



December 10th 2024



Price Forecasting using Machine Learning Analysis

Predicting Microsoft stock prices with
an LSTM model.

Alexander Cimini
ECNS 460

Research Question:



Can an LSTM accurately forecast Microsoft stock's next-day closing price?

Key Features:

- Utilized historical price data and market indicators.
- Model: Long Short-Term Memory (LSTM) Neural Network.
- Goal: Develop a reliable stock price forecasting model.

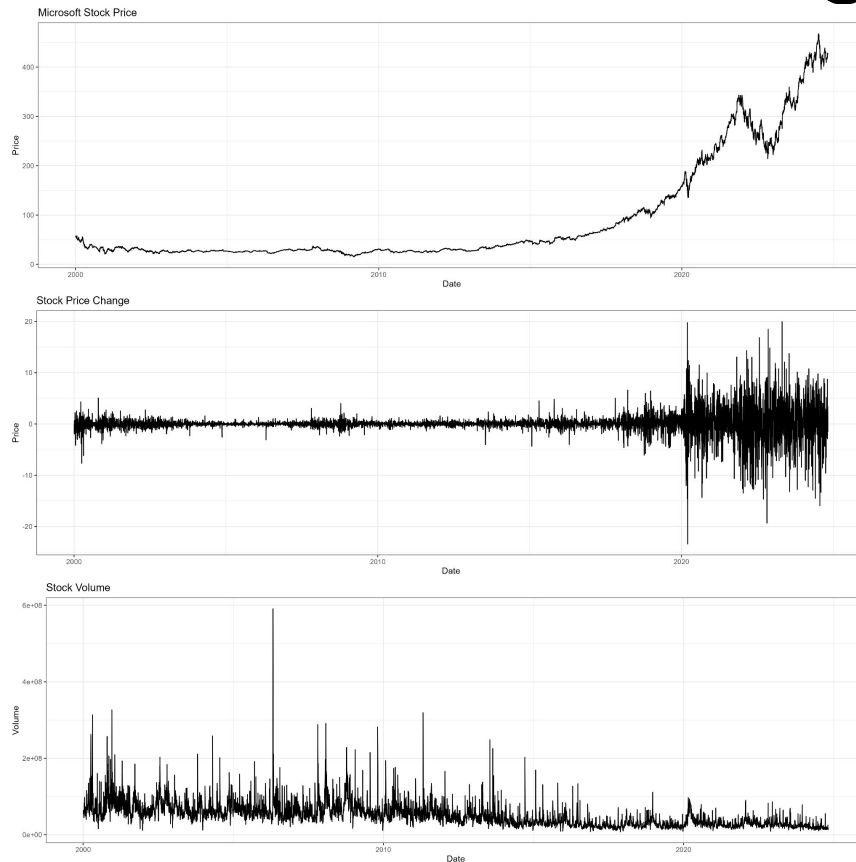
Data Overview

Timeframe: January 1, 2000 – October 10, 2024.

Variables included:

- Microsoft closing price, trading volume (Barchart).
- S&P 500 indicators (Alpaca).
- Volatility Index (VIX, Barchart).
- Corporate events (Nasdaq).

Why Microsoft? Microsoft was chosen because its consistently high trading volume helps reduce noise, making it easier to identify underlying trends in the data. Additionally, its long trading history provides a good dataset for training and testing the model.





Model Selection: LSTM – specialized for time series data

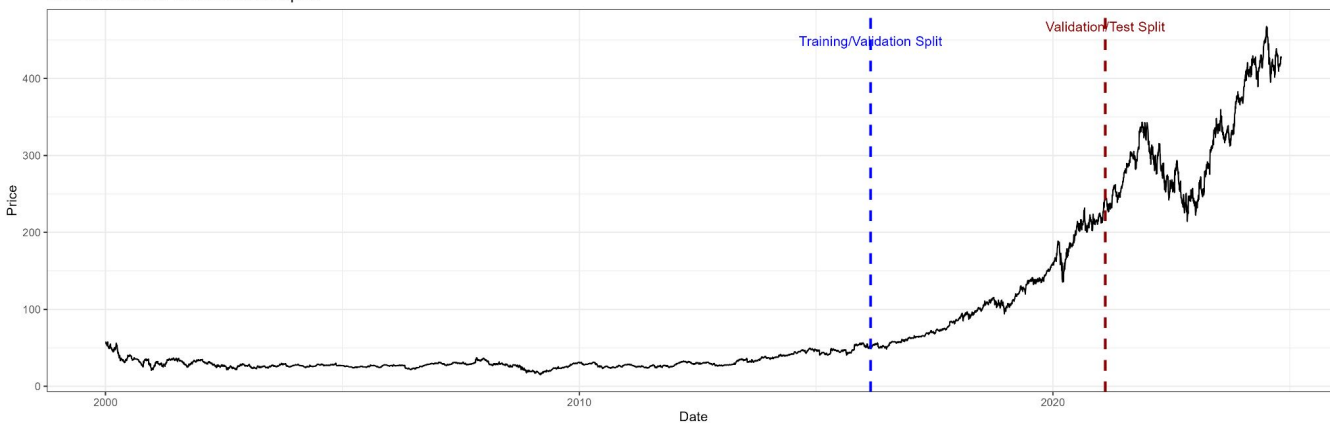
Data Preprocessing:

- Normalized variables for better convergence.
- Feature engineering: Percent changes, rolling averages.
- Data split: 65% training, 20% validation, 15% test.

Training Details:

- Optimizer: Adam.
- Loss Function: Mean Squared Error (MSE).
- Structure: Single-layer LSTM with 50 epochs (complete pass through the entire training dataset), batch size 32 (the number of samples processed before the model updates its parameters).

Microsoft Stock Price with Data Splits



Results

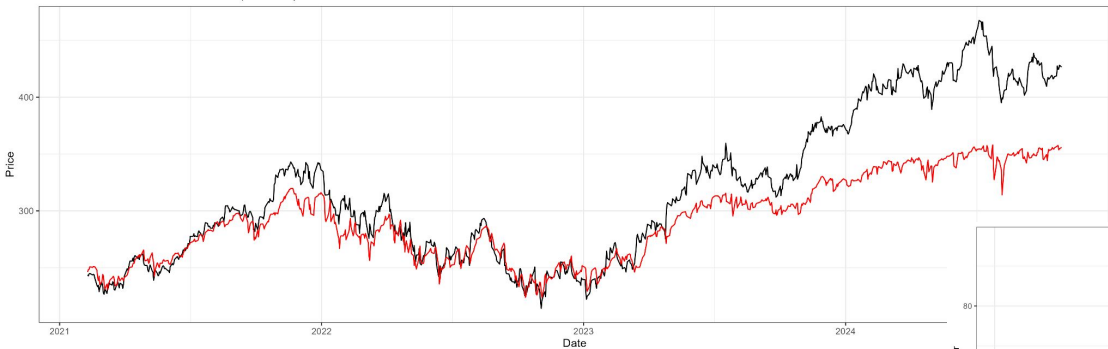


Validation MSE: 0.138 (balanced trade-off), without spikes over Epochs

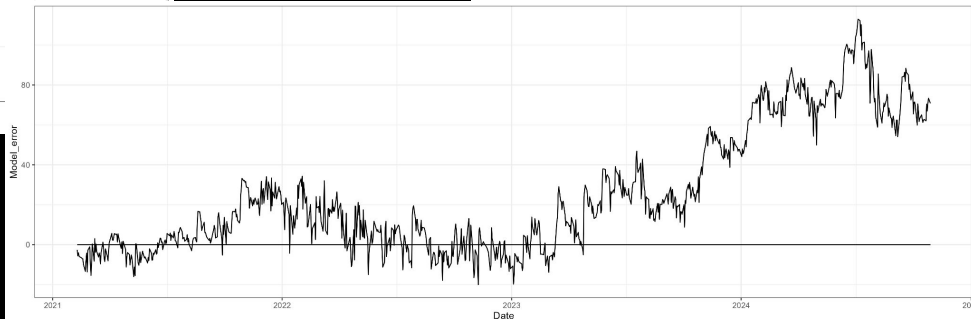
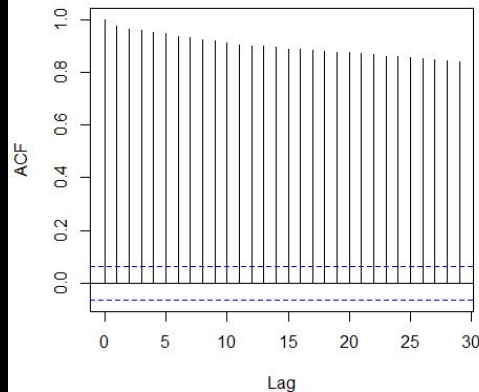
Test Performance:

- Predicted vs. actual stock prices.
- Error Analysis:
 - Errors closely follow stock price trends.
 - Autocorrelation in residuals indicates model issues.

Actual vs Predicted Stock Prices (Test Set)



Autocorrelation of Model Errors





Discussion

Challenges Identified:

- Errors are non-stationary and highly correlated.
- Indicates failure to fully capture patterns in the data.

Implications:

- Potential overfitting.
- Missing key trends or features.
- Changing market dynamics

Potential Steps to Enhance the Model:

- Gather more contextual data (economic/sector-specific indicators).
- Improve feature engineering (differencing, smoothing).
- Experiment with preprocessing methods for stationarity.



Conclusion

- Predicting stock prices remains a challenging task.
- LSTM showed potential but struggled with systematic issues.
- Insights gained will inform future modeling efforts.

Final Thought: Data quality and thoughtful preprocessing are as important as model choice



References

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

Allaire, J., & Chollet, F. (2023). keras: R interface to Keras. R package version 2.12.0. <https://keras.rstudio.com/>

Allaire, J., Tang, Y., & Ostblom, J. (2023). tensorflow: R interface to TensorFlow. R package version 2.12.0. <https://tensorflow.rstudio.com/>

GeeksforGeeks. (n.d.). Adam optimizer. GeeksforGeeks. <https://www.geeksforgeeks.org/adamoptimizer/>

GeeksforGeeks. (n.d.). Deep learning: Introduction to long short-term memory. GeeksforGeeks. <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>

Nasdaq. (n.d.). Microsoft Corporation (MSFT) press releases. Nasdaq. Retrieved December 3, 2024, from <https://www.nasdaq.com/market-activity/stocks/msft/press-releases>

Alpaca. (n.d.). Alpaca: Commission-free API-first stock brokerage. Retrieved December 3, 2024, from <https://alpaca.markets/>

Barchart. (n.d.). Market data APIs, feeds, and digital solutions. Retrieved December 3, 2024, from <https://www.barchart.com/>



Thank you
– Questions?