Comparison of deep learning models for Stock Price Prediction

Jiarui Zha
j1zha@ucsd.edu

Mentor Name (Lily Weng)
lweng@ucsd.edu

1 Introduction

Stock price forecasting is a challenging task in financial markets, where volatility and randomness introduce high uncertainty. Traditional models like Long Short-Term Memory (LSTM) have been widely used for time series forecasting due to their ability to capture temporal dependencies. However, recent advancements in deep learning, such as transformer-based architectures and probabilistic forecasting models, have introduced more sophisticated methods like Lag-Llama and Time Series Transformer (TST), which can better handle long-range dependencies and capture uncertainty in predictions.

This study aims to compare the effectiveness of LSTM, Lag-Llama, and Time Series Transformer (TST) in predicting NVIDIA (NVDA) stock prices. We evaluate the models based on their ability to predict future stock prices and measure their performance using Root Mean Squared Error (RMSE) for deterministic forecasts and Continuous Ranked Probability Score (CRPS) for probabilistic forecasts. Through this comparison, we assess which model is more suitable for short-term stock price forecasting and uncertainty quantification.

2 Methods

The dataset consists of historical daily stock prices for NVIDIA (NVDA). It includes columns such as Open, High, Low, Close, and Volume, but for this study, we primarily focus on the closing price. The dataset spans several years and is preprocessed to remove missing values and standardize the format.

To understand the dataset's properties, we performed Exploratory Data Analysis (EDA). The stock price series exhibits strong trend and seasonality, where long-term upward trends are observed along with periodic fluctuations. A rolling-window moving average and autocorrelation function (ACF) analysis indicate that past values significantly influence future stock prices.

For model training, we split the dataset into training and testing. The data was scaled using MinMax normalization to improve convergence during training.

This study evaluates three models: LSTM, Lag-Llama, and Time Series Transformer (TST). Each model is trained using NVIDIA stock price data, with the goal of predicting future stock prices based on historical data. The prediction horizon is set to seven days, aligning with short-term stock forecasting needs.

LSTM is a recurrent neural network (RNN) architecture that effectively captures sequential dependencies in time series data. It takes a rolling window of 60 past time steps as input and predicts the next seven days. The LSTM model is trained using Mean Squared Error (MSE) loss and optimized with the Adam optimizer.

Lag-Llama is a transformer-based probabilistic forecasting model designed for time series data. Unlike LSTM, it does not require an explicit rolling window but instead leverages context lengths to model long-range dependencies. We experiment with context lengths of 32, 64, 128, and 256, observing how different historical window sizes affect the model's forecasting accuracy. The model generates 100 probabilistic samples per prediction, which are aggregated to compute prediction intervals and CRPS values.

The Time Series Transformer (TST) applies self-attention mechanisms to capture complex temporal dependencies. The transformer processes the entire sequence in parallel to enable scalability and efficiency. The TST model is trained with positional encodings and multiple attention layers, where the input consists of historical price data transformed into embeddings. The model is optimized using MSE loss and evaluated based on RMSE for deterministic accuracy.

To compare the models, we use the following metrics: Root Mean Squared Error (RMSE): Measures the standard deviation of residuals between predicted and actual values. Continuous Ranked Probability Score (CRPS): Evaluates the probabilistic performance by measuring how well the predicted probability distribution aligns with actual observations. Lower CRPS values indicate more reliable uncertainty estimation.

3 Results

The results indicate that the LSTM model, which produces deterministic forecasts, achieved an RMSE of 9.16, suggesting that its predictive accuracy is relatively low. In contrast, the zero-shot forecasting approach using Lag-Llama demonstrated varying levels of performance depending on the context length. The lowest CRPS was observed at 0.0363 when the context length was set to 64, indicating the most reliable probabilistic forecasts. A shorter context length, such as 32 and 64, generally outperformed longer ones like 128 and 256, where the CRPS values increased to 0.0545 and 0.0459. This suggests that shorter historical contexts are more effective in capturing patterns within financial time series, while longer contexts may introduce unnecessary noise and reduce forecast reliability.

4 Conclusion

In this study, we compared the performance of an LSTM model and the zero-shot forecasting approach using Lag-Llama for time series prediction on NVIDIA stock prices. Our results show that Lag-Llama outperforms LSTM in terms of probabilistic forecasting, as indicated by its lower CRPS values, while LSTM struggles with an RMSE of 9.16.

Among the different context lengths used in Lag-Llama, a context length of 64 yielded the best results. Longer context lengths (128 and 256) resulted in slightly worse CRPS values, indicating that excessive historical data may introduce noise rather than improving accuracy. This study highlights large language models for time series forecasting. Future work can explore hybrid approaches combining LLM-based forecasting with traditional deep learning models to further enhance predictive accuracy.

Appendix: Updated Project Proposal

Our project extends the evaluation of **Lag-Llama** by applying it to new datasets and conducting a comparative analysis with traditional time series forecasting models. The project is divided into two main phases:

Phase 1: Comprehensive Evaluation on Stock Market Data

- Run **Lag-Llama** on real-world **stock price datasets** (e.g., NVIDIA stock prices from Yahoo Finance).
- Evaluate **short-term volatility forecasting** and **long-term trend prediction** using **different context lengths**.
- Compare the **performance metrics** (CRPS, RMSE) of Lag-Llama across different hyperparameter settings.
- Develop **visualizations** such as **forecast uncertainty intervals**, **error distributions**, and **performance heatmaps**.

Phase 2: Benchmarking Against Traditional Models (LSTM & Transformer)

- Implement **LSTM-based** and **Transformer-based** models for stock price forecasting.
- Compare their performance against **Lag-Llama** in terms of **forecasting accuracy and uncertainty estimation**.
- Evaluate the impact of **context length** on performance across models.
- Conduct statistical analysis to determine whether **Lag-Llama provides significant advantages over traditional deep learning approaches**.

Primary Output

Deliverables

1. **Comprehensive Report:**
   - Performance analysis of **Lag-Llama** on stock data.
   - Comparison with **LSTM** and **Transformer-based models**.
   - Analysis of **context length impact** on different models.
   - Discussion of practical implications in financial forecasting.
2. **Visualizations:**
   - **Prediction vs. actual trends**, including uncertainty intervals.
   - **CRPS and RMSE trends** across different context lengths.
   - **Heatmaps illustrating lag importance and model performance**.

Justification for Success

Phase 1: Stock Market Forecasting

- **Stock price data** is readily available via **Yahoo Finance API**.
- **Lag-Llama** has demonstrated strong performance in zero-shot forecasting on other datasets, making stock data an ideal test case.
- Evaluation through **context length tuning** helps determine the model's optimal setup for financial applications.

Phase 2: Comparative Study

- **LSTMs and Transformers** are widely used in financial forecasting, making them suitable benchmarks for Lag-Llama.
- By comparing **CRPS, RMSE, and forecasting stability**, we can assess whether **Lag-Llama provides meaningful improvements** in financial prediction tasks.
- The results will provide insights into **whether large-scale foundation models for time series (like Lag-Llama) can outperform traditional deep learning models in finance**.

5 Contributions

I do this project all by myself.