Forecasting Stock Prices Using LSTM and Attention-Based Deep Learning Models

Victoria Bayang, Jacqueline Sanchez

College of Natural Science and Mathematics, University of Houston

COSC 4368 - Fundamentals of Artificial Intelligence

Dr. Yang

April 10, 2025

Contents

I. Introduction and Problem Description

a . Background

Forecasting stock prices is a challenging task due to the sequential and volatile nature of financial markets. Stock data naturally forms a time series, where today's behavior is influenced by prior trends, external events, and investor sentiment. Unlike standard machine learning tasks that assume data points are independent, stock prediction demands models that can capture dependencies across time.

Feedforward neural networks, which treat inputs independently, are ill-suited for such tasks. Effective stock prediction requires architectures capable of remembering and modeling sequential patterns, preserving temporal structure.

b. Challenges in Time-Series Prediction

While Recurrent Neural Networks (RNNs) introduced memory into machine learning models, they are prone to the **vanishing gradient problem** when dealing with long sequences. As gradients backpropagate across many time steps, they can shrink exponentially, making it difficult for RNNs to learn long-term dependencies, a critical need in financial forecasting, where older events may still influence future prices.

c. Solution Approach

To overcome these challenges, **Long Short-Term Memory (LSTM)** networks were developed. LSTMs address the vanishing gradient problem through specialized gating mechanisms including forget gates, input gates, and output gates which enable the network to retain, update, and control information flow across longer sequences. This architectural innovation makes LSTMs particularly effective for tasks like stock prediction, where both short-term fluctuations and long-term patterns matter.

Beyond LSTM alone, **Attention mechanisms** provide an additional layer of enhancement. Attention allows the model to dynamically assign different importance weights to different time steps in the input sequence, enabling it to focus on the most relevant events when making a prediction. This can lead to improved predictive performance and better interpretability.

II. Project Objective and Methodology

In this project, we aim to build and compare two models: a standard LSTM model and an LSTM model enhanced with an Attention mechanism.

Both models will be trained to predict the next day's closing stock price based on historical sequences of Open, High, Low, Close, Adjusted Close, and Volume data. The dataset consists of daily trading records for NASDAQ-listed companies, sourced from Yahoo Finance using the yfinance Python package, and spans data up to April 2020. The stock being used is owned by Apple inc. Model performance will be evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and visual analysis of predicted versus actual stock price trajectories.

Models used:

1. LSTM Model

The Long Short-Term Memory (LSTM) model is a type of recurrent neural network (RNN) capable of learning long-term dependencies. Its architecture includes forget gates, input gates, and output gates, which allow the model to selectively retain or discard information across time steps. In this project, an LSTM model is used to predict the next day's closing stock price based on a sequence of 30 prior days of trading data.

2. LSTM with Attention Mechanism

The Attention mechanism extends the basic LSTM by allowing the model to dynamically focus on different parts of the input sequence. Instead of treating every time step equally, Attention assigns different weights to each day's information, helping the model prioritize significant market events when making a prediction.

3. Evaluation Metrics

Model performance is assessed using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) on the testing set. In addition, predicted closing prices are plotted against actual closing prices to visually assess model tracking accuracy.

III. Literature Review

Ensemble Learning with Sliding Window and Sentiment Features

In the study **"A Novel Ensemble Learning Approach for Stock Market Prediction Based on Sentiment Analysis and the Sliding Window Method"** (IEEE, 2022), the authors investigate how combining sentiment analysis of financial news with technical stock indicators can enhance predictive performance. A key methodological contribution is their use of a Sliding Window technique to restructure both sentiment scores and historical price data into fixed-length sequences, capturing short-term temporal patterns crucial for modeling market behavior. This use of Sliding Windows directly aligns with the approach adopted in this project to structure sequential stock features for LSTM-based modeling. Their ensemble model, integrating Decision Trees, Support Vector Machines, and Random Forests, outperformed single models, highlighting the effectiveness of combining temporal structuring with diverse input features to improve financial forecasting accuracy.

Enhancing LSTM for Stock Prediction with Attention Mechanisms

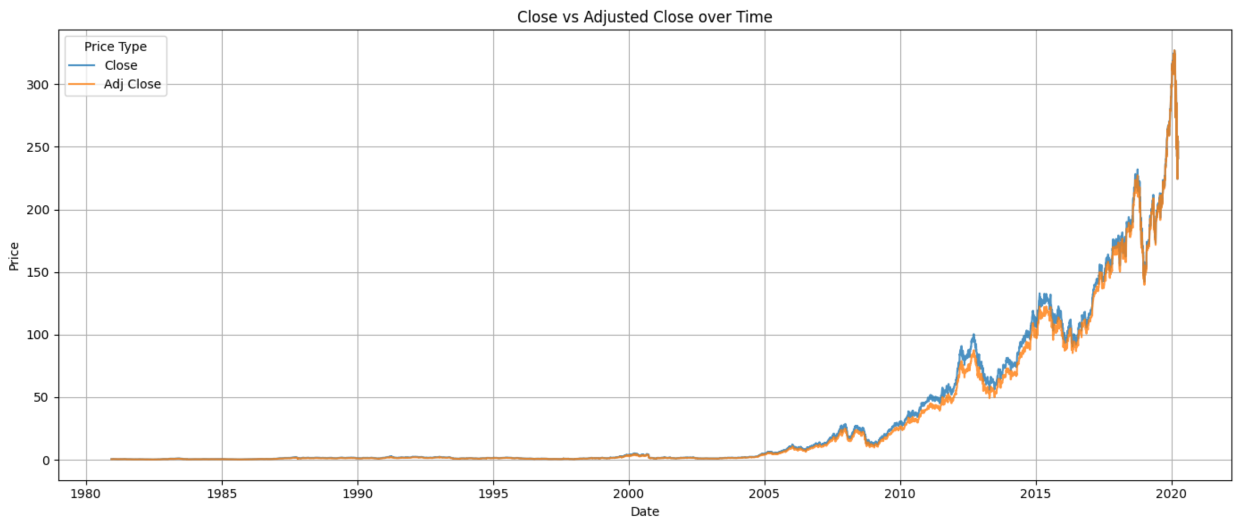
In **"A Novel Variant of LSTM Stock Prediction Method Incorporating Attention Mechanism"** (Mathematics, 2024), the authors propose an improved LSTM architecture aimed at enhancing the stability and accuracy of stock price predictions. They modify the traditional LSTM by coupling the forget and input gates and simplifying the forget mechanism, helping the model manage information flow more effectively across long sequences. Additionally, they integrate an Attention mechanism to dynamically prioritize more influential time steps when making predictions. This combination addresses common issues in standard LSTM training, such as overfitting and instability, while improving convergence speed. Their results demonstrate that the Attention-enhanced LSTM variant outperforms both traditional LSTM models and earlier attention-based designs. This study reinforces the idea that selectively focusing on key parts of a historical sequence, as achieved through Attention, can significantly improve stock forecasting, a principle directly influencing the model designs explored in this project.

[third article goes here]

IV. Results

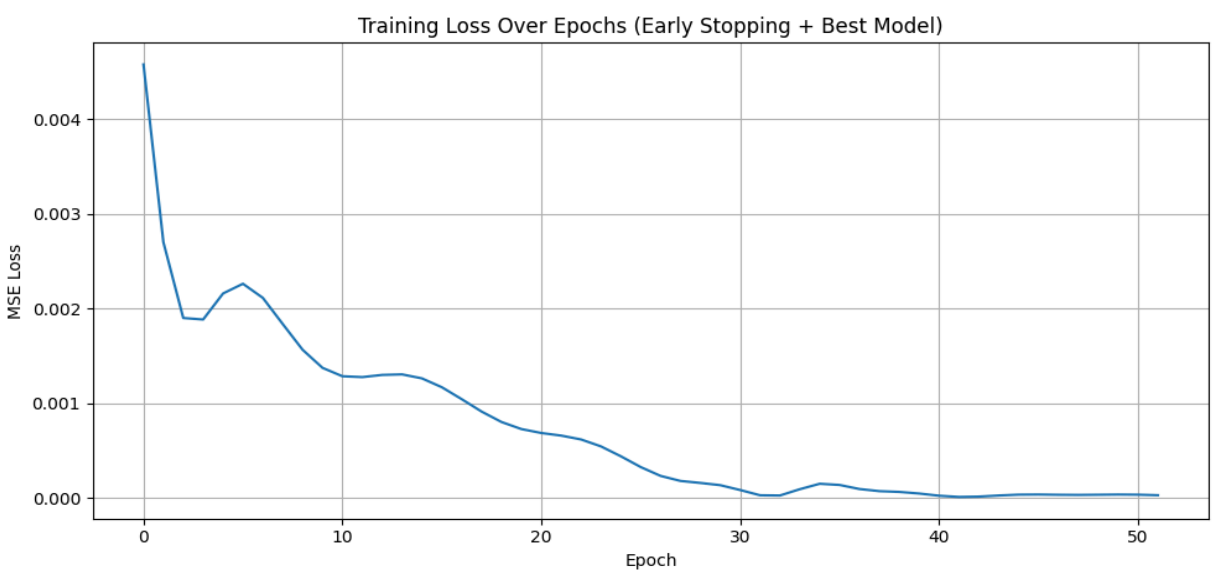
Long Short Term Memory (base model)

EDA



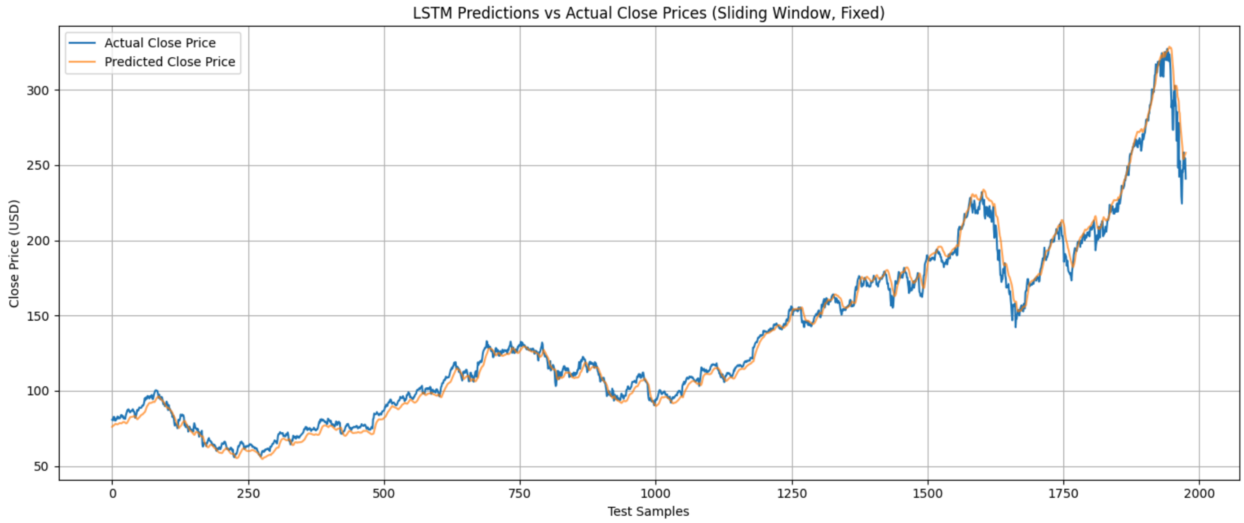
[Some insight]

Training Loss



[Some insight]

Predictions



[Some insight]

Long Short Term Memory with Attention

Cited Work

Sliding Window Method on LSTM stock price prediction

Chiong, R., Fan, Z., Hu, Z., & Dhakal, S. (2023). A novel ensemble learning approach for stock market prediction based on sentiment analysis and the sliding window method. *IEEE Transactions on Computational Social Systems*, 10(5), 2613–2623. https://doi.org/10.1109/TCSS.2022.3182375

Enhancing LSTM for Stock Prediction with Attention Mechanisms

Sang, S., & Li, L. (2024). A Novel Variant of LSTM Stock Prediction Method Incorporating Attention Mechanism. *Mathematics*, *12*(7), 945. https://doi.org/10.3390/math12070945