https://algotrading101.com/learn/yfinance-guide/> for getting data
  + Choose **any stock** (freely) that you find interesting for analysis.