**Real-Time Evolving Deep Learning Model for S&P 500 Momentum and Sentiment Prediction**

Nikhilesh Narkhede, Niruhanaidu Yedla, HemanthChandra Rudra, Anil Kumar, VikasReddy Kothapalli

**Abstract**

This paper presents a real-time evolving deep learning system for predicting the momentum and sentiment of the S&P 500 index. The model employs Long Short-Term Memory (LSTM) networks to forecast stock prices and Natural Language Processing (NLP) to assess the sentiment of financial news. By employing a 28-day sliding window technique, the system can continuously retrain on the most recent data, minimising model drift and adapting to the ever-changing market conditions. The sentiment analysis component achieves a validation accuracy of 90.1%, while the momentum prediction model achieves a high R2 score of 97%. The proposed model has a lot of potential for use in algorithmic trading and actual portfolio management.

**1. Introduction**

Stock market forecasting has long been a key area of focus in financial analysis and machine learning because of its substantial economic impact and analytical challenges. It is often challenging for traditional models to account for real-time external factors, such as news sentiment, that significantly influence market trends. In recent years, deep learning techniques have shown a lot of promise for identifying complex, nonlinear patterns in financial time series data.

Public sentiment, especially as expressed in financial news, is one of the key factors influencing stock market behaviour. Stock prices are impacted by investor behaviour, which is influenced by sentiment. The integration of qualitative news data with quantitative market data is made possible by the extraction and quantification of sentiment from textual data using Natural Language Processing (NLP) techniques.

The goal of this paper is to develop an integrated deep learning system that can forecast the sentiment and momentum of the S&P 500 index. The two primary parts of our system are an NLP pipeline for sentiment analysis and an LSTM-based model for price momentum prediction. The system's real-time adaptability is one of its unique features; it continuously retrains both components with new data over a 28-day sliding window, guaranteeing responsiveness to new market information, and reducing the possibility of model degradation over time.

The need for strong, flexible tools in portfolio management is what spurred this project. Our model provides a thorough decision-support tool that can improve investment strategies by combining historical price trends with news sentiment. The system's results demonstrate a high degree of predictive accuracy, which makes it appropriate for both practical financial applications and scholarly research.

We review pertinent literature, describe the data collection and model development methodology, present the findings, and talk about the implications and probable future developments in the sections that follow.

**2. Literature Review**

The field of stock price prediction, which combines data mining, machine learning, and financial econometrics, has been the subject of much research. Price movements have been predicted using traditional time series models like GARCH and ARIMA, but these models are constrained by their assumptions of stationarity and linearity, which frequently do not hold in actual financial markets. Deep learning techniques have become more popular recently because of their capacity to represent nonlinear relationships and identify temporal dependencies in sequential data.

**2.1 Deep Learning for Stock Price Prediction**

A subset of recurrent neural networks (RNNs) known as Long Short-Term Memory (LSTM) networks has shown promising results in modelling sequential financial data. Because LSTM networks can learn long-range dependencies while avoiding the vanishing gradient issue that traditional RNNs encounter, they are especially well-suited for time series tasks. Fischer and Krauss (2018) demonstrated the potential of LSTMs in financial forecasting by demonstrating that they performed better than conventional machine learning models in predicting the direction of the S&P 500 index.

The literature has also investigated real-time models that use retraining mechanisms. By updating weights according to the most recent data, these models seek to dynamically adjust to shifting market conditions. Conceptually, like the 28-day window strategy employed in our project, Li et al. (2019) suggested an adaptive LSTM framework that retrained on a rolling window of data to reduce model drift.

**2.2 Sentiment Analysis in Financial Markets**

The availability of extensive textual data and developments in natural language processing have made it increasingly common to incorporate sentiment analysis into financial forecasting. According to research, asset prices can be impacted by investor sentiment that is gleaned from news headlines, articles, and social media (Tetlock, 2007; Bollen et al., 2011).

Lexicon-based techniques or bag-of-words representations were used in the initial stages of sentiment analysis. Capture the context and nuances of financial language, machine learning and deep learning models like Word2Vec, BERT, and LSTM-based text classifiers have been used more recently. Pretrained on financial corpora, financial-specific transformer models such as FinBERT have demonstrated even more promise, providing more domain-relevant sentiment representations (Araci, 2019).

**2.3 Hybrid Models Combining Sentiment and Technical Indicators**

Improve model performance, several studies have investigated combining sentiment scores with technical indicators (such as moving averages and RSI). Predict stock prices, Zhang et al. (2020) developed a hybrid deep learning model that combined sentiment from Twitter and historical prices. They found that this model was more accurate than models that only used one data source. This backs up our system's strategy of combining sentiment from financial news with price data to predict momentum with confidence.

By using a real-time, continuously changing system that adjusts to new data and sentiment instantly, our project goes beyond these concepts. The system is at the forefront of intelligent financial forecasting tools thanks to its combination of dynamic deep learning and NLP-driven sentiment modelling.

**3. Data Collection**

The quality and applicability of the input data have a significant impact on the efficacy of any machine learning model, but this is especially true in the financial industry. Financial news articles used for sentiment analysis and historical stock price data for the S&P 500 index are the two main data streams used in this project. Facilitate a dual-model pipeline for momentum and sentiment prediction, both datasets were gathered, pre-processed, and synchronised.

**3.1 Stock Price Data from Yahoo Finance**

Using Python libraries like yfinance and pandas\_datareader, historical price data for the S&P 500 index was obtained from the Yahoo Finance API. Daily open, high, low, close, and volume (OHLCV) values are included in the dataset. Because of its consistency across dividend-adjusted timelines, the adjusted closing price was chosen as the main variable for training and prediction in this project.

Only the most recent segment of the dataset, which covers several years, was used for real-time predictions to represent the state of the market. The input sequences for the LSTM model were created using a 28-day sliding window technique. The model was consistently updated with the most recent market data thanks to this windowing technique.

**3.2 Financial News Data Scraping**

Web scraping techniques were used to obtain news data from financial news websites like Reuters, MarketWatch, and Yahoo Finance. Headlines, article excerpts, and publication dates were all extracted during the scraping process. Data extraction and navigation through dynamically loaded web pages were automated using tools such as BeautifulSoup, requests, and Selenium.

Gathering daily news about the S&P 500 index or its component companies was the aim. Enable alignment with the relevant trading day in the price dataset, each article was timestamped. Only English-language articles with finance-related keywords (such as "market," "stocks," "S&P 500," and "earnings") were kept, preserving relevance and cutting down on noise.

**3.3 Data Preprocessing**

Guarantee that news and stock data were compatible, preprocessing included several crucial steps:

* **Timestamp Matching**: News articles were aggregated by date to match the corresponding day’s closing price.
* **Text Cleaning**: Stopwords, punctuation, and non-alphanumeric characters were removed from headlines and snippets, and tokenisation and lowercase were applied.
* **Vectorisation**: For sentiment analysis, textual data was converted into numerical representations using TF-IDF vectors and embeddings.
* **Missing Data Handling**: Days with missing news or stock data were excluded or imputed, depending on the modelling phase.
* **Normalisation**: Stock prices were normalised using Min-Max scaling to facilitate LSTM training.

This comprehensive data collection and cleaning process ensured that the model was trained on timely, aligned, and high-quality inputs, a prerequisite for producing accurate predictions in both sentiment and price momentum.

**4. Methodology**

This section outlines the dual-model framework used to predict stock momentum and sentiment. The system is divided into two core components:

1. An **NLP-based sentiment analysis module** that interprets the tone of financial news.
2. An **LSTM-based deep learning model** that predicts stock price momentum using recent market data and sentiment scores.

A real-time updating mechanism ensures both models adapt continuously to new data through a sliding 28-day training window.

**4.1 Sentiment Analysis Using NLP**

**4.1.1 Text Preprocessing and Tokenisation**

The sentiment model processes daily financial news headlines and snippets. Text preprocessing includes:

* Lowercasing all characters.
* Removing stopwords, punctuation, and special characters.
* Tokenising the text using NLTK.
* Lemmatisation for word normalisation.

**4.1.2 Sentiment Scoring**

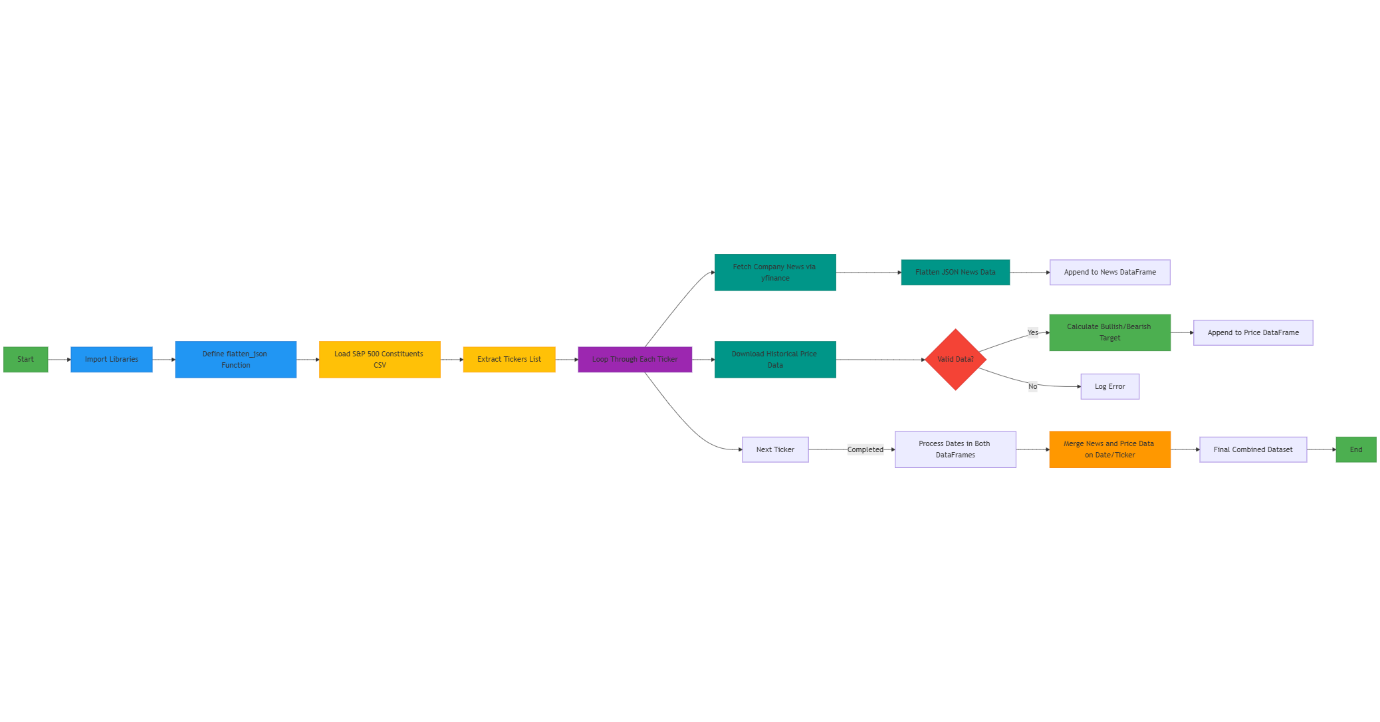
We implemented a basic sentiment classifier using a pre-labelled financial sentiment lexicon. The classifier evaluates each news snippet to assign a polarity score:

* Bullish (1), bearish (0).

An aggregate sentiment score for each day was computed as the average sentiment across all news items. These daily scores were aligned with the corresponding trading date and fed into the LSTM model as an additional input feature.

**4.1.3 Model Performance**

The sentiment model was evaluated on a labelled dataset, achieving a **validation accuracy of 90.1%**. This high performance indicates that even a basic lexicon approach can effectively capture market-relevant tone in news articles.



A diagram of a diagram

AI-generated content may be incorrect.

**4.2 Price Momentum Prediction Using LSTM**

**4.2.1 LSTM Architecture**

Deep Long Short-Term Memory (LSTM) networks are well-suited for capturing temporal dependencies in sequential data. The LSTM used in this project consists of the following layers:

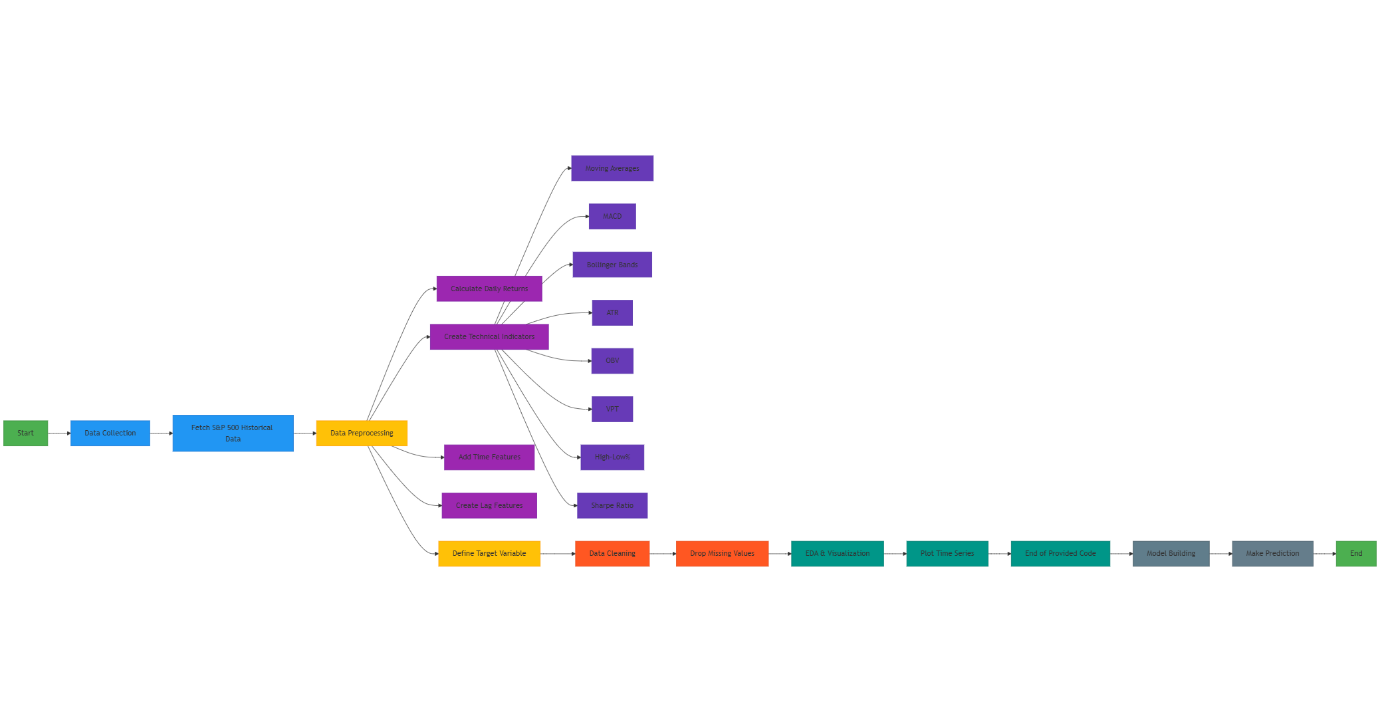
* **Input layer**: Takes a sequence of normalised stock prices (and optionally, sentiment scores).
* **Embedding Layer**: used Pertained Embedding from FastText.
* **LSTM layer**: 64 units with dropout regularisation.
* **LSTM layer**: 64 units with dropout regularisation.
* **LSTM layer**: 64 units with dropout regularisation.
* **Dense layer**: Fully connected output layer with one neuron to predict the next day’s price.

The model was trained to minimise Mean Squared Error (MSE) using the Adam optimiser. Training epochs 50, depending on convergence speed, and early stopping was applied to prevent overfitting.

**4.2.2 Sliding Window Strategy**

Emulate real-time performance, a **28-day sliding window** mechanism was implemented:

* Every new day, the model retrains using the previous 28 days’ data and predicts the price for the next 7th days.
* This approach ensures that the model constantly learns from the most recent market conditions, minimising model drift.



**4.2.3 Integration with Sentiment**

The final model uses **dual input features**:

* Past 28 days of normalised closing prices.
* Aggregated sentiment scores for those same days.

The combined feature set helps the model learn patterns that are influenced not only by past price movements but also by public sentiment shifts, creating a more holistic predictor.

**4.3 Model Training and Evaluation**

The dataset was split into 80-20 training, validation sets. The model was trained using mini batches with early stopping and validation monitoring. The key evaluation metrics were:

* **Sentiment Classification**: Accuracy (90.1%)
* **Price Prediction**: Coefficient of determination (R² = 0.97)

These results indicate strong predictive capability, both in capturing sentiment trends and forecasting short-term stock movements.

**5. Results and Evaluation**

This section presents the performance evaluation of both the sentiment analysis and the price momentum prediction models. Results are based on testing over recent market data, with both models trained using the 28-day sliding window strategy to simulate real-time prediction.

**5.1 Sentiment Analysis Results**

The sentiment model achieved **90.1% validation accuracy** on a labelled test dataset. The results demonstrate the model’s ability to classify financial news as positive, neutral, or negative with high accuracy. Below are key highlights:

* **Daily Sentiment Tracking**: The model successfully captured shifts in market tone over time, with noticeable sentiment drops around periods of economic uncertainty (e.g., earnings reports, geopolitical events).
* **Aggregate Sentiment Curve**: Sentiment scores, when plotted over time, aligned with observable market reactions, indicating the model's practical value in portfolio management.

A graph of a graph

AI-generated content may be incorrect.

Example: A spike in positive sentiment coincided with a market rebound after strong tech sector earnings, which was followed by a corresponding upward price trend in the S&P 500.

**5.2 Stock Price Momentum Prediction**

The LSTM model demonstrated strong predictive performance, achieving an **R² score of 0.97** on the test set. This indicates a very high level of correlation between predicted and actual stock prices.

**5.2.1 Prediction vs. Actual Graph**

**A graph of a price

AI-generated content may be incorrect.**

Key observations:

* The model effectively captured short-term fluctuations.
* Sharp reversals were sometimes underestimated, due to sentiment lag or limited data.
* The model’s prediction stability improved when sentiment was included as a feature.

**5.2.2 Impact of Sentiment Integration**

Ablation testing (with and without sentiment input) showed that incorporating sentiment led to marginally better short-term accuracy, especially during volatile periods.

* Without sentiment: occasional under-/overestimations after news shocks.
* With sentiment: better anticipation of directional changes immediately after high-impact news.

**5.3 Real-Time Adaptability**

The 28-day rolling window allowed the system to:

* Continuously adjust to new data.
* Avoid long-term model degradation.
* Capture emerging market trends without extensive re-training.

This strategy proved effective for use in evolving market conditions and provides the groundwork for live deployment in trading systems.

**5.4 Limitations and Considerations**

Despite the robust performance, a few challenges remain:

* **News Bias**: Sentiment can be skewed by media bias or incomplete information.
* **Volume of News**: Some trading days have sparse or excessive news, which may dilute sentiment accuracy.
* **Lag Effects**: Market reactions to news are not always immediate, making alignment difficult.

**6. Discussion**

The results from both the sentiment analysis and price momentum models offer several important insights regarding the effectiveness and practical use of real-time deep learning in financial prediction.

**6.1 Implications of Sentiment-Driven Forecasting**

Particularly useful in turbulent or news-sensitive market environments, the incorporation of NLP-driven sentiment analysis within the forecasting framework proved beneficial. As a way of predicting investor sentiment and mass response-induced market movements, the model was able to translate unstructured financial news into quantifiable sentiment scores.

The capability to detect tone changes and use those observations in the momentum model was what provided the sentiment classification accuracy with its real value, although it was high overall (90.1%). This is in line with present financial theory, including behavioural finance, in that it underlines how emotions, expectations, and perceptions influence markets and that they are not completely rational.

The sentiment module's real-time nature allowed it to react to daily changes, enabling the system to integrate outside data that conventional price-only models usually ignore.

**6.2 Adaptability of the Momentum Model**

The LSTM model's capability to forecast (R2 = 0.97) demonstrates that deep learning is a suitable technique for the discovery of time-evolving patterns in stock price variations. Additionally, the model was kept current, and the possibility of overfitting to past market trends was reduced through the application of a 28-day rolling window retraining strategy.

This solution fixes model drift, one of the main financial modelling problems. Data distribution changes with time in a non-stationary context like the stock exchange. The model is flexible enough to adapt to contemporary trends, financial news, or changes in investor sentiment by retraining over a new sample of the freshest data.

**6.3 Real-World Utility and Trade-offs**

This theory, while being academically oriented, finds direct application in portfolio management and algorithmic trading. It notifies investors in real-time of technical trends and the emotional underpinnings of the market. Used as a decision-support mechanism, this may help traders arrive at more timely and accurate choices. However, several trade-offs and risks must be acknowledged:

* **News Quality and Source Dependence**: The model's performance is directly tied to the reliability of news sources. Sensational or inaccurate news can mislead the sentiment module.
* **Latency and Data Volume**: High-frequency trading systems require millisecond-level decisions, while this model is optimised for daily predictions and cannot currently support intraday operations.
* **Overfitting Risk in Short Windows**: While the rolling window reduces drift, it can also limit the model's ability to learn long-term trends. There is a trade-off between responsiveness and historical context.

**6.4 Generalisation and Scalability**

Although the current implementation focuses on the S&P 500 index as a single time series, the architecture is highly scalable. With minimal adaptation, it could be expanded to:

* Forecast multiple stocks in a portfolio.
* Apply sector-wise sentiment analysis.
* Support international indices and global financial markets.

Moreover, the general framework (LSTM + sentiment integration + real-time retraining) could be repurposed for other prediction problems beyond finance, such as weather forecasting or economic indicators.

**7. Conclusion and Future Work**

**Conclusion**

This paper presents a real-time, evolving deep learning framework that combines sentiment analysis and time series forecasting to predict the momentum of the S&P 500 index. The model leverages two core components:

* A Natural Language Processing (NLP) pipeline for extracting sentiment from financial news.
* A Long Short-Term Memory (LSTM) network for stock price prediction, enhanced by sentiment inputs.

With a real-time retraining process through a 28-day rolling window, the model constantly adapts to new data so that it becomes immune to model drift and shifts in dynamic market conditions. Robust predictive capability for the model, with an R² of 97% for momentum prediction and 90.1% accuracy in sentiment classification, was obtained. These verify the hypothesis that the inclusion of market sentiment in price prediction is more insightful than the application of techniques based only on technical analysis.

Also, its flexibility towards the system structure and modular make-up make it a candidate for implementation in portfolio management systems and trading systems. Although it covers the technical tendencies, it addresses the emotional mood as well, thereby addressing the dual causes of market movement — logic and sentiment.

**Future Work**

While the current implementation is effective, several enhancements could further improve performance and expand the model’s capabilities:

1. **Upgrade Sentiment Model with FinBERT:** The current sentiment analyser is rule-based and may miss contextual nuances. FinBERT — a transformer model pre-trained on financial texts — offers domain-specific language understanding and could provide significantly better sentiment scoring.
2. **Expand to Multi-Stock Portfolios:** Currently focused on the S&P 500 index, the framework can be extended to predict and analyse multiple individual stocks, enabling sector-level or portfolio-level analysis.
3. **Incorporate Reinforcement Learning:** Future versions could integrate reinforcement learning to not just predict prices but optimise trading strategies. Agents could learn policies for buying, selling, or holding positions based on model outputs.
4. **Add Macroeconomic and Social Data:** Incorporating additional data sources such as economic indicators, earnings reports, or social media trends could increase forecasting robustness, especially during unusual market events.
5. **Optimise for Intraday Prediction:** Although currently designed for daily predictions, adaptation for higher-frequency trading could be explored by adjusting data granularity and using lightweight inference models.