# Comparative Analysis of LSTM and LSTM with Multi-Head Attention for Stock Market Price Prediction

Ananth Karkala Pai(paia@iu.edu), Manishkumar Alagumalai(maalag@iu.edu)

**Abstract**

Stock market price prediction has always been a challenging problem, as many factors influence the movement of stock prices. Due to the advent of advanced deep-learning techniques predicting future stock price movements is an area of ongoing research and development. Advanced deep-learning techniques, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory Networks (LSTMs), and Convolutional Neural Networks (CNNs), have been employed to analyze historical stock price data and attempt to forecast future stock price movements. The study aims to assess the suitability of LSTM-based models for predicting future stock price movements. Additionally, this paper analyzes the effectiveness of the Multi-Head Attention layer in the LSTM model to predict future price trends, by comparing it against the LSTM model without the Attention layer. Finally, the study explores the impact of incorporating market conditions as auxiliary input on enhancing prediction accuracy.

## 1 Introduction

### 1.1 Background and Literature Review

In the constantly evolving world of financial markets, the pursuit of forecasting stock movements remains a challenge. While there are traditional ways to predict the trend, it has not always been successful and always comes down to a gamble, Not to mention the amount of data that needs to be understood to predict the flow. However, due to advances in the field of Data Analytics and Deep learning, a lot of new techniques have come up and are changing the landscape of how the data is being processed to get insights into making future decisions. RNN-based models serve diverse applications leveraging time-series data. For instance, Chung et al. (2014)[**1**] employed an RNN network to predict music sequences, using consecutive music segments as inputs to anticipate subsequent sequences. In a study by Zhang et al. (2018)[**2**], RNN-based models, particularly LSTM, were utilized for Short-Term Load Forecasting (STLF) in predicting electricity load values for forthcoming hours or days. Additionally, Bu et al. (2020)[**3**] explored the application of Multi-head attention RNN models to forecast residential energy consumption. Forecasting residential energy usage involves analyzing multiple sensor signals within specific intervals to extract features for an accurate consumption prediction model.

### 1.2 Objective

This study aims to explore and understand the complexities of stock market price prediction, focusing on the evaluation of LSTM-based models' ability to foresee future stock price movements. The primary objective is to thoroughly assess the suitability of LSTM architectures for their ability to capture long-term dependencies and to understand their performance in the context of stock market forecasting. Additionally, the study explores the performance of LSTM models both with and without the Attention layer, the research aims to understand and provide some insights on whether using attention layers with LSTM helps to forecast future stock movements.

Furthermore, we investigate whether using market conditions as an auxiliary input has any impact on changes to the model's prediction. This is usually performed since changes in market dynamics for stocks have a proportional cause and effect on the direction of the market trend.

In summary, the goal of this research is to advance our understanding of LSTM-based models in the context of stock market price prediction, assess the potential benefits of incorporating attention mechanisms, and investigate the impact of auxiliary input variables, thereby contributing valuable insights to the ongoing discourse on leveraging deep learning for financial forecasting.

# 2 Methodology

Recurrent Neural Networks (RNNs) have gained prominence for their effectiveness in handling sequence generation and time series data. Among the variants of RNNs, Long Short-Term Memory (LSTM) networks have demonstrated exceptional performance, particularly in the domain of time series data prediction **Y**. Wang et al. 2019 [**4**]. In this section, we will explore various LSTM-based model architectures employed for time series data prediction along with all the dataset preprocessing tasks and model evaluation metrics.

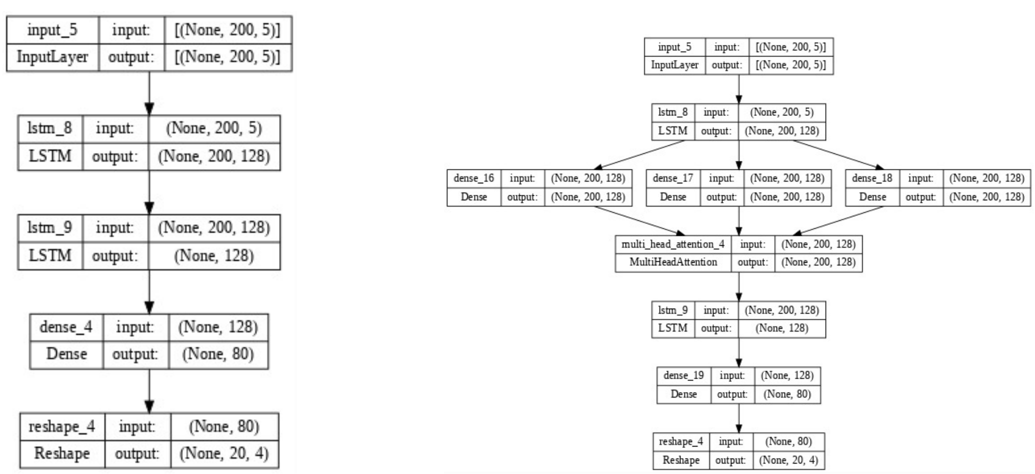
## 2.1 Model Architecture

To evaluate the suitability of LSTM models for time series data prediction, a vanilla LSTM model is developed which will be used as the baseline model Van Houdt et al. 2020 [**5**]. This vanilla LSTM baseline model will serve as the reference point, allowing us to establish a benchmark for performance comparison. Furthermore, to enhance the predictive accuracy of the baseline model, auxiliary inputs and Attention mechanisms will be incorporated into the computational framework. All the discussed Model Architecture will be using **Adam** Optimizer with a learning rate of **0.0002** and Mean Squared Error**(MSE)** loss function.

### 2.1.1 Vanilla LSTM Model Architecture

The Vanilla LSTM Architecture serves as the baseline model for the time series data prediction task. This model takes an input of shape **(200, 5)**, which are **200 days** of historical Open, Close, High, Low, and Volume values of a particular stock under consideration, and predicts future stock Open, Close, High, and Low values for **20 days**. The Vanilla LSTM implementation has two LSTM layers with 128 units in each layer. The tanh activation function is used for LSTM layers in the architecture. The subsequent dense layer has 80 units which will be then reshaped into **(20, 4)**.

This model, with a total of **210,512** trainable parameters, establishes our baseline for evaluating subsequent enhancements. The rationale behind this architecture lies in its ability to comprehend sequential information within the given time series data. The stacked LSTM layers capture the long-term dependencies, while the dense layer and reshaping operations contribute to the model's ability to distill relevant patterns for prediction. Figure [**1**]**(a)** shows the block diagram of Vanilla LSTM Architecture.




(a) Vanilla LSTM Model Architecture

(b) LSTM Model with Attention Layer Architecture

Figure 1: Model Architecture Block Diagram

### 2.1.2 LSTM Model with Attention Layer Architecture

To enhance the predictive capabilities of the Vanilla LSTM model Multihead Attention mechanism is introduced to the model. This novel iteration incorporates a Multi-Head Attention layer after the initial LSTM layer. The input sequence undergoes three distinct linear projections to serve as queries, keys, and values, respectively which will be input for the forthcoming Attention layer. The Multi-Head Attention layer, with eight heads and a key dimension of **128**, facilitates capturing intricate dependencies within the sequence. The resulting attention-weighted context is then processed by a subsequent LSTM layer. The model retains a similar output shape to that of the Vanilla implementation to maintain consistency with the task requirements. Figure [**1**](**b**) shows the block diagram of the model with the Attention Layer.

The LSTM model with the Multi-Head Attention layer has a total of **787,536** parameters, indicating a significant augmentation from the Vanilla LSTM model. This increment aligns with the additional complexity introduced by the attention mechanism. The attention layer's ability to selectively attend different parts of the input sequence, coupled with the LSTM's sequential processing capability, enables the model to effectively capture intricate temporal dependencies.

### 2.1.3 LSTM model with Attention Layer and Auxiliary Input Architecture

To further improve the model performance, market index data is introduced as auxiliary input along with stock price data. The input sequences, denoted as **inputs-5** and **inputs-6**, are concatenated along the temporal axis, resulting in an input shape of **(400, 5)** where 1st 200 data points belong to stock price history and subsequently 200 data points belong to Market index history for 200 days. This concatenated sequence undergoes LSTM processing, capturing intricate temporal dependencies. Simultaneously, the sequence is projected into query, key, and value representations, serving as inputs to a Multi-Head Attention layer with eight heads and a key dimension of **128**. Figure [**2**] shows the block diagram of the LSTM model with Attention Layer and Auxiliary input.

The LSTM model with both an attention layer and auxiliary input encompasses a total of **787,536** parameters, mirroring the complexity introduced by the attention mechanism in the previous model iteration. The concatenation of an auxiliary input aims to provide the model with additional contextual information, while the attention mechanism enriches the model's ability to focus on pertinent aspects of the input sequence dynamically presenting a holistic approach to refining the LSTM model for improved time series data prediction accuracy.

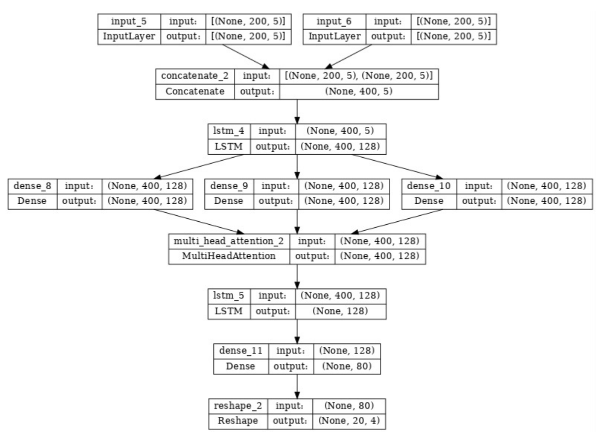



Figure 2: LSTM model with Attention Layer and Auxiliary Input Architecture.

## 2.2 Data Preprocessing

Data Preprocessing utilizes the min-max scaling function. Its primary objective is to transform a set of scalar points into a standardized range usually **0 - 1**, ensuring uniformity among the data points while maintaining their original magnitude. This normalization process is pivotal as it ensures consistency across the dataset, preventing certain features from overpowering others.

The model benefits from balanced contributions of each feature without domination by any particular attribute. This uniformity optimizes the learning process, facilitating the model's ability to discern patterns efficiently while avoiding undue bias towards specific features.

# 3 Experiments

## 3.1 Datasets

Training the model involves using datasets from Yahoo Finance. This platform allows data extraction based on specified start dates, end dates, and target companies, providing comprehensive access to historical market trends and metrics. This study uses datasets from three entities: Amazon, Apple, and Duke Energy. The aggregation of the datasets results in a final total comprising 17,166 records.

To initialize the training set, the records were divided into 'input' and 'prediction' sets. Specifically, each sequence consisted of **200** consecutive days of historical stock data as input, and the prediction set with subsequent **20** days. The objective of the model is to discern and learn different features to improve its predictive capabilities.

Moreover, for comprehensive model evaluation, the test set was divided into three distinct time frames, detailed in Table **(figure 3)**. The prediction capability of the model is cross-checked using the time frames.

| Input | Prediction(output) |
|---|---|
| 200 days(Jan-2023 to July-2023) | 20 days (Aug-01-2023 to Aug-20-2023) |
| 200 days(Jan-2023 to Sep-2023) | 20 days (Sep–09-2023 to Oct-13-2023) |
| 200 days(Jan-2023 to Oct-2023) | 20 days (Oct-11-2023 to Nov–10-2023) |

Figure 3: Overview of Test Dataset

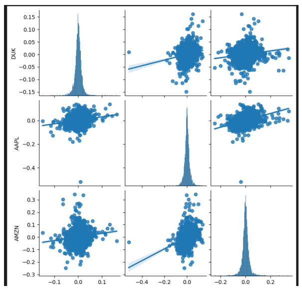
The research further explores correlations among the datasets. Correlation, typically ranging from **-1 to 1**, represents a statistical measure quantifying the extent to which two variables co-vary. It illuminates the directional relationship between these variables, signifying how they move concerning each other. However, it's essential to note that correlation doesn't inherently establish causation between the data points, merely indicating the degree and direction of their association. From **Figure 4**, it can be seen that the correlation between Apple and Amazon seems interesting with a positive value of **0.37**. This means that there is a positive chance that if Apple's price goes up, Amazon's might go up as well.

Finally, to understand the impacts of using auxiliary input with attention-based LSTM layers, a separate dataset called **S and P 500** is used as an auxiliary input to the attention layers. The dataset consists of a total of **6023** records as input.

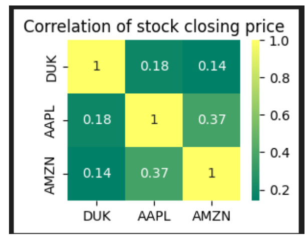
## 3.2 Experiment Setup

The system used to train and evaluate the model had the following specifications. The system had an Intel Core i7-8750H CPU operating at 2.2GHz with 8GB RAM and an NVIDIA GTX 1050Ti GPU with 4GB memory capacity. Model development took place in the Anaconda IDE, leveraging the power of deep learning frameworks such as Tensorflow and Keras. Additionally, essential machine learning libraries, including Numpy, Pandas, and Matplotlib, were employed throughout various stages of the implementation.

In terms of model architecture, the experiments were designed to access the model performance on diverse stock price datasets from three different companies. The evaluation encompassed a Vanilla LSTM model, an LSTM model with an Attention layer, and an LSTM model with an Attention Layer and Auxiliary input. The performance of all three models was assessed for all three datasets. The training phase spanned approximately **10 to 15** epochs for each model, providing a comprehensive evaluation of their performance and adaptability to the unique characteristics of the different stock datasets.




(a) Scatter Plot Showing Correlation for Closing Price of Datasets

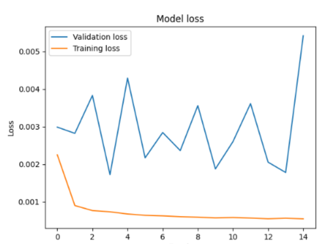



(b) Heat Map Showing Correlation Values for Closing Price of Datasets

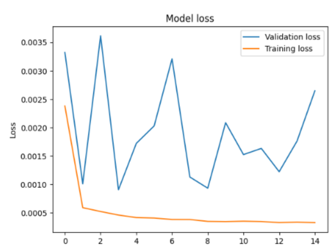
Figure 4: Correlation Plots

# 4 Results

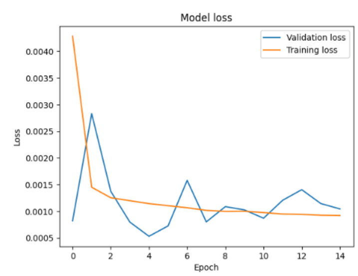
## 4.1 Vanilla LSTM Model Performance (Baseline Model)




(a) Amazon



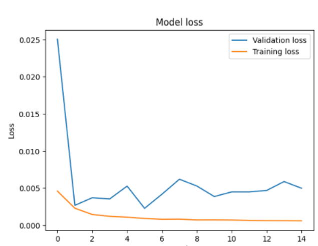

(b) Apple



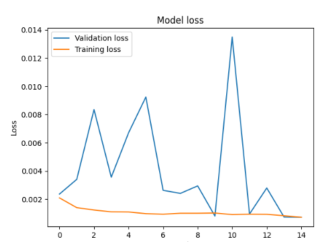

(c) Duke Energy

Figure 5: Vanilla LSTM Model Performance (Baseline Model)

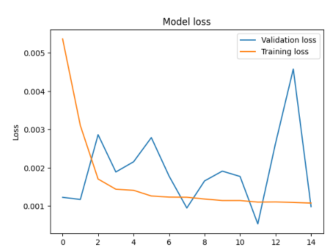
## 4.2 LSTM with Attention Layer Model Performance




(a) Amazon



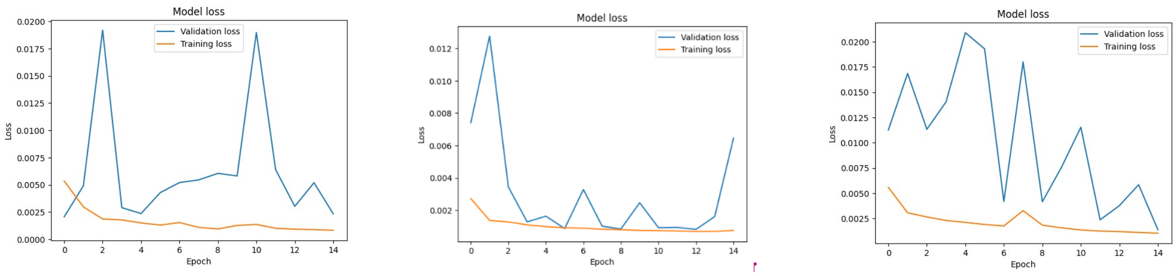

(b) Apple




(c) Duke Energy

Figure 6: LSTM with Attention Layer Model Performance

### 4.3 LSTM Model with Attention Layer and Auxiliary Input Performance




(a) Amazon (b) Apple (c) Duke Energy

Figure 7: LSTM Model with Attention Layer and Auxiliary Input Performance

# 5 Discussions

## 5.1 Limitations

While the proposed models exhibit good performance in predicting future stock price trends and prices, there exists some room for improvement. Notably, the model's current limitations become evident in scenarios of swift stock price fluctuations induced by external factors, such as breaking news or other unforeseen financial indicators. The model's inability to promptly capture these rapid movements introduces a degree of inaccuracy. Addressing this limitation could enhance the model's adaptability to dynamic market shifts.

## 5.2 Future Directions

To mitigate the aforementioned limitation, the adoption of early stopping techniques emerges as a prospective avenue. This involves halting model training or reverting to a prior training epoch by saving weights from the epoch with optimal performance.

Furthermore, there lies potential in the exploration of larger and more sophisticated models. Introducing advanced financial indicators, including analyses of companies' new product lines, product reviews, and comprehensive examinations of quarterly and annual financial reports, among other factors, could enrich the model's predictive capabilities.

# 6 Conclusion

During the course of this research, various versions of LSTM-based models were developed, trained, and validated on diverse datasets to evaluate individual model's performance.

In conclusion, the study helps to understand the suitability of the LSTM model for stock market price prediction. Additionally, the research also explores the effectiveness of using LSTM Network with Multi-Head Attention Layers to check if it provides better performance than Vanilla LSTM Network. We further research the possibility that the model with attention layers will outperform previous models in performance.

# References

[1] Chung, J., et al. (2014). "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling." arXiv preprint arXiv:1412.3555.

[2] Zhang, S., et al. (2018). "Short-Term Load Forecasting using Recurrent Neural Networks: An Extreme Learning Approach." IEEE Transactions on Smart Grid, 9(1), 477-485.

[3] Bu S-J, Cho S-B. Time Series Forecasting with Multi-Headed Attention-Based Deep Learning for Residential Energy Consumption. Energies. 2020; 13(18):4722. https://doi.org/10.3390/en13184722

[4] Y. Wang, S. Zhu and C. Li, ”Research on Multistep Time Series Prediction Based on LSTM,” 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 2019, pp. 1155-1159, doi: 10.1109/EITCE47263.2019.9095044.]

[5] Van Houdt, G., Mosquera, C. Nápoles, G. A review on the long short-term memory model. Artif Intell Rev 53, 5929–5955 (2020). https://doi.org/10.1007/s10462-020-09838-1