Deep Learning-Driven Predictive Modeling and Trading Strategy Optimization for AAPL

Introduction

Leveraging machine learning models for equity price prediction has become an inherent need of algorithmic trading strategies. This report details the development, implementation, and results of a sophisticated deep-learning model that predicts AAPL stock prices using 5-minute interval data. The objective was to use these predictions to identify profitable trading strategies and optimize options pricing, enhancing trading performance and decision-making.

Data Collection and Preprocessing

Data was fetched using the yFinance API, focusing on AAPL's 1-minute interval prices over a 7-day period. The 'Close' prices were extracted and reshaped into a format suitable for time series analysis. Normalization was performed using MinMaxScaler to scale the data from 0 to 1.

Given the interval data, a sequence length of 60 was chosen, approximately equivalent to one trading day. This sequence length allowed the model to capture intraday patterns and trends. The data was split into input sequences (x) and corresponding target values (y), which were then reshaped to fit the input requirements of an LSTM model.

Model Architecture

The core of this model is a Long Short-Term Memory (LSTM) network, which is well-suited for time series forecasting due to its ability to capture long-term dependencies. The model architecture includes:

- Two LSTM layers with 50 units each.
- Dropout layers with a dropout rate of 0.2 to prevent overfitting.
- A Dense output layer with a single unit to predict the closing price.

The model was compiled using the Adam optimizer with a learning rate 0.001 and a loss function of mean squared error (MSE). This setup was chosen to balance the need for efficient learning and accurate prediction.

Cross-Validation and Training

To validate the model, a TimeSeriesSplit cross-validator with 5 splits was employed. This method ensures that the model is tested on unseen data from different periods, thus mimicking real-world trading scenarios. Each fold involved training the model on a subset of the data and testing it on the subsequent subset.

To enhance training, early stopping and learning rate reduction callbacks were integrated. Early stopping monitors the validation loss and halts training when no improvement is observed for 10 epochs. In comparison, the learning rate reduces by a factor of 0.1 if the validation loss plateaus for 5 epochs.

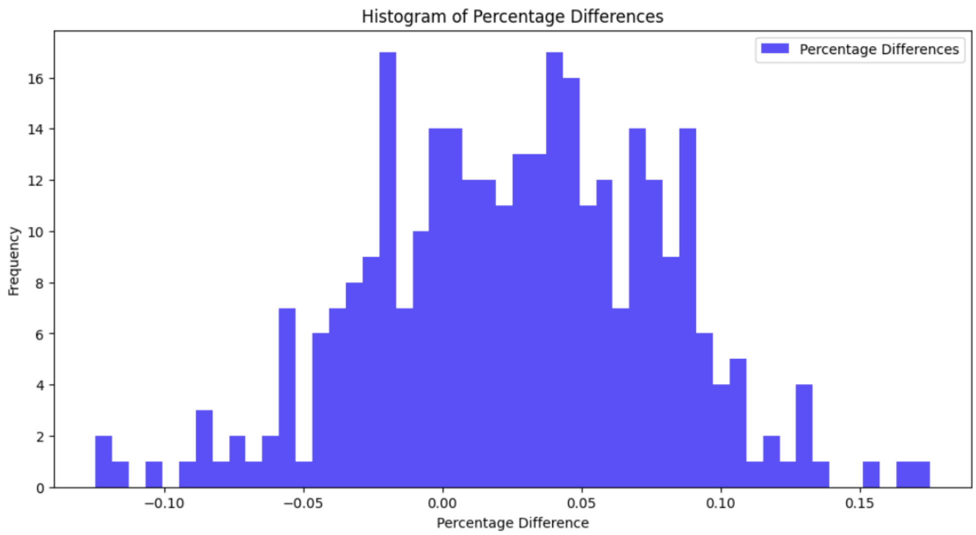
Model Performance and Evaluation

The model's performance was evaluated for each fold using Mean Squared Error (MSE) and Mean Absolute Error (MAE). The predictions were scaled back to original prices for comparison with actual prices. The results showed high precision of the model's prediction results.

- Overall Mean Squared Error (MSE): 0.091
- Overall Mean Absolute Error (MAE): 0.26
- Mean Difference: 0.0099
- Median Difference: 0.0075
- Max Difference: 1.45
- Min Difference: -1.045
- Mean Percentage Difference: 0.0048%
- Median Percentage Difference: 0.0030%
- Max Percentage Difference: 0.74%
- Min Percentage Difference: -0.55%

Analysis of Percentage Differences

A histogram of percentage differences between actual and predicted prices was generated to further evaluate the accuracy of the model. This visualization (see below) provides a frequency distribution of the percentage differences, illustrating that most predictions fall within a narrow range around zero. The concentration of data points near zero indicates high predictive accuracy, with few outliers demonstrating significant deviations.



Timestamps GMT	Actual Prices	Predicted Prices	Differences	Percentage Differences
2:30	189.4400024	189.47046	-0.030456543	-0.016077145
2:31	189.4400024	189.47574	-0.035736084	-0.018864064
2:32	189.4799957	189.47728	0.002716064	0.001433431
2:33	189.4349976	189.47765	-0.042648315	-0.02251343
2:34	189.4799957	189.4755	0.004501343	0.00237563

Options Pricing and Strategy Optimization

The predicted prices were then used to optimize options trading strategies. By applying binomial pricing method with Heston model and Implied volatility, Vertical Put Spread and Long Straddle, were evaluated. The Vertical Put Spread strategy, for instance, involved buying a put option at a lower strike price and selling one at a higher strike price, achieving a net premium of -4.0 and a maximum profit of 26.5. Based on the model's predictions, this strategy was identified as both the most profitable and cost-effective.

Conclusion

Implementing a deep learning model for predicting AAPL stock price has demonstrated significant potential in enhancing trading strategies and optimizing options pricing. Using LSTM networks to capture intraday price movements, combined with rigorous cross-validation, has resulted in a robust predictive model with high accuracy. Integrating these predictions into financial strategies has further showcased the practical applications of machine learning in trading.

This project highlights the importance of data preprocessing, model architecture design, and validation techniques in developing effective predictive models. The insights gained from this model can be extended to other stocks and financial instruments, paving the way for more advanced and profitable trading strategies in the future.

Future Work

Future enhancements could include exploring additional features such market sentiment data and tick level bid/ask prices to improve prediction accuracy.