Predicting Stockmarket with stacked LSTM and NLP techniques

Bekhbat Davaatseren, Kandasamy Illanko
Department of Electrical, Computer, and Biomedical Engineering
Toronto Metropolitan University, Toronto, Canada
bekhbat.davatseren@torontomu.ca, killanko@ryerson.ca

***Abstract*—The paper proposes a hybrid deep learning model to forecast the next-day closing price of Tesla stock (TSLA) using historical price data and financial news articles. The model architecture integrates an attention based stacked bidirectional long short-term memory (LSTM) for sequential stock price data with a DistilBERT-based module for news sentiment analysis. The research utilizes the Finsen dataset (2011–2023), which contains stock prices and related financial news, to establish correlations between news sentiment and stock price movements. Experimental results show that the proposed hybrid model outperforms two baseline approaches: a DistilBERT-only model and a Bi-LSTM only model. The hybrid model achieved the lowest root mean square error (RMSE) of 9.71 and the mean absolute error (MAE) of 7.17, as well as the highest R^2 of 0.971. These results indicate that integrating textual sentiment with historical price data substantially improves forecasting accuracy.**

***Index Terms*—LSTM, Stock Prediction, BERT, Recurrent Neural Network**

I. Introduction

Stock market prediction is a daunting task that has been the research focus for a long time due to the important economic implications. Classic forecast methods, including technical and fundamental analysis, have been of some utility but are limited in their ability to cope with the complexity. Technical analysis relies on historical price patterns and statistical indicators (e.g., moving averages, oscillators) to forecast future price action. Fundamental analysis, on the contrary, analyzes financial reports, economic indicators, and news wires to estimate the underlying value of a stock. The inherent complexity of the task highlights the need for more advanced, data-driven modeling methods in predicting stocks.

Deep learning models have emerged as strong alternatives to financial time series modeling in recent years [11]. While linear models struggle with nonlinear relationships and multivariate correlations, deep network architectures can automatically discover high-order patterns in big data. Specifically, recurrent ones, such as LSTM, are more capable of handling the temporary relationships and long-term correlations.

Another essential progress in predicting the stock is based on the integration of natural language processing (NLP) to process unstructured textual data. Incorporating these textual cues as a type of fundamental analysis has been demonstrated to improve the prediction model. Some of the early attempts to leverage news sentiment in prediction used simplistic sentiment scores or bag-of-words representations, potentially neglecting to account for context and subtleties. Transformers and pre-trained language representations such as BERT (Bidirectional Encoder Representations from Transformers) transformed the world of NLP by preserving the context and semantic meaning of words in a piece of text. BERT's bidirectional transformer encoding architecture makes the model capable of capturing long-range relationships and subtle patterns in language in text. Using a transformer, there is a possibility to combine qualitative data (e.g., news titles, earnings call transcripts) with quantitative price data in a single predictive framework.

In this work, we introduce a multi modal prediction model that combines a stacked LSTM network with attention mechanisms for time-series analysis and a BERT-based NLP module for textual analysis of reports and news. Through the integration of these two modules, the model aims to leverage both the historical price dynamics and market sentiment or information from news, offering a more unified view of predicting stock movements.

Similar works in the topic are explored to arrive at a better model that is presented here. Then, the data set and the problem statement are presented. The detailed model presented in this paper is showcased in Section IV. Finally, empirical evaluation establishes that the proposed approach performs better compared to baseline models that utilize only time-series or textual inputs, highlighting the potential of technical and textual combination analysis.

II. Literature Review

A. An Hybrid Stacked LSTM Approach to Forecast the Stock Prices [1]

This study tackles stock market prediction by combining historical data with technical indicators to boost accuracy. It introduces a hybrid model that uses Convolutional Neural Networks (CNN) for feature extraction and stacked LSTM networks with dropout for forecasting. Tested on five years of data from HDFC Bank and YES Bank, the model outperforms traditional machine learning approaches in both accuracy and RMSE.

B. Price Wise a Deep Learning Approach to Stock Price Prediction [2]

This study compares deep learning models—specifically LSTM, Bidirectional LSTM, and Stacked LSTM—for forecasting Apple Inc. stock prices from 2017 to 2023. Stacked LSTM delivers the highest accuracy and best captures complex patterns, although it requires more computation time. The research also notes potential in hybrid models and further hyperparameter optimization.

C. Stock Price Prediction Based on LSTM Neural Network: the Effectiveness of News Sentiment Analysis [3]

This paper explores the relationship between public sentiment and stock price prediction by analyzing New York Times news articles. It employs sentiment analysis on news headlines and text, quantifying the sentiment to enhance predictions of future stock prices and returns using an LSTM neural network. The results indicate that, compared to models that only rely on historical stock data, models that incorporate sentiment analysis improve the accuracy, exhibiting a lower RMSE.

D. BELT: A Pipeline for Stock Price Prediction Using News [4]

The paper presents a data-driven approach called BERT-LSTM (BELT) to enhance stock price prediction by integrating real-time Twitter news. BERT is a natural language processing model that was used to analyze relevant tweets, extracting features that indicate potential stock price movements. This model reportedly outperforms existing methods by providing a more accurate prediction of stock price directions, thereby aiding investors during market volatility.

E. Emotion Analysis of News and Social Media Text for Stock Price Prediction using SVM-LSTM-GRU Composite Model [5]

This study explores the relationship between news sentiment polarity and stock prices, demonstrating the effectiveness of using BERT and Support Vector Machine (SVM) regression. This sentiment score, combined with historical stock data, is then analyzed using an LSTM-GRU forecasting model to predict stock prices. The proposed method has a low mean error rate of 13%.

F. Analysis of the Effect of News Sentiment on Stock Market Prices through Event Embedding [6]

This study uses sentiment analysis using a Random Forest classifier to analyze news headlines, achieving an accuracy of 84.92% in predicting sentiment polarity. The stock price data set is normalized and structured into input-output sequences for a bidirectional LSTM (Bi-LSTM) time series forecasting model. Showcasing the potential of NLP techniques in financial predictions.

G. A Comparative Analysis of Share Price Prediction and Trend Direction Using Sentiment Analysis of Financial News Articles [7]

The study focuses on integrating emotional and contextual information into machine learning models, using 30-day stock projections. It utilizes data from reliable sources and employs various machine learning algorithms, including the VADER model for sentiment analysis and the BERT model for pattern identification. The findings highlight the superior performance of the Convolutional Gated Recurrent Unit model with LSTM at second, which achieved an accuracy rate of 84% and an R2 value of 95%.

H. Stock Prediction Using Evolutionary Attention-based LSTM [8]

The study presents a novel deep learning approach for stock prediction called Evolutionary Attention-Based LSTM (EA-LSTM), which improves traditional statistical and machine learning models. EA-LSTM utilizes an attention layer and a competitive random search mechanism, inspired by genetic algorithms, to optimize the attention weights applied to multivariate time series data. The methodology revealed that EA-LSTM outperforms existing models like RNN, GRU, and standard LSTM in terms of prediction accuracy, as measured by RMSE.

I. Prediction of Exact Price of Stock and Direction of Stock Market Using Statistical and LSTM Model [9]

The paper reviews various models for analyzing stock market behavior, including Statistical Analysis, Regression Analysis, SVM, and LSTM networks. The study evaluates these models, finding that the LSTM model outperforms existing methods in forecasting stock market trends.

J. A Survey of Stock Market Prediction - Based on Machine Learning Techniques [10]

The review highlights the application of machine and deep learning techniques, discusses performance metrics, and identifies gaps in current research. The study concludes that models incorporating both technical and deep learning techniques outperform those relying solely on one, with hybrid models, especially LSTM and neural networks, achieving the best results.

III. Dataset Description and Problem Statement

This research addresses the problem of forecasting Tesla (TSLA) stock prices using both financial news text and historical market data. The main issue is whether incorporating textual sentiment and news information can improve the accuracy of stock price prediction compared to using historical prices alone. Tesla was selected as the focus due to its high market volatility and strong media presence. Formally, the task is a time series regression: given recent historical prices and related news, predict the closing price for the next day of TSLA as accurately as possible (minimizing prediction error).

The data set integrates two primary data sources: Financial news data from the FinSen dataset and market price data from Yahoo Finance [12]. The FinSen dataset is a comprehensive collection of economic and financial news articles (spanning many years and countries), including temporal information and sentiment annotations. It has titles, tags, content, and date as columns with around 130000 records in various stocks. For this study, we processed the FinSen data to retrieve Tesla-relevant news, such as articles mentioning *"Tesla"* or *"TSLA"*. This rich textual corpus is well-suited to capture market sentiment and context around the times of Tesla's stock movements. In parallel, we obtained historical stock data for Tesla via Yahoo Finance API, and other sources provide a free range of the daily closing stock market prices. The range chosen for the test was 2021-07-15 to 2023-01-15 for a total of 628 days. The numerical data was normalized using the target, which was the next day's price. Ending with only 2 input parameters, the news content and the normalized numerical value.

By combining these datasets, we form a multimodal time-aligned dataset: for each trading day, we have Tesla's stock metrics and related financial news up to that day. The research goal is to train a predictive model on the data that minimizes the RMSE between the predicted and actual stock prices.

IV. Model Description

We propose a hybrid deep learning architecture that fuses an attention-based bidirectional LSTM network with a BERT-based natural language processing module. An overview of the model's design is as follows. The architecture consists of two components: a DistilBERT-based text encoder processing news articles, and a stacked Bidirectional LSTM (Bi-LSTM) processing the time-series stock data. Followed by an attention layer. Then, to integrate both data types, the model combines the output of the textual and numerical branches using a residual fusion mechanism, followed by a regression layer to generate the final stock price prediction.

For processing textual data, we utilize DistilBERT, a lightweight transformer model distilled from BERT. DistilBERT strikes a balance between performance and efficiency, making it well-suited to our computational constraints while still capturing key language features needed for financial news analysis. We feed the financial news text into DistilBERT to obtain a contextualized embedding. This is a 768-dimensional vector that encapsulates the semantic content and sentiment of the input text. By fine-tuning DistilBERT on our dataset during training, the model learns to emphasize features in the text.

For the numerical modality, we use a stacked Bidirectional LSTM network to process the sequence of recent stock data. The Bi-LSTM layers read the data in both forward and backward directions, capturing patterns that may not be evident when scanning time in only one direction. In the implementation, a two-layer Bi-LSTM stack is used: the first Bi-LSTM layer takes 768 from DistilBert and outputs hidden state vectors (256 total) (for both forward and backward directions), and the second Bi-LSTM layer takes those as input to refine the temporal representation. Each LSTM layer has a hidden state size of 128 units. We also apply dropout regularization between LSTM layers to prevent overfitting, given the relatively limited size of training data in comparison to model parameters.

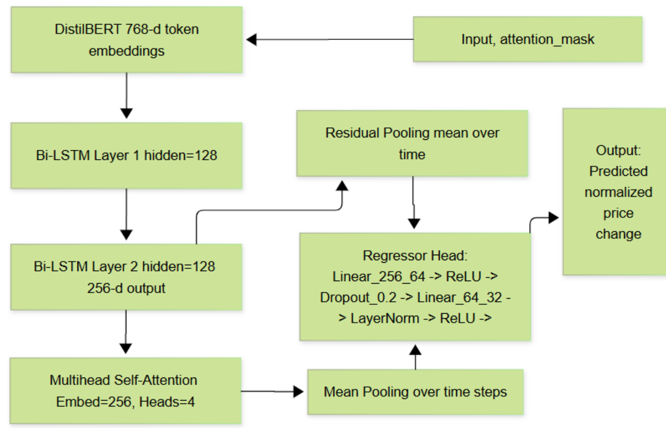

Fig. 1. Model Description

On top of the final Bi-LSTM layer, we incorporate a multi-head attention mechanism. The sequence of hidden states from the Bi-LSTM (one vector per day in the look-back window) is fed into a multi-head self-attention layer. It consists of multiple attention heads (each an independent attention mechanism) that learn to weight the importance of each time step when aggregating the sequence information. For example, attention might learn to place higher weight on more recent days or on days where significant price changes occurred, as these could be more informative for the next day prediction. The outcome of this multi-head attention is a single fixed-size vector (same dimensionality as the LSTM hidden state) that summarizes the historical price movement in an attention-weighted fashion. This attention-enhanced summary vector provides the model with a focused representation of the recent state market.

The core of the approach lies in the fusion of text and time-series representations. Attention is averaged out to context vectors. In addition, the average of the raw LSTM outputs is calculated as residual vectors. Add them together so that on the days where news is very influential, the text vector will carry strong signals that shift the fused representation in a way that affects the prediction. On quieter days, the numerical trend might dominate.

Finally, the model's output is produced by a regressor head. The final regressor head (256/64→ReLU→32→ReLU→1) predicts the next-day stock price. The entire network, from DistilBERT and Bi-LSTM through to the output, is trained using a mean squared error loss function. Model training utilized the Adam optimizer with a learning rate of 2e-5, weight decay of 1e-4, batch size of 32, maximum sequence length of 128 tokens, and a 70/30 train-test split over 15 epochs due to computational constraints.

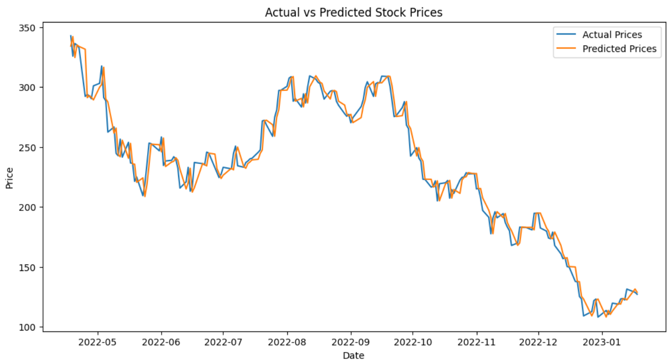

Fig. 2. Testing Data Set Graph

V. Results

We evaluated the proposed hybrid model on the Tesla stock dataset and compared its performance against two baseline models. The first baseline is a Bi-LSTM-only model (essentially the numerical branch of the model with attention but without a BERT text branch). The second baseline is a DistilBERT model (without the stacked LSTM). These comparisons allow us to quantify the contribution of each modality and the effectiveness of the fusion. The main evaluation metrics are RMSE and MAE, which measure the prediction error in the same units as the stock price (RMSE penalizing larger errors more), and the coefficient of determination (R^2), which indicates the proportion of variance in the actual prices that is explained by the predictions. All three models will receive the same input with the same epoch. There was hyperparameter tuning; however, not much testing was done to improve the models further due to computational limitations. Table 1 summarizes the results of each model in the test set.

TABLE I
Prediction Model Results

Techniques	***Proposed Model***	***DistilBERT***	***Stacked Bi-LSTM***
RMSE	9.71	9.80	9.99
R^2	0.971	0.970	0.969
MAE	7.17	7.19	7.34

These results indicate that the fusion model outperforms both single-modality baselines. The hybrid model achieves the lowest error rates. For example, the RMSE of the hybrid model is substantially lower than that of the Bi-LSTM-only model (by 5%), indicating that adding the news information helped reduce prediction uncertainty. It should be noted that they are all within a reasonable range. As seen in “Fig. 2”, the predictions for the proposed model match the actual prices closely. The other models also show very similar graphs. However, improving the model, even by a small margin, becomes a challenge at this point.

The isolation in analyzing baselines provides additional insight. The Bi-LSTM-only model (trained purely on historical prices and technical features) performs the worst, followed by the DistilBERT model. This is likely due to the DistilBert model processing not only the text but also the numerical price. The model does use a regressor head with 2 layers (768/256→1). This outcome suggests that transformers that process news and numerical data outperform traditional LSTM models. The DistilBert model indicates that the sentiment of the news contains some predictive information (the model was able to explain $\tilde{9}7\%$ of the variance).

It is important to note that these results were obtained under certain computational limitations. Training the hybrid model, which involves fine-tuning a transformer and training a deep LSTM simultaneously, is computationally intensive. The model was trained on a single high-end GPU, and each training epoch took on the order of 7 minutes given our dataset size. We also limited the number of training epochs to keep the training time reasonable. These constraints mean that we did not exhaustively search all hyperparameter combinations, so further improvements might be possible with more computational resources (for example, using a larger language model or performing more extensive hyperparameter tuning). Despite these limits, the strong performance of the hybrid model highlights its effectiveness. The results validate our hypothesis that the combination of financial news with price history can improve stock prediction accuracy. In a practical sense, the improved metrics suggest that an investor using our model would have better estimates on the next-day price of Tesla than one relying only on past prices or only on news sentiment.

VI. Conclusion

This study introduced a hybrid deep learning model for stock price prediction that combines historical market data with financial news, using Tesla (TSLA) as a case study. The model integrates an attention-based Bi-LSTM for learning time-based patterns and a DistilBERT-based NLP module to extract sentiment and insights from news articles. By learning from both sources, the model generates more accurate predictions than single-source approaches. Our results show that this dual-input framework significantly improves forecasting performance, achieving lower prediction errors. Key contributions include a novel architecture that fuses text and time-series data, an evaluation of real-world data, and the use of attention mechanisms to improve the interpretability by highlighting key time points in the data. The approach is flexible and can be applied to other stocks or markets where the news heavily influences the price. Future developments will focus on scaling the model to multiple stocks, incorporating new data sources such as social media and earnings call transcripts, and exploring more advanced models like FinBERT and Transformer-based time-series predictors. In general, this research demonstrates how combining natural language processing with time-series analysis can enhance financial predictions and support more informed decision making in dynamic market environments.

References

[1] A. K. Parida, M. Rout, A. K. Jena, N. Parida and R. K. Parida, ”An Hybrid Stacked LSTM Approach to Forecast the Stock Prices,”

2023 OITS International Conference on Information Technology (OCIT), Raipur, India, 2023, pp. 411-415

[2] A. Bansal, A. Singh, S. Roy and K. Agarwal, "Price Wise a Deep Learning Approach to Stock Price Prediction," *2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE), Rourkela, India,* 2024, pp. 1-6

[3] Y. Guo, "Stock Price Prediction Based on LSTM Neural Network: the Effectiveness of News Sentiment Analysis," *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), Chongqing, China, 2020*, pp. 1018-1024.

[4] Y. Dong, D. Yan, A. I. Almudaifer, S. Yan, Z. Jiang and Y. Zhou, "BELT: A Pipeline for Stock Price Prediction Using News," *2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020*, pp. 1137-1146.

[5] R. Kumar, C. M. Sharma, V. M. Chariar, S. Hooda and R. Beri, "Emotion Analysis of News and Social Media Text for Stock Price Prediction using SVM-LSTM-GRU Composite Model," *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India,* 2022, pp. 329-333.

[6] S. Sridhar and S. Sanagavarapu, "Analysis of the Effect of News Sentiment on Stock Market Prices through Event Embedding," *2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS), Sofia, Bulgaria,* 2021, pp. 147-150.

[7] H. Singh and M. Malhotra, "A Comparative Analysis of Share Price Prediction and Trend Direction Using Sentiment Analysis of Financial News Articles," *2023 Global Conference on Information Technologies and Communications (GCITC), Bangalore, India,* 2023, pp. 1-7.

[8] X. Chen, "Stock Prediction Using Evolutionary Attention-based LSTM," *2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA), Dalian, China*, 2023, pp. 1062-1067.

[9] A. Mishra, R. Singh, A. Agrawal, P. Kumar Arya and A. Sharma, "Prediction of Exact Price of Stock and Direction of Stock Market Using Statistical and LSTM Model," *2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), Bristol, United Kingdom,* 2024, pp. 981-985.

[10] K. W. M., S. Allagi and M. Laddi, "A Survey of Stock Market Prediction - Based on Machine Learning Techniques," *2024 Global Conference on Communications and Information Technologies (GCCIT)*, BANGALORE, India, 2024, pp. 1-5.

[11] Chen, P., Boukouvalas, Z. & Corizzo, R., "A deep fusion model for stock market prediction with news headlines and time series data," *Neural Comput & Applic* 36, 21229–21271 (2024).

[12] Liang, Wenhao, Zhengyang Li, and Weitong Chen. "Enhancing Financial Market Predictions: Causality-Driven Feature Selection." *International Conference on Advanced Data Mining and Applications*. Singapore: Springer Nature Singapore, 2024.