Title

DEEP-LEARNING BASED STOCK PRICE FORECASTER WITH TECHNICAL INDICATORS AND SENTIMENT ANALYSIS

Author

Dash, Swaraj

Publication Date

2026-06-15

DEEP LEARNING-BASED STOCK PRICE FORECASTER WITH TECHNICAL INDICATORS AND SENTIMENT ANALYSIS

By

Swaraj Dash

A capstone project submitted for Graduation with University Honors

May 07, 2026

University Honors

University of California, Riverside

APPROVED

Dr. Xiaoqian Liu

Department of Statistics

Dr. Begoña Echeverria, Howard H Hays, Jr. Chair

University Honors

TABLE OF CONTENTS

ABSTRACT

Accurate short-term equity forecasting is challenging due to nonstationarity and noisy market signals. This project develops an end-to-end forecasting prototype that predicts short-horizon price movement (future return over a user-selected horizon) for a selected stock. Historical open–high–low–close–volume (OHLCV) data are retrieved using yahoo finance API and transformed into a compact set of core technical indicators, including daily returns, moving averages, RSI, Bollinger percent-B, average true range (ATR), and volume/range-based measures. To incorporate qualitative market information that may not be fully captured by prices alone, the system also retrieves recent news headlines for the chosen ticker and computes VADER sentiment scores, producing a sentiment signal used to generate sentiment-adjusted forecasts for demonstration. The forecasting model is a Long Short-Term Memory (LSTM) network, a recurrent neural architecture designed to learn patterns and dependencies in sequential time-series data, and it is used because it can model temporal structure in engineered financial features across fixed-length historical windows. Model uncertainty is summarized using holdout residual variability to provide a confidence interval for predicted prices. The trained model, scaler, and metadata deployed through an interactive Streamlit application that displays results including predicted returns, uncertainty range, sentiment summary, and optional backtest comparisons.

ACKNOWLEDGEMENTS

In conclusion, this project has been a great chance to take my passion for finance and my skills in data science and turn them into something tangible and functional. I want to extend my deepest gratitude to Professor Xiaoqian Liu for her invaluable guidance throughout this process. Her mentorship ensured that while my project was creative, it remained grounded in technical rigor.

I would also like to thank the University Honors faculty for providing important lessons, workshops, and information regarding the capstone project. I would like to express my sincere gratitude to Ms. Latoya Ambrose, my counsellor at the Honors program for her support and guidance throughout the process.

INTRODUCTION

Starting my 3rd year in UCR, I developed a strong interest in the field of investment management. I was curious to learn what drives the value of a stock and how to assess it. This curiosity is what led me to join Highlander Financial Group (HFG), UCR's student managed finance club. During my time there, eventually serving as Vice President and Portfolio Manager, I gained valuable insights into what drives stock prices. The market often feels like a chaotic sea of random numbers, but beneath the surface, there are currents: patterns, behaviors, and sentiments, that can be deciphered if you have the right tools.

My creative project, titled "Deep Learning-Based Stock Price Forecaster with Technical Indicators and Sentiment Analysis," is the intersection of my passion for finance and my technical skills in Data Science. My goal was to predict stock price using sentiment data and by discovering hidden patterns beneath numerical data. While I am well aware that it isn't possible to predict market behavior with 100% certainty, I believe we can use deep learning to identify the most probable path forward by processing information at a scale no human analyst could match.

Before diving into neural networks and sentiment scores, it is vital to establish what a "stock" actually represents in the context of this project. A stock is a fractional share of ownership in a corporation. The stock price is effectively a real-time, high-stakes tug-of-war between buyers and sellers. It is influenced by a dizzying array of variables such as company earnings, interest rates, global politics, and even a stray tweet from a CEO.

Many people view price changes as purely random, the "Random Walk" theory (Fama, 1965). While it is true that short-term volatility can look like "noise," prices are driven by specific, measurable variables. In this project, I categorized these into two main streams:

- **Quantitative Data:** These are the “hard numbers,” including the Open, Close, High, Low, and Volume of each trading day.
- **Qualitative Data:** This is the “vibe check”, the news headlines and public discussions that influence how investors feel and act.

By combining these, the project aims to support a more data-driven financial analysis rather than relying on “guess and hit” methods that often miss hidden patterns in large datasets.

METHODOLOGY: DATA AND FEATURE ENGINEERING

The core objective of this project is to forecast future stock price movements over a 5-day window. To achieve this, I designed a multimodal pipeline that processes information much like a human analyst would, but with the added rigor of deep learning.

The model begins its journey by gathering data from two primary sources: Yahoo Finance for historical market figures and the News API for text-based sentiment data.

- **Market Data:** We collect the standard OHLCV (Open, High, Low, Close, Volume) metrics.
- **Text Data:** We pull headlines and public market discussions related specifically to the stock being analyzed.

Before the model ever sees this data, it undergoes a rigorous cleaning process. I remove duplicates, handle missing entries, which are common in financial sets due to market holidays, and most importantly, align the numerical data with the sentiment data by date.

Technical Indicators: To give the model more context, I calculated several Technical Indicators, which can be interpreted as summaries of recent behavior of numbers.

- **Moving Averages:** These smooth out price data to identify the broader trend (Varma P et al., 2024).

- **Relative Strength Index (RSI):** This helps the model understand if a stock is "overbought" (due for a dip) or "oversold" (due for a bounce) (Varma P et al., 2024).
- **Volume:** This is a crucial feature because it tells us how many shares were traded. High volume usually means a price move is significant, while low volume suggests a lack of market conviction (Varma P et al., 2024) .

Sentiment Analysis: Then comes sentiment analysis, the process of converting human language into a measurable signal. In this model, news headlines are analyzed and assigned a daily score, positive, neutral, or negative (Xiao & Ihnaini, 2023). This is added as a feature so the model can learn how market "mood" correlates with price behavior. For example, a positive news score might precede a price jump even if the previous day's price was trending down (Xiao & Ihnaini, 2023).

THE DEEP LEARNING MODEL: LSTM ARCHITECTURE

When choosing the mathematical "brain" for this project, I chose a Long Short-Term Memory (LSTM) neural network. Standard linear regressions look at data points in isolation, but stock data is a Time Series, a sequence where chronological order is everything (Pan, 2024). What happened 60 days ago influences the context of what happens today (Pan, 2024).

An LSTM is specifically designed to have a "memory". It processes sequences and decides which information is important to keep and which can be forgotten. Mathematically, we represent the input at any time t as a vector X, containing our price data P, volume V, technical features F, and sentiment scores S.

$$X = [P, V, F, S]$$

The model then looks at a sequence of these observations to produce Y(t+1), which is our best estimate for the future price movement.

APPLICATION AND PROFESSIONAL CONTEXT

The final phase of this project was the creation of a live application demo, hosted locally on my system. This demo allowed me to show, in real-time, how the model takes a specific ticker, processes the last 60 days of sequences, and outputs a 5-day forecast.

The larger meaning of this work is that financial prediction is not just a numerical problem; it is a behavioral and information-driven problem. By combining market history with human sentiment, we create a stronger, more "human-aware" forecasting system.

In the broader professional world, the forecasting system I have built is a functional microcosm of the sophisticated algorithmic engines used by major global financial institutions every single day (Sang & Li, 2024). Investment corporations utilize similar multimodal pipelines to maintain their competitive edge (Sang & Li, 2024). These organizations recognize that because stock prices change rapidly and are notoriously hard to predict, they must combine market history with vast streams of market sentiment signals to strengthen their forecasting capabilities. While the Long Short-Term Memory (LSTM) network was my primary choice due to its sequence-handling capabilities and its ability to link price behavior with market mood, the industry is constantly exploring even more exotic architectures. For instance, many professionals now experiment with Gated Recurrent Units (GRUs) to achieve faster training times or use state-of-the-art Transformer models that utilize attention mechanisms to focus on specific historical events with incredible precision (Sang & Li, 2024). Some firms even lean into Temporal Fusion Transformers or complex hybrid models that blend classical autoregressive statistical methods with deep learning to better implement risk controls (Yang et al., 2025). Engaging with this project confirmed my belief that financial prediction is not solely a numerical puzzle to be solved with a calculator, but a complex behavioral and information-driven challenge that requires a deep understanding of human sentiment. Real-world forecasting is rapidly moving toward these adaptive, real-time systems that can learn from prices, news, and macro signals simultaneously.

Future Work

Looking back at this creative journey, there are several areas where I believe accuracy could be improved. As markets move toward real-time, multimodal regimes, future versions of this predictor should include:

- **Macroeconomic Signals:** Integrating interest rate changes and inflation data.
- **Richer Features:** Analyzing the nuance of entire earnings call transcripts rather than just short headlines.
- **Model Tuning:** Implementing more robust risk controls and hybrid pipelines that combine deep learning with traditional statistical methods.

CONCLUSION AND REFLECTION

Reflecting on this creative journey, the transition from classroom theory to a functional application demo was a profoundly transformative experience that fundamentally shifted my perspective on the practical utility of data science. My academic foundation in Stat 146 (Time-series modelling) and CS 171 (Artificial Intelligence and Data Mining) served as the bedrock for the technical decisions I made throughout this process. Stat 146 helped me understand the true gravity of chronological order, the idea that a stock's past carries important information, a sequence of observations that influences future behavior. Within those lectures, I learned to appreciate the nuances of 60-day sequences as a means to capture meaningful market patterns. Conversely, CS 171 taught me the importance and raw power of deep learning architectures that can learn complex, non-linear relationships directly from raw data without the need for manual, rigid rules. This project allowed me to finally synthesize these two worlds, taking the abstract mathematical representations of LSTM layers, specifically the interaction between price-related inputs and hidden states, and applying them to the volatile reality of the stock market where traditional linear regression or simple "guess and hit" methods often miss the most important hidden signals. I learned that data collection and preprocessing are perhaps the most critical yet overlooked steps. Cleaning the data by removing missing, duplicate, or inconsistent entries is the only way to ensure the model is learning from truth rather than noise. This arduous process taught me that a model is truly only as good as the features it consumes, and manually engineering technical indicators like moving averages and RSI allowed me to provide the neural network with the necessary "hints" it needed to decipher market activity and volatility.

Through this journey, I realized that while accuracy can always be improved through better data quality and model tuning, the ultimate goal is to move past the "guess" and use deep learning to learn the patterns that help support data-driven financial analysis. Financial prediction will likely never be a "solved" science, but with a bit of curiosity, a lot of data, and a few thousand lines of code, we can certainly get closer to understanding the rhythm of the market.

Thank you for the opportunity to present this work through the University Honors program.

REFERENCES

Sezer, Omer Berat, et al. "Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review: 2005–2019." *Applied Soft Computing*, vol. 90, 2020, p. 106181, https://doi.org/10.1016/j.asoc.2020.106181.

Vijh, Munish, et al. "Stock Closing Price Prediction Using Machine Learning Techniques." *Procedia Computer Science*, vol. 167, 2020, pp. 599-606, https://doi.org/10.1016/j.procs.2020.03.326.

Fama, Eugene F. "Random Walks in Stock Market Prices." *Financial Analysts Journal*, vol. 21, no. 5, 1965, pp. 55–59. *Taylor & Francis Online*, https://doi.org/10.2469/faj.v21.n5.55.

P, Krishna Kanth Varma, et al. "Stock Market Price Prediction Using Neural Networks (LSTM) and Technical Indicators." *AIP Conference Proceedings*, vol. 3131, no. 1, 19 Sept. 2024, article 020003. *AIP Publishing*, https://doi.org/10.1063/5.0229761.

Xiao, Qianyi, and Baha Ihnaini. "Stock Trend Prediction Using Sentiment Analysis." *PeerJ Computer Science*, vol. 9, 2023, article e1293. *PeerJ*, https://doi.org/10.7717/peerj-cs.1293.

Pan, Yixuan. "Compare of Linear Regression Model and LSTM Neural Network in Machine Learning." *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, SciTePress, 2024, pp. 218–223. https://doi.org/10.5220/0012871700004547.

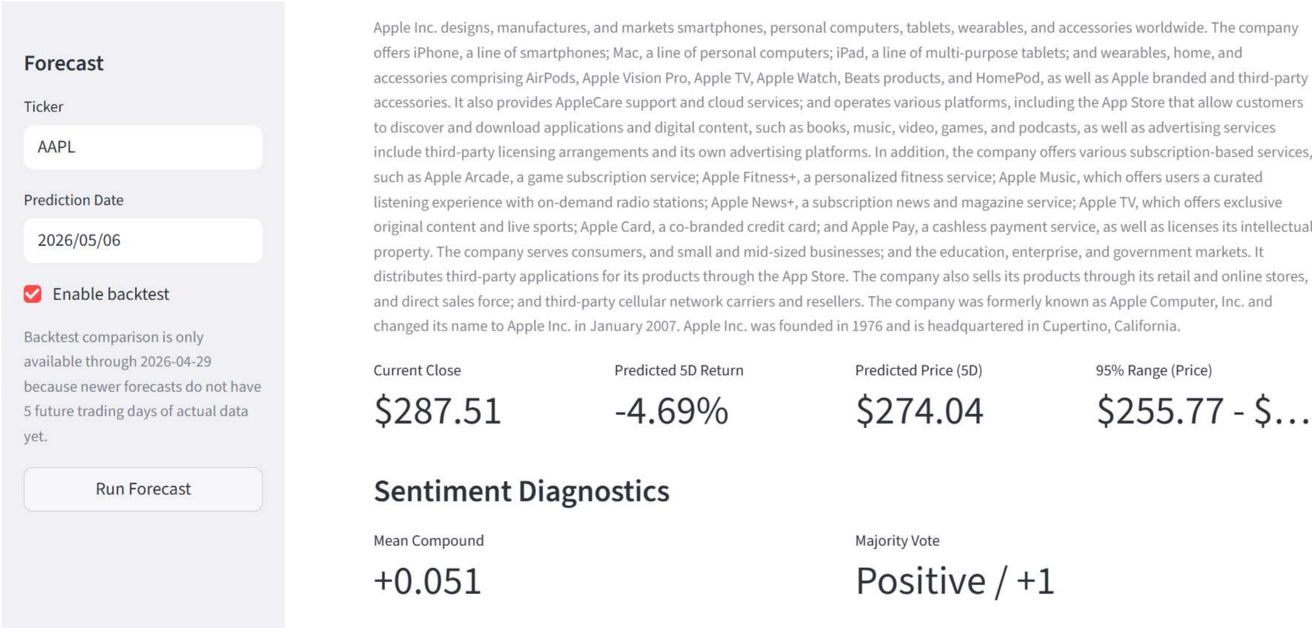
Sang, Shuai, and Lu Li. "A Novel Variant of LSTM Stock Prediction Method Incorporating Attention Mechanism." *Mathematics*, vol. 12, no. 7, 2024, article 945. *MDPI*, https://doi.org/10.3390/math12070945.

APP DEMO

- Initial Page showing options to choose stock ticker, set the t-5 days prediction date, and to enable backtesting for Prediction date that was in the past.

- Selected AAPL (Apple) ticker. Shows numeric info.

- Shows company info and more metrics along with sentiment scores as you scroll down.



- Shows news headlines with sentiment score for each along with date and source.

- Shows price history + Forecast market aiding in visual understanding.

Price History + Forecast Marker