TUS
Technological University of the Shannon: Midlands Midwest
Ollscoil Teicneolaíochta na Sionainne: Lár Tíre Iarthar Láir

Faculty of Engineering & Informatics

Finding the lost time by using hourly data in the Stock Market

Aqib Tareen[1], Mark Daly[2], Michael Russell[2]

[1] Final Year Student, Master of Science in Software Engineering
[2] Supervisor, Department of Computer and Software Engineering, Technological University of The Shannon

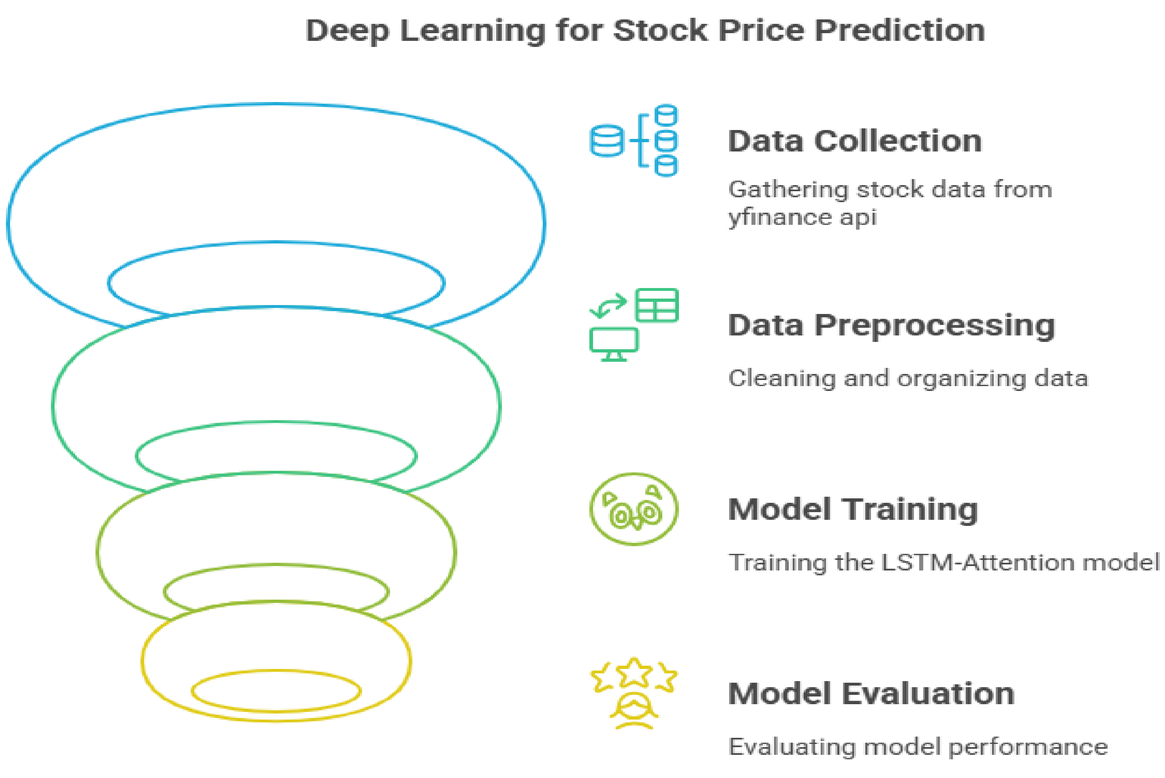
LEARN MORE

Abstract: Financial time series tend to be dynamically complicated at high frequencies. This research presents a model of intra-hour stock price "super-resolution," to forecast 2-minute price movements for any particular hour. The model takes historical hourly data as background, and it utilizes an LSTM-Attention framework. The explicit constraint mechanism is the distinctive feature that makes the predicted 2-minute sequence begin and terminate at the real observed 2-minute values of that particular hour.

Introduction: High-frequency trading and financial analysis can be based on high-resolution data at minute or even second frequencies. Such high-resolution historical data, however, are not always readily available as many datasets consist only of lower-frequency records, e.g., hourly or daily price records. This lack of data creates a growing need for methods that can effectively estimate price movements between these coarse observations.

This work bridges this gap by framing the problem as a super-resolution problem well adapted to intra-hour patterns in stock prices. The objective is to generate a natural series of 2-minute OHLC (Open, High, Low, Close) values for a given hour using only historical hourly data for context.

The proposed approach is based on a deep learning model recording intra-hour trends under strict boundary constraints having the predicted sequence begin and finish with the actual observed 2-minute values for the hour. Combining trend sensitive modeling with enforced consistency at anchor points, the method is designed to produce more realistic intra-hour price paths than the traditional interpolation approaches.



Methodology

Data Collection: Stock data at 1-hour and 2-minute intervals was downloaded daily using a Python script scheduled via a cron job on Google Cloud. Data was retrieved through the Yahoo Finance API and stored in separate directories based on resolution.

Preprocessing: Data was cleaned by removing duplicates and filling missing values using linear interpolation. A custom loader script organized and paired 1-hour data (for training) with corresponding 2-minute data (for evaluation).

Model Architecture: An LSTM-Attention model was used to predict intermediate 2-minute OHLC values within a given hour based on prior hourly context.

Encoder: Multi-layer LSTM processes hourly sequences.

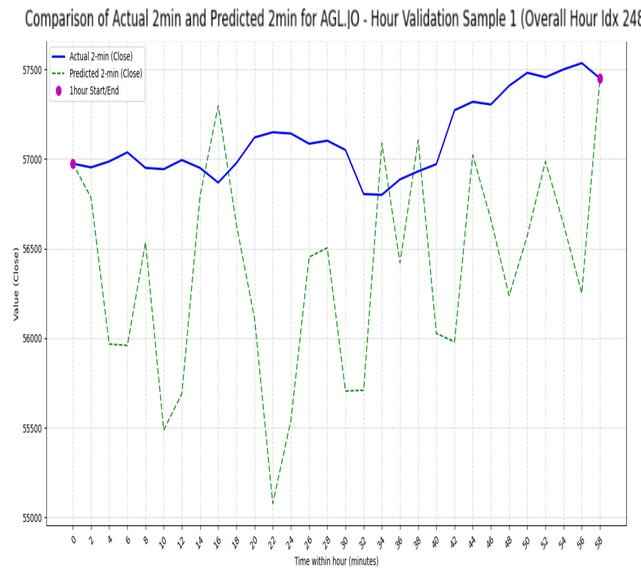
Attention: Multi-Head Attention focuses on relevant time steps.

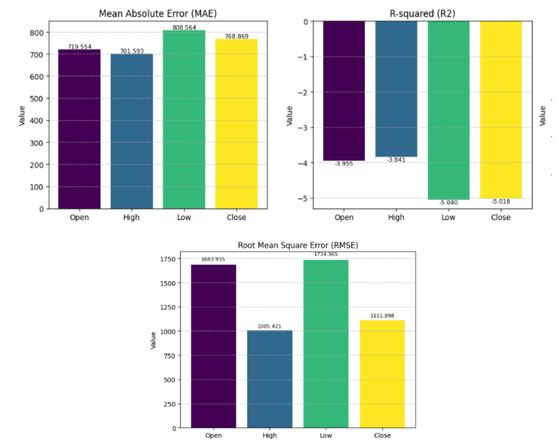
Decoder: Fully Connected layers reconstruct the intra-hour sequence. Only intermediate values are predicted; start and end 2-minute points are fixed using boundary constraints.

Training & Evaluation: The model was trained using MSE loss on intermediate points, optimized with Adam, and regulated using a learning rate scheduler. Performance was validated against actual 2-minute data using MAE, RMSE, R^2, and trend correlation.

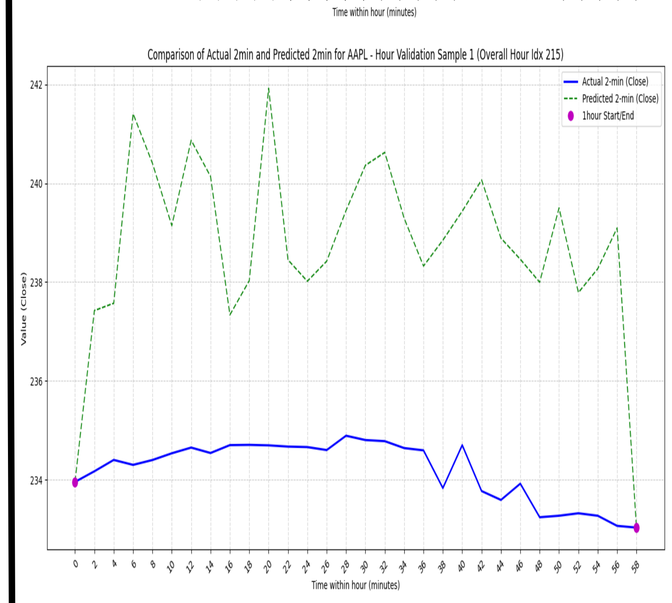
Research Question:- Can low-resolution stock data approximate high-resolution data be using historical trends in stock market?

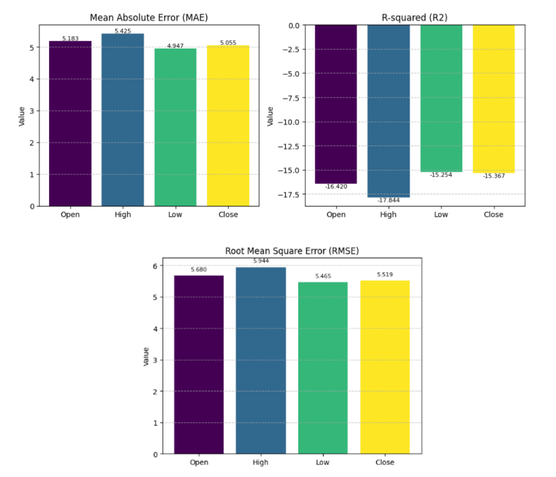
Results and Discussion: The LSTM-Attention model was applied to hourly OHLC data for stocks like AGL.JO and AAPL, aiming to generate intra-hour 2-minute close price sequences. In both figures the predicted paths (green dashed lines) correctly align with the enforced start and end points (magenta) but diverge notably from the actual 2-minute price movements (blue lines).While the predictions occasionally follow the overall direction, they often fail to capture fine-grained intra-hour dynamics and tend to overestimate volatility, producing unrealistic oscillations. The evaluation metrics, including negative R^2 and low trend correlation, confirm that the model underperforms relative to even a basic mean-based baseline. These results highlight the challenges of using only past hourly OHLC data to reconstruct high-frequency price patterns with accuracy.





Intra-Hour Prediction Model Performance Metrics





Intra-Hour Prediction Model Performance Metrics

Conclusion: This study explored the use of an LSTM-Attention model for generating 2-minute close price paths from hourly stock data. While the model successfully adheres to start and end constraints, ensuring alignment with known hourly values, its overall predictive performance was poor. Key metrics, including a negative R^2, indicate that the model fails to learn meaningful intra-hour dynamics. The output often exaggerates volatility and lacks consistency with actual trends, performing worse than a simple average-based baseline. This suggests that the current architecture and input features are insufficient for accurate high-frequency price reconstruction.

Future Work

1. Add More Features: Incorporate volume, volatility, or integrate sentiment for richer market signals.
2. Try New Architectures: Explore Transformers or diffusion models for better pattern learning.
3. Use More Historical Data: Extend training beyond the initial two months to improve generalization.
4. Flexible Context: Dynamically adjust input length based on volatility or market structure.