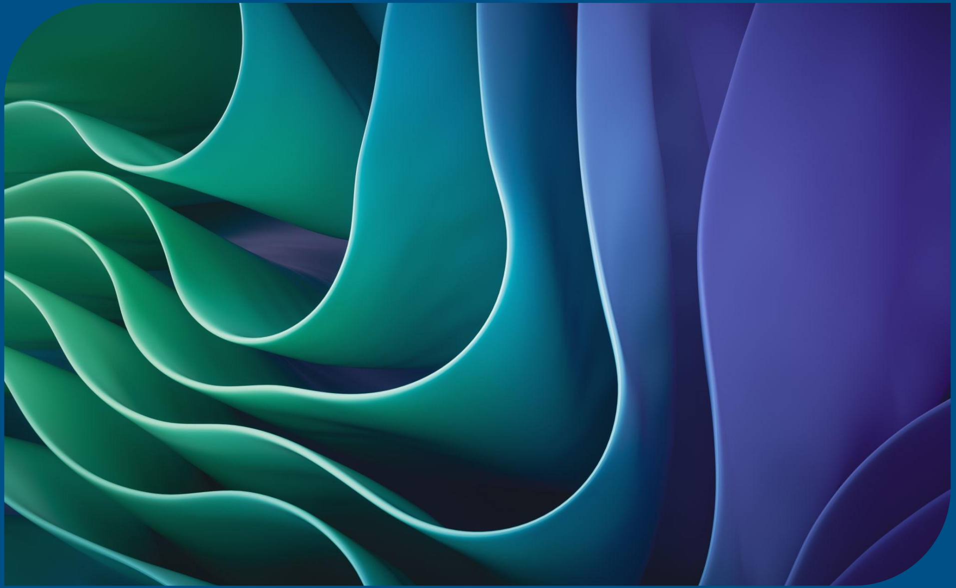
12.09.2024

Stock Market Price Prediction

By: Leo St Amour, Roshan Ravindran, Mohammad Heydari, and Pranesh Ambokar



1. Introduction
2. Related Works and Alg. Selection
3. Methodology
4. Results and Analysis
5. Conclusion and Future Work

1. Introduction
2. Related Works and Alg. Selection
3. Methodology
4. Results and Analysis
5. Conclusion and Future Work

Our Project's Focus

- **Our Objective**
  - Evaluate how well algorithms handle the complexities of financial data to inform better stock market forecasting
  - Compare/contrast two machine learning algorithms:
    - *Linear Regression*
    - *Long Short-Term Memory (LSTM)*

- **Important Distinction**
  - Only evaluating on a day to day basis
  - Given an opening price, what is the predicted closing price?

- **Real-World Implication**
  - Understand which algorithm provides more reliable predictions for short-term trading decisions

1. Introduction
2. **Related Works and Alg. Selection**
3. Methodology
4. Results and Analysis
5. Conclusion and Future Work

Linear Regression

Overview:

- A fundamental statistical technique modeling relationships between dependent and independent variables.
- Represents the relationship as a straight line: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_2 x_2 + \epsilon$

Advantages:

- Computationally efficient and interpretable.

Limitations:

- Struggles with capturing non-linear and dynamic trends in financial data.
- Performance deteriorates in volatile market conditions.

LSTM

Overview:

- LSTM networks, a subtype of RNNs, use memory cells to retain relevant information over extended time periods.
- Designed to model sequential data, capturing long-term dependencies and complex temporal patterns.

Advantages:

- Handles noisy and non-linear data effectively.
- Flexible integration of input features like technical indicators and macroeconomic variables.

Challenges:

- Computationally intensive and requires extensive training.
- Performance depends on effective feature engineering and hyperparameter tuning

Similar Studies

- **Overview**
  - Multiple studies confirmed our rationale for choosing these specific algorithms

- **Why Linear Regression?**
  - Emioma and Edeki (2021) successfully applied linear regression for short-term stock price prediction using factors like opening and closing prices

- **Why LSTM?**
  - Bhandari et al. (2022) demonstrated that LSTMs outperform traditional methods like ARIMA and linear regression for forecasting indices like S&P 500
  - Saboor et al. (2020): Achieved high accuracy in predicting indices like S&P 500 and NIFTY 50

Data and Preprocessing

100 Stocks

- 50 large cap
- 50 mid cap

Historical stock data

- Date
- Opening price
- Closing price
- Daily high
- Daily low

Partition into input and output

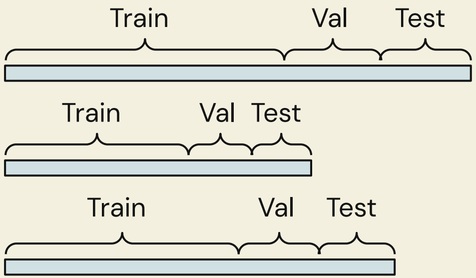
- Input: open, low, and high
- Output: close

Scaling

- Scale prices to a value between 0 and 1
- Improves model efficiency

Training/Validation/Testing Sets

- Training: 70%
- Validation: 15%
- Testing: 15%



- Aggregate training and validation data sets

Training and Validation

Training

- Train both models using aggregate training set

Validation

- Fine-tune LSTM hyper-parameters using aggregate validation set
- Number of units in LSTM layer: 100
- Recurrent dropout rate: 0.1
- Dropout layer rate: 0.3
- Learning rate: 0.005

Testing

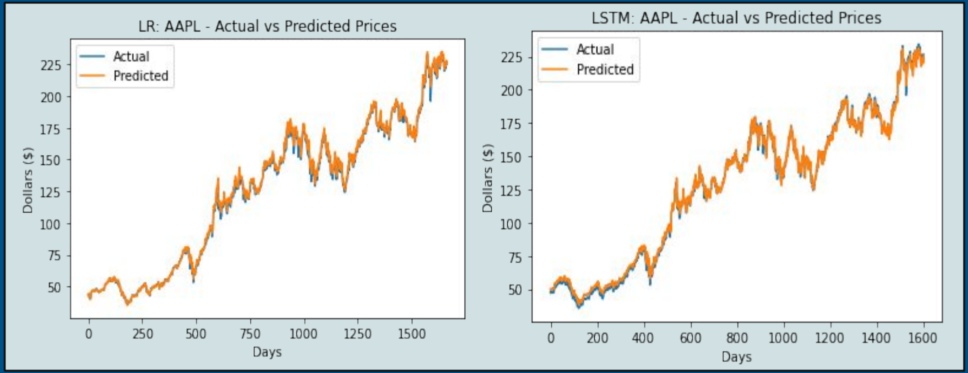
Test on testing sets for each individual stock

- Mean squared error
- Mean absolute error
- R^2 value
- Linear regression coefficients/correlations

Day trading simulation

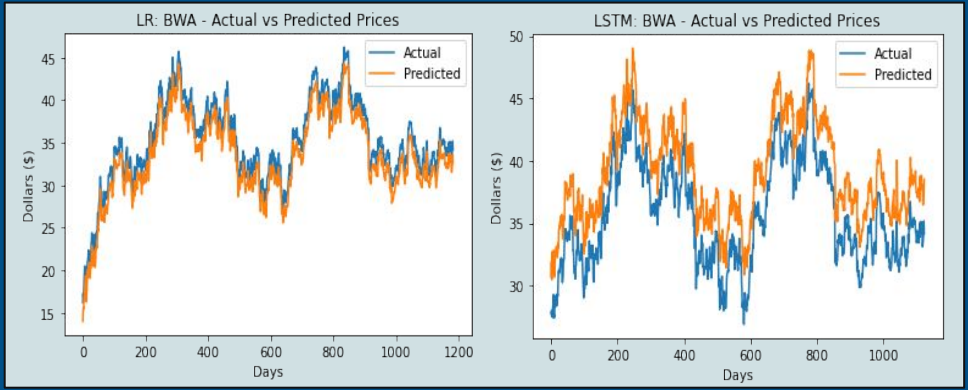
- Randomly select stocks
- For a random day, make a prediction about which stock would produce greatest yield
- Compare predicted to actual

Model Comparison



High Market Cap (AAPL):

- Both models perform well with clear trends
- Shows stability in predictions for established stocks



Mid Market Cap (BWA):

- More volatility and complex patterns
- LR shows surprisingly competitive performance despite simplicity

Performance Metrics

Metric (Average)	Linear Regression	LSTM
R squared	0.977	0.935
MSE	1.45e-8	3.89e-08
MAE	0.000076	0.000110
Training time	Less than 1 second	1.17 hours

- **Both models perform well.**
  - Linear regression performs slightly better.
- **Big trade-off on model complexity**
  - LSTM takes a significantly higher computation power.

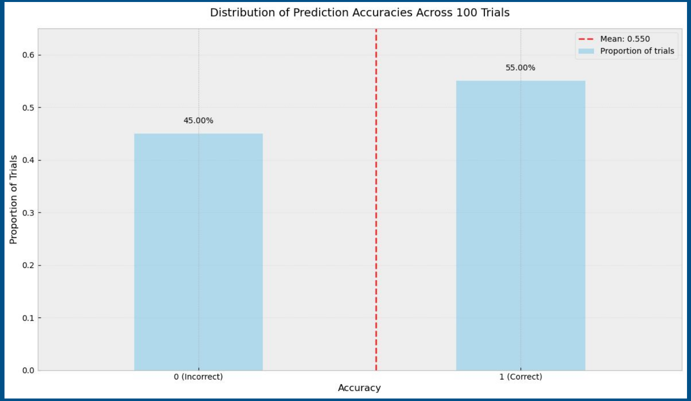
➔ Linear Regression: Simple, assumes linear relationships between features and target

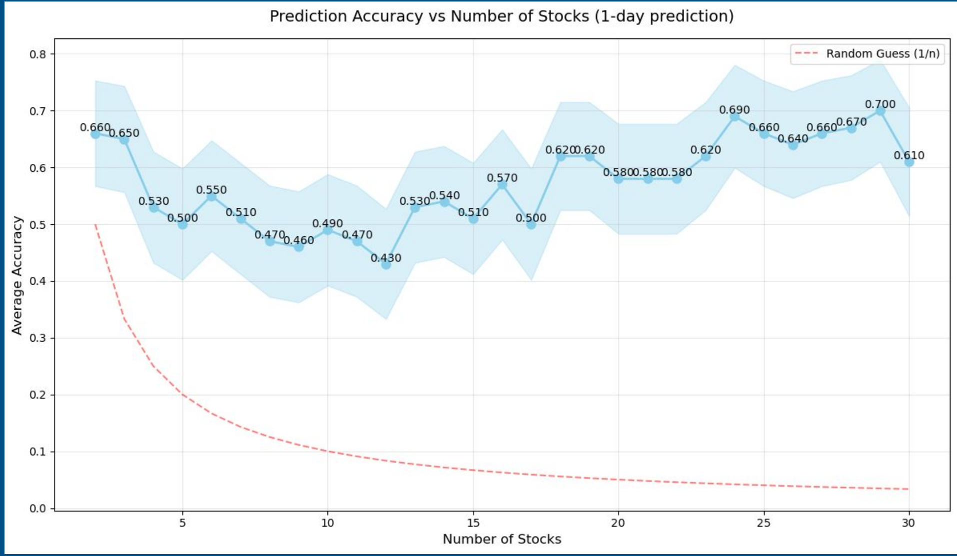
➔ LSTM: More complex, capable of learning sequential patterns and dependencies in time series data

Implication

- **Two Stocks Prediction Accuracy:**
  - Tested over 100 different trials
  - Success = Correctly identifying the stock with highest price increase
  - Model achieves 55.0% accuracy in picking best performer

- **Multiple Stock Comparison (100 trials):**
  - Compares predicted vs actual daily closing prices for N stocks (N = 2 to 30).
  - Particularly valuable for day traders comparing multiple options.
  - Performance stays stable choosing between a high number of stocks.





1. Introduction
2. Related Works and Alg. Selection
3. Methodology
4. Results and Analysis
5. **Conclusion and Future Work**

Conclusion

Both models performed well, showcasing their strengths in stock price prediction.

Linear Regression:

- Competitive performance, especially for high-market-cap stocks.
- Surprising edge in **stability**.
- Achieved a higher **R^2 value (0.977)** compared to LSTM.
- Significantly shorter training time (**less than 1 second**).

LSTM:

- Demonstrated the ability to learn **complex temporal patterns**.
- Achieved a strong **R^2 value (0.935)** despite the complexity.
- Training time was significantly longer (**1.17 hours**), reflecting its higher computational requirements.

Key Takeaway:

- Trade-offs between accuracy and complexity quantified.

Future Work

1. **Future Price Prediction**: Extend the model to predict the closing price of future dates, evaluating the algorithms' ability to handle longer-term forecasts and adapt to temporal patterns over multiple days.

2. **Dataset Expansion**: Increase the size and diversity of the dataset to evaluate the models' performance across a broader range of stocks and market conditions.

References

[1] Hum Nath Bhandari, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Ke shab R Dahal, and Rajendra KC Khatri. 2022. Predicting stock market index using LSTM. Machine Learning with Applications 9 (2022), 100320.
[2] Svetlana Borovkova and Ioannis Tsiamas. 2019. An ensemble of LSTM neural networks for high-frequency stock market classification. Forest (2019). Https: //doi.org/10.1002/for.2585 First published: 21 March 2019.
[3] Kai Chen, Yi Zhou, and Fangyan Dai. 2015. A LSTM-based method for stock returns prediction: A case study of China stock market. In 2015 IEEE International Conference on Big Data (Big Data). IEEE, XXX–XXX. https://doi.org/10.1109/
BigData.2015.7363946
[4] CC Emioma and SO Edeki. 2021. Stock price prediction using machine learning on least-squares linear regression basis. In Journal of Physics: Conference Series, Vol. 1734. IOP Publishing.
[5] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. 2017. LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems 28, 10 (2017), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924
[6] MEAG Hiransha, E Ab Gopalakrishnan, Vijay Krishna Menon, and KP Soman. 2018. NSE stock market prediction using deep-learning models. Procedia computer science 132 (2018), 1351–1362.
[7] Hani A.K. Ihlayyel, Nurfadhlina Mohd Sharef, Mohd Zakree Ahmed Nazri, and Azuraliza Abu bakar. 2018. An enhanced feature representation based on linear regression model for stock market prediction. Intelligent Data Analysis 22, 1 (2018), 45–76. https://doi.org/10.3233/IDA-163316
[8] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. 2023. Linear regression. In An introduction to statistical learning: With applications in python. Springer, 69–134.
[9] Rianchal Jha, Nitin Dixit, Rakhi Arora, Rishi Soni, Vijay Prakash Sharma, and Shreshtha Kinger. 2024. Predicting Stock Market Movements with Linear Regression and LSTM Machine Learning Model. In 2024 IEEE International Conference on Contemporary Computing and Communications (InC4). IEEE, Bangalore, India, https://doi.org/10.1109/InC460750.2024.10649281
[10] Mahinda Mailagaha Kumbure, Christoph Lohrmann, Pasi Luukka, and Jari Porras. 2022. Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications 197 (2022), 116659. https://doi.org/10.1016/j.eswa.2022.116659
[11] Oleh Onyshchak. 2020. Stock Market Dataset. https://www.kaggle.com/datasets/jacksoncrow/stock-market-dataset. Accessed October 29, 2024.
[12] Ogulcan E. Orsel and Sasha S. Yamada. 2022. Comparative Study of Machine Learning Models for Stock Price Prediction. arXiv preprint arXiv:2202.03156 (2022). https://doi.org/10.48550/arXiv.2202.03156