Pertanika J. Sci. & Technol. 33 (2): 823 - 843 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Stock Trend Prediction Using Multi-attention Network on Domain-specific and Domain-general Features in News Headline

Phaik Ching Soon*, Tien-Ping Tan, Huah Yong Chan and Keng Hoon Gan

School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

ABSTRACT

In stock market prediction, using news headlines to anticipate stock trends has become increasingly important. Analyzing sentiment from these headlines makes it possible to predict the stock price trends of the targeted company and profit from the resulting trades. This study examines the impact of company-related news headlines on stock price trends. The objectives of this study are as follows: First, we propose a multi-attention network that incorporates the strength of long short-term memory (LSTM) and bidirectional encoder representations from transformers (BERT) to model domain-specific and domain-general features in news headlines to predict the stock price trend of companies. Second, the proposed model can model and evaluate the effect of news on the stock price trend of different companies. Third, we construct the Bursa Malaysia news headline dataset and automatically align headlines with target companies and their stock price trend. This study proposes that the LSTM WITH ATTENTION +BERT model should use domain-specific and domain-general features to predict stock price trends using news headlines. The proposed model is compared to several convention models and deep learning models. The LSTM WITH ATTENTION +BERT model achieved an accuracy of 50.68%, showing notable improvements over other approaches. It surpassed the Decision Tree by 11.2%, Naïve Bayes by 20.13%, and Support Vector Machine by 5.12%. Compared to the CNN, LSTM, and BERT models, the proposed model is 4.27%, 2.91% and 1.64% higher, respectively, in terms of accuracy. These results highlight the strength of the proposed model.

Keywords: Domain-general features, domain-specific features, multi-attention network, news sentiment analysis, stock price trend

ARTICLE INFO

Article history:
Received: 07 May 2024
Accepted: 18 November 2024
Published: 07 March 2025

DOI: https://doi.org/10.47836/pjst.33.2.13

E-mail addresses:
phaikchingsoon@student.usm.my (Phaik Ching Soon)
tienping@usm.my (Tien-Ping Tan)
hychan@usm.my (Huah Yong Chan)
khgan@usm.my (Keng Hoon Gan)
* Corresponding author

INTRODUCTION

Classification is a fundamental task in machine learning, and deep neural networks have demonstrated notable advancements in classification. In text classification,

ISSN: 0128-7680
e-ISSN: 2231-8526

pre-trained large language models such as bidirectional encoder representations from transformers (BERT) have been used for transfer learning in tasks like sentiment analysis (Kabbani & Duman, 2022) and translation (Prachyachuwong & Vateekul, 2021). The idea of pretrained transformers in natural language processing is then introduced in image recognition, where we see the Vison Transformer in object recognition (Dosovitskiy et al., 2021) and facial recognition (Vaswani et al., 2017). In addition, in speech processing, similar developments can be seen where models such as Wav2Vec and WavLM have achieved remarkable performance in speech processing tasks such as automatic speech recognition (Karpagavalli & Chandra, 2016) and speaker diarization (Vaswani et al., 2017). However, despite these successes, applying deep learning models in financial markets, such as stock trend prediction, remains challenging. Financial markets involve complex dynamics, non-linear patterns, and high-frequency data due to changing government policies, companies' earnings, raw prices, and external events, making accurate prediction and classification difficult.

Investing and Trading

In general, investing in the stock market involves purchasing high-quality, undervalued stocks and holding them for the long term, as opposed to short-term trading. The underlying principle of this approach is the belief that the market is not efficient. Thus, it has not accurately assessed the true worth of these investments, creating an opportunity for future capital appreciation. Furthermore, by conducting a rigorous analysis of a company's financial statements, earnings, assets, and competitive advantages, value investors endeavor to uncover opportunities where the intrinsic value of a stock has been overlooked by the market (Azhikodan et al., 2019)

On the other hand, trading involves more frequent buying and selling of stocks. The main idea in stock trading is to identify the trend of stock so that a trader can make maximum profit by buying it when the price is going to go up and selling it before the price of the stock goes down.

An example of trading using stock price analysis is candlestick patterns, which help to identify potential uptrends or downtrends in stock price (Figure 1). For instance, the hammer pattern indicates a possible uptrend, while the shooting start pattern indicates a possible downtrend. Besides, technical indicators such as simple moving average (SMA), Moving Average Convergence

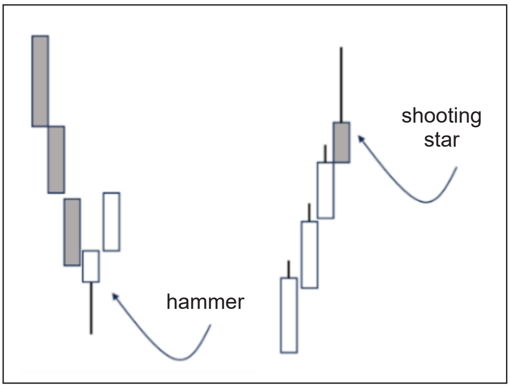

Figure 1. Example of candlestick chart patterns for stock trend prediction (Santur, 2022)

Divergence (MACD) and Relative Strength Index (RSI) are also used by traders to predict the stock price trend. Many machine learning studies in stock trend prediction also computed technical indicators as features for their models (Nguyen & Yoon, 2019)

News Headlines

News articles may influence the direction of stock prices (Hu et al., 2018; Akita et al., 2016; Liu & Carrio, 2018). In general, positive sentiment news will lead to an uptrend in the stock price and vice versa. There are many types of news. One of the types of news that have a major influence on stock price is company-related news. Here, we define company-related news as news that mentions one or more publicly listed companies. Before delving into stock trend prediction, it is essential to first understand the different types of company-related news. The company-related news can be categorized into three groups: quarterly reports, financial analyst reports, and events. Table 1 illustrates company news headlines and their corresponding stock price patterns on Bursa Malaysia.

Quarterly results refer to the financial performance of a company over three months, typically aligned with the company's fiscal quarters. Consistent revenue growth may indicate a healthy business environment and potentially lead to a positive stock trend. For instance, the news headline, "Zhulian 3q profit jumps 138%," indicates a substantial increase of 138% in profit compared to the corresponding period in the previous year. The term "profits" normally indicates a positive sentiment in the specified company. On the other hand, the news headline, "Zhulian falls 11.01% after 4q earnings decline 44%," indicates a negative sentiment, which is reflected in the stock trend.

Table 1
Examples for all news types and stock trends

Type of Company-Related News	Examples of News Headlines	Company Involved	Stock Trend
Quarterly Result	Zhulian 3q profit jumps 138%	ZHULIAN	Up
	YTL Power 4q lifted by deferred tax credit recognition	YTLPOWR	Flat
	Zhulian falls 11.01% after 4q earnings decline 44%	ZHULIAN	Down
Financial Analyst Report	AppAsia may climb higher, says RHB Retail Research	APPASIA	Up
	AirAsia among airlines able to weather COVID-19 crisis, says Brand Finance	AIRASIA	Flat
	Zecon at higher high but ended weak, says AllianceDBS Research	ZECON	Down
Events	AirAsia increases the number of flights to meet holiday demand	AIRASIA	Up
	Yusri Md Yusof is the new Green Ocean MD as Tan See Meng quits post	GOCEAN	Flat
	Zelan falls 4.76% after unit faces RM16.12m arbitration claim	ZELAN	Down

In addition, financial analyst reports are documents prepared by financial analysts who study and analyze companies, industries, and financial markets. Analyst reports offer insights into the company's financial prospects, investment opportunities, and risks, helping investors and stakeholders make informed decisions about buying, selling, or holding a particular security. Table 1 shows the headline "AppAsia may climb higher, says RHB Retail Research." RHB Retail Research suggested that AppAsia may climb higher, indicating an uptrend in the stock price. This could be based on their analysis of the company's financials, market conditions, or other relevant factors that lead them to believe the stock is likely to rise. A financial analyst may also give an underweight rating on a stock, with a headline such as, "Zecon at higher high but ended weak, says AllianceDBS Research ."Financial analyst reports may give a rating on a particular stock based on the performance of the company and the stock prices. These recommendations can influence investor sentiment and impact the company's stock price trend.

A company event is a significant occurrence or development reported in the media and may potentially impact the specified companies and industries. These news events can include corporate mergers or acquisitions, product launches, legal disputes and natural disasters. Positive news, such as successful product launches, can drive stock prices higher and vice versa. For example, in Table 1, "AirAsia increases the number of flights to meet holiday demand," the news snippet indirectly suggested that increasing flights may positively impact the company's revenue and reflect the stock price. On the other hand, the news headline, "Zelan falls 4.76% after unit faces RM16.12m arbitration claim," gave investors a negative sentiment toward Zelan due to possible loss in revenue due to the lawsuit. Nevertheless, not all company news has an impact on the stock price. As shown in Table 1, "Yusri Md Yusof is the new Green Ocean MD as Tan See Meng quits post. "The market might not see the leadership change as a substantial factor influencing the company's financial prospects or overall direction. In such cases, the stock may experience a flat or neutral trend.

Research Questions and Contributions

Company-related news headlines have a significant impact on stock price trends, with positive news leading to an increase in stock prices and negative news leading to a decrease in stock prices. The sentiment and timing of the news publication influence the magnitude and direction of the stock price movement. This hypothesis sets the foundation for exploring the relationship between news sentiment and stock price fluctuations.

The research questions of this study are as follows:

- How do we model and predict stock price trends given a news headline?
- How do we model and predict stock price trends of different listed companies given a news headline?

- How does company-related news affect the stock price trend of listed companies in Bursa Malaysia?

The main contributions of this study are summarized as follows:

- We propose a multi-attention network that incorporates the strength of LSTM and BERT to model domain-specific and domain-general features in news headlines to predict the stock price trend of a company.
- The proposed model can model and evaluate the effect of news on the stock price trend of different companies. For example, in "Samsung sues Apple," the effect of this news headline will be different for Samsung and Apple.
- We construct the Bursa Malaysia news headline dataset by collecting news headlines using a web crawler and automatically aligning headlines with target companies and their stock price trends from Bursa Malaysia. The dataset will be released to GitHub.

News Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique used to evaluate the sentiment or emotional tone expressed in a text. Before conducting sentiment analysis, it is crucial to perform feature extraction on news data. Feature extraction can be broadly classified into frequency-based and vector-based approaches, each offering distinct advantages for capturing the context within news content.

Frequency-based feature extraction is well-suited for traditional machine learning models like Decision Trees, Support Vector Machines (SVM), and Naïve Bayes. This approach relies on normalized word counts and statistical measures such as frequency, probability, and term frequency-inverse document frequency (TF-IDF) as feature representations. For example, news articles or headlines are often transformed into a bag-of-words model, from which feature vectors are derived. These vectors are then used in machine learning algorithms for tasks like sentiment analysis or stock trend prediction. Agarwal et al. (2020) extracted word n-grams as features from the news; Tipirisetty (2018) implemented VADER to extract news features for model training. Similarly, Jariwala et al. (2020) utilized TF-IDF for news feature extraction.

Vector-based feature extraction, on the other hand, typically involves the use of word embeddings, which are often combined with neural network models. Word embeddings represent words as high-dimensional vectors that capture contextual meaning. These embeddings are generated using methods like autoencoders, which predict neighboring words (e.g., skip-gram) or target words (e.g., Continuous Bag of Words, CBOW). Compared to frequency-based methods, word embeddings are more effective at capturing semantic relationships between words. For example, Kaur (2017) implemented a neural network with pre-trained Word2Vec embeddings for sentiment analysis of news related to a specific company. Hu et al. (2018) used CBOW embeddings to extract context from news articles.

Akita et al. (2016) implemented a combination of Paragraph Vectors for distributed representations of news articles and Long Short-Term Memory (LSTM) to model temporal effects on stock price. The Paragraph Vector model was utilized to generate continuous distributed representations for each news article and map variable-length text to fixed-length vectors using the Distributed Memory Model (PV-DM) and the Distributed Bag of Words model (PV-DBOW) simultaneously. Articles about 10 companies were represented as vectors and concatenated into a group article vector, with fixed positions assigned to each company; if no article is available for a company at a given timestep, a zero vector is inserted, and multiple articles for the same company are averaged. This approach assumes one article per company per timestep, which is addressed by inserting zero vectors or averaging multiple articles. Experiments conducted on data from fifty companies on the Tokyo Stock Exchange demonstrated that this method outperforms numerical-only and Bag-of-Words-based approaches. LSTM effectively captures time series influences, and considering industry-wide correlations improves stock price prediction.

Although CNN is commonly used on spatial data, such as image categorization and image segmentation, Liu and Carrio (2018) applied it to stock trend prediction. The researchers proposed a joint model combining the TransE model (knowledge graph into a continuous vector space) for representation learning and a CNN for feature extraction from financial news to improve stock price prediction. It also integrated daily trading data and technical indicators. Evaluated with SVM and LSTM, the model achieved 97.66% accuracy in news sentiment classification and 55.44% accuracy in predicting Apple's stock movement using the S&P 500 index. The joint learning approach outperformed traditional methods, providing better decision support for investors.

Similarly, Hu et al. (2018) utilized Hybrid Attention Networks (HAN) with a self-paced learning mechanism. HAN integrated several components to address the challenges posed by the volatile nature of the stock market and the variability in online news quality. It started with news embedding, where news articles were transformed into vectors using pre-trained Word2Vec embeddings. The news-level attention mechanism assigned importance weights to each news item based on its relevance, using a one-layer network and softmax function. For sequential modeling, Gated Recurrent Units (GRU) encoded the temporal sequences of news vectors, capturing both past and future contexts through bi-directional outputs. The temporal attention mechanism further refined this by weighing news over time to enhance classification. To tackle the difficulty of varying news quality, the framework employs Self-Paced Learning (SPL), which gradually incorporates challenging and noisy samples into the training process using a linear regularizer. The model's performance had an accuracy of 48%. However, it focused only on specific companies.

However, traditional word embeddings have a limitation: each word is represented by a single static vector, which does not account for varying contexts. Contextualized word embeddings have been developed to address this. Examples include Embeddings from Language Models (ELMo) and Bidirectional Encoder Representations from Transformers (BERT), which generate dynamic word embeddings based on the context in which a word appears, offering improved performance in sentiment analysis. Transformers leverage attention mechanisms to capture relationships within a sequence, enabling more effective handling of context and dependencies in text. Transformer-based models, such as Bidirectional Encoder Representations from Transformers (BERT), have recently emerged as a powerful financial news sentiment analysis tool.

Kabbani and Duman (2022) implemented a fine-tuned BERT model (FinBERT) for news sentiment analysis into trading strategies by calculating sentiment scores to enhance trading decisions. The process involved using a rule-based method to identify relevant headlines for an asset, applying the FinBERT model to determine the sentiment (positive, negative, or neutral) of each headline, and then averaging the sentiment scores for each asset daily. If the positive sentiment outweighed the negative, a score of +1 is assigned; otherwise, -1 is given. Neutral sentiments were disregarded under the assumption that any news mentioning an asset affects its price.

On the other hand, Prachyachuwong and Vateekul (2021) processed each headline using the BERT model with Multilingual Cased pre-training weights to adapt it to a specific headline corpus. The researchers used the first token from the BERT-Base output, typically employed for classification tasks, to represent the headlines. The model was trained until the convergence of the loss function, adhering to the BERT architecture workflow. Lastly, the optimal embedding values are retained and consolidated for each sector to represent the most relevant headlines of the day. Although BERT captures rich, context-dependent word meanings and performs well in classifying sentiments, it lacks domain-specific features.

Referring to Table 2, the limitations of those models are that they are trained on specific datasets without focusing on company names during feature extraction on news headlines, and the limitations of BERT are that they are trained on more generalized data. Liu and Carrio (2018) used a CNN + LSTM model focused on domain-specific tasks, excelling in spatial feature extraction and integrating news and stock price features during model training, though it was limited to sequence modeling and single-company predictions. Similarly, Akita et al. (2016) employed a Paragraph Vector + LSTM model to handle company-specific news, effectively combining news and stock price features, but encountered challenges when training data lacked news for certain companies. On the domain-general feature extraction, Hu et al. (2018) introduced HAN, benefiting from self-paced learning to enhance generalization but faced difficulties with non-financial news introducing noise. Kabbani and Duman (2022) leveraged BERT to capture nuanced

Table 2
Strengths and weaknesses of reviewed models

Author	Model	Domain	Strengths	Weaknesses
Liu and Carrio (2018)	CNN + LSTM	Domain-specific	Spatial feature extraction. Extract news features and stock price features on model training	Limited sequence modelling. Stock prediction on a single company
Akita et al. (2016)	Paragraph Vector + LSTM	Domain-specific	Handles news paragraph for selected company Extract news features and stock price features on model training	There is no news for a few companies on model training.
Hu et al. (2018)	HAN	Domain-general	Self-paced learning improves generalization.	Non-financial news causes noise.
Kabbani and Duman (2022)	BERT	Domain-general	Captures nuanced word meanings. Pretrained model with a large set of data.	Lack of local news context (such as company names)
Prachyachuwong and Vateekul (2021)	BERT	Domain-general	Multilingual pre-trained data	Specific for Thai news feature extraction

word meanings and took advantage of pre-trained models on large datasets, though it lacked sufficient local news context, particularly in company names. Prachyachuwong and Vateekul (2021) also utilized BERT with a multilingual dataset, focusing specifically on Thai news feature extraction.

Overall, vector-based approaches extract more context in vector form compared to frequency-based context extraction, but they still encounter challenges in addressing deeper contextual relationships within financial news regarding company names. However, without an attention mechanism for company names, the model cannot focus on the most critical information about company names appearing on the news headlines for stock trend prediction.

In addition, most researchers primarily concentrate on predicting stock trends for a limited number of specific companies. However, Akita's study is an exception, as it forecasts stock trends for all companies listed on the Tokyo Stock Exchange using data from the Nikkei Newspaper. We will employ a direction similar to that of Akita to forecast stock trends for all companies listed on Bursa Malaysia, utilizing news data from The Edge Newspaper.

MATERIALS AND METHODS

This study is conducted on companies listed in Bursa Malaysia to model the relationship between news headlines and stock price trends. We only focus on company news in this study. The reasons are threefold: firstly, company news often affects the stock price of the target companies. Thus, it is easier to analyze the causal and effect. Secondly, the target companies that are related to a news headline are often mentioned, and it is easier to find out the target companies. Thirdly, the number of listed companies in Bursa Malaysia is more than eight hundred, and the amount of news data is not a lot. Thus, the amount of data may not be enough to robustly model the many-to-many relationship between a company and some news and between news and the stock price trend. Nevertheless, the approach proposed can also be applied to another type of news, given that sufficient data is available.

News Data

We crawl the financial news from The Edge Markets (which has been renamed The Edge Malaysia) since it is the leading financial news publisher in Malaysia, and the latest events are published very quickly online compared to other media. After a news article is crawled, we extract the news headlines and publication date and time from it. The target companies mentioned in the news headline are extracted using regular expressions. The news headlines are crawled and downloaded from The Edge Markets, September 2018 until September 2021. The news articles are first downloaded, and then the news headlines are extracted. The news headline must contain at least one company to retrieve the news headline into our data. Overall, 24388 news headlines with company names are extracted for stock trend prediction.

Stock Trend Annotation

The end-of-day (EOD) stock prices of listed companies in Bursa Malaysia, which consists of 883 companies, are retrieved from Yahoo Finance from March 2016 until March 2021.

Table 3 delineates distinct conditions used to label the stock price trend based on the news publication date and time. The general intuition of labeling a stock price trend (as uptrend, downtrend, or flat) is to get the publication time of news and then get the stock price before and after publication. If the relative change is +1% or more, then it is labeled as an uptrend; if the change is -1% or less, it is labeled as a downtrend. Otherwise, it is flat.

In the first scenario, if the news is reported during weekdays (D_i) between 9 am and 5 pm (that is Bursa Malaysia operating hours), the stock prices of the target companies, consisting of the Open price (D_i Open) and Close price (D_i Close) of the day, are retrieved and set as Price Before and Price After, respectively.

Table 3
Published time conditions to retrieve stock price

Published Day	Conditions	Price Before	Price After
D_i, 9 am–5 pm	D_i: Weekday	D_i Open	D_i Close
D_i, 5 pm–D_{i+1}, 9 am	Both days: Weekday (except Friday and holiday)	D_i Close	D_{i+1} Close
D_i–D_j, 9 am	D_i: Holiday, D_j: Weekday	D_{i-n} Close	D_j Close
D_i, 5 pm–D_j, 9 am	D_i: Friday, D_j: Monday	D_i Close	D_j Close

* Day: D, Open: Open Price, C: Close Price

In the second scenario, if the news is reported outside Bursa Malaysia's operating hours, which are between 5 pm and 9 am on the subsequent day (D_{i+1}), during weekdays but not Fridays and holidays, the stock prices of the target companies consist of the Close price of D_i (D_i Close) and the Close price of D_{i+1} (D_{i+1} Close) are retrieved and used as the stock Price Before and Price After, respectively, for calculating the stock price trend.

In the third scenario, if the news is reported between a holiday (D_i) and the subsequent weekday (D_j), the stock prices of the target companies, consisting of the Close price on the holiday (D_{i-n} Close) and the Close price on the subsequent weekday (D_j Close), are retrieved and set as the Price Before and Price After.

In the fourth scenario, if the news is reported from 5 pm on Friday (D_i) until 9 am on Monday of a new week (D_j), the stock prices of the target companies, consisting of the Close price on Friday (D_i Close) and the Close price on Monday (D_j Close), are retrieved and set as the Price Before and Price After.

With the Price Before and Price After, the relative Change formula in Equation 1 calculates the percentage change in a stock price before and after the publication of a news article. The following criteria are employed to classify news headlines according to stock trends: a relative change greater than 0.01 is designated as an uptrend, while a relative change less than -0.01 is categorized as a downtrend. A relative change falling between -0.01 and 0.01 is considered a flat trend.

$$Relative\ Change = \frac{Price\ After - Price\ Before}{Price\ Before} \quad [1]$$

To illustrate, the quarterly report "Zhulian 3q profit jumps 138%" was published on 2017-10-11 at 8:08 pm. The second scenario applies to the time the news was published. The Price Before the closing price of the stock on 2017-10-11 at 5 pm was RM 1.64, and the Price After was the closing price on the following day, 2017-10-12, which was RM 1.79. Applying the relative change equation, we got Relative Change = (1.79 – 1.64) /1.64 = 0.01; a relative change of 0.01 is classified as an uptrend. This example demonstrates how the specified conditions in Table 3 and Equation 1 are utilized to calculate and interpret stock price movements.

News Trend Analysis

After annotating the stock price trend for the news headlines, we do a preliminary analysis by dividing the company news into the following groups: quarterly results, financial analyst reports, and news events.

The 24388 data samples have 2954 annual reports, 1707 financial analyst reports, and 19727 events. In news headlines, the company names are extracted by pattern matching the name in the company name list extracted from Bursa Malaysia 2021. If there is more than one name listed in the news headlines, each of the companies in the news headlines is analyzed as a sample. For example, On December 21, 2020, the news “Kejuruteraan Asastera (KAB) wins three contracts worth RM57 m from Mah Sing” involved more than one company: KAB and Mah Sing. From this news headline, a total of 2 samples are created for the headline with different target company names. KAB has secured three contracts from Mah Sing Group, totaling RM57 million. This suggested positive developments for KAB as it had successfully won new business. The stock trend was calculated from stock price by referencing the company ticker, date, and time of the news. This news positively affected KAB company, as it was an uptrend for KAB with a relative change of 0.01942, but Mah Sing showed a downtrend.

Table 4 shows the analysis of various types of company news headlines and their minimum and maximum magnitude changes. The Absolute Relative Change is calculated using Equation 2 to quantify the changes in stock price due to news.

$$Absolute\ Relative\ Change\ = Abs[\frac{(Price\ After - Price\ Before)}{Price\ Before}] \qquad [2]$$

From our analysis, news related to annual reports makes a maximum of 0.5111 magnitude change. This shows that the most influential financial disclosures can result in a more than 50% change in stock prices, underscoring their significance in shaping investor perceptions. The lowest magnitude of 0.000 indicates that certain annual reports have no apparent impact on stock prices. On average, annual reports led to a relatively modest 2.34% change, highlighting the variability in market reactions.

Financial analysis reports, with the highest magnitude of 0.2964, point to a 29.64% potential change in stock prices following the most impactful analyses, showcasing the substantial influence of expert financial assessments. Instances where financial analysis reports have no impact are denoted by the lowest magnitude of 0.000, revealing the diverse responses to such analyses. The average magnitude of 0.0191 implies a comparatively smaller change on average.

Events exhibited the biggest potential magnitude change of 2.600, indicating more than two increases/decreases in stock prices following the most impactful event occurrences. The average magnitude of 0.0275 suggests that an event causes more than 2% of changes in the stock price of the target company. This analysis gives an insight into the potential

Table 4
Data analysis on different types of company news headlines

Type of News Headlines	Highest Magnitude	Lowest Magnitude	Average Magnitude
Annual Reports	0.5111	0.000	0.0234
Financial Analysis Report	0.2964	0.000	0.0191
Events	2.600	0.000	0.0275

rewards and risks associated with different types of news. It also shows the importance of stock trend prediction using news.

Proposed Multi-attention Network

This study proposed a multi-attention network with domain-specific and domain-general features to predict stock price trends given a news headline and the target company. The model learns the effect of the news on the target companies' stock price trend, recognizing that the effect of a headline like 'Samsung sues Apple' varies for Samsung and Apple.

The steps to predict the stock price trend given a news headline and a company name are as follows. First, a simple name matching using regular expressions matches the names of Bursa-listed companies with the news headline. If one or more companies are found in the headline, both the news headline and company names are input to the proposed multi-attention network consisting of an LSTM with an attention model and a BERT-based model to learn the relation between the stock price trend of the target company. LSTM with attention is implemented to model domain-specific features, where financial domain word embedding is trained using the financial text. On the other hand, the BERT is a pretrained large language model where we leverage the general-domain contextualized word embedding features to learn the relation between the stock price trend of the target company and the news headline associated across multiple positions through the mechanism of attention heads and layers. This complex problem is well-suited for Transformers that capture long-range dependencies and global patterns effectively. The predictions from the LSTM-based model and BERT-based model are normalized and combined to produce the outcome. The following paragraphs explain the proposed model in detail, referring to Figure 2.

The news headline and company name are converted to domain-specific word embedding vectors and input into LSTM encoders. Refer to the left neural network in Figure 2, 2a. A domain-specific word embedding model is first pretrained with texts from the financial domain using a word embedding to train the embedding vectors v_h and v_c from the text corpus. In $LSTM_h$, it inputs the v_h and outputs the hidden states h. In $LSTM_c$, it receives the v_c as input and outputs the hidden states c. The company attention distributions are calculated with the hidden states from the news headlines encoder and target company encoder using a dot product that produced an attention matrix with a size of N x M, where N is the length of news headlines vectors, and M is the length of the company name vector. This shows the importance of the words/tokens in the news headline. The attention weights are then normalized using the SoftMax function and then concatenated to the hidden states output from the LSTM news headline encoder. It weighs the importance of company names on news headlines.

From the LSTM encoder, the output vectors go through a dense layer followed by a convolutional neural network (CNN) consisting of a 1D convolutional layer and a max pooling layer. The dense layer processes the flattened tensor by connecting each element to every neuron, enabling the model to learn complex relationships and patterns across the input. The output of the dense layer is then input into the Softmax layer for label classification, transforming logits into a probability distribution over three classes (downtrend, flat, and uptrend), followed by normalization.

At the same time, the news headline and company name are used to fine-tune Bidirectional Encoder Representations from Transformers (BERT) to predict stock price trends. BERT is a powerful pre-trained transformer-based language model that captures contextual information and relationships within text. To model the target company name and news headline, we concatenate the company name with a delimiter (e.g., '|') and then append it to the news headline. Refer to Figure 2, 2b; following this concatenation, the resulting combined representation undergoes processing through a series of standard transformer blocks. These blocks consist of self-attention and feedforward layers, normalization layers, and residual connections (Jurafsky & Martin, 2022). Subsequently, the output vector passes through a dense layer, applying Softmax and normalization.

The final output of the network is derived through a two-step process. First, the normalized Softmax output from both the LSTM-based model and the BERT-based model are combined using an element-wise addition. This operation results in a fused probability distribution across the classes, referring to Equation 3. λ is the normalization weight, where

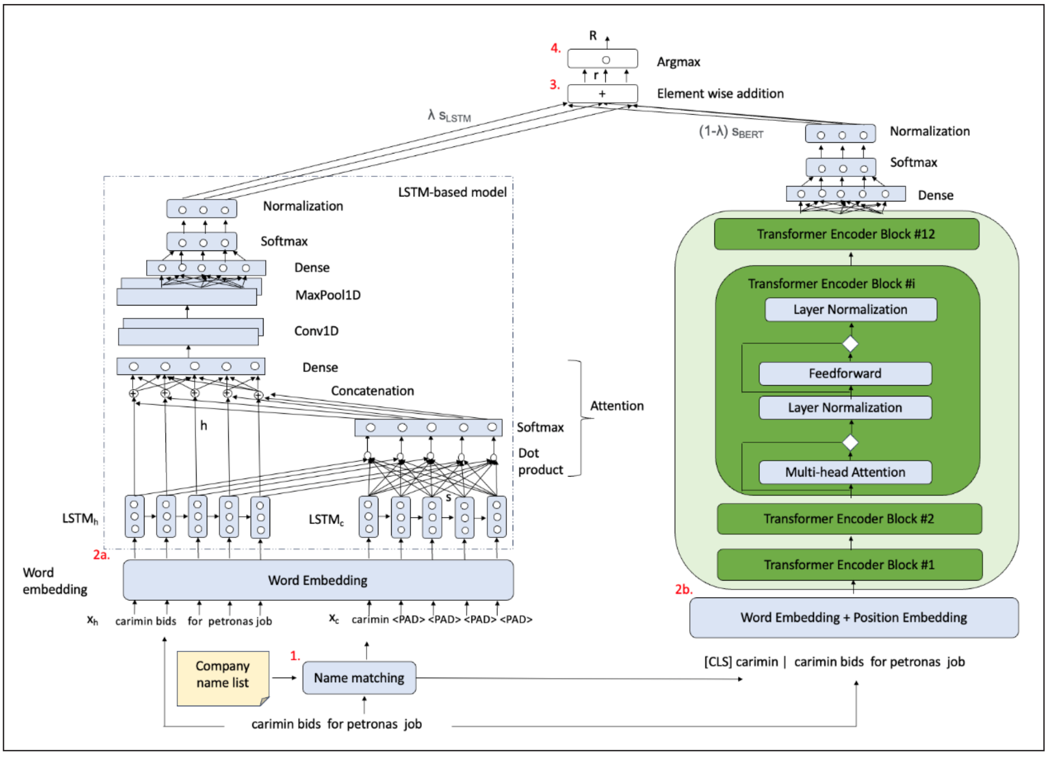

Figure 2. Proposed multi-attention network (LSTM + BERT) with domain-specific and domain-general features

$0 \leq \lambda \leq 1$, which can be determined experimentally from the validation data. λs_{LSTM} and $(1-\lambda)s_{BERT}$ are the probability distributions of the classes predicted from the LSTM-based and BERT models, respectively. Subsequently, the argmax function is applied to determine the joint prediction, referring to Equation 4. The class with the highest probability is selected as the final output (R). The proposed model allows the integration of predictions from both models, providing a comprehensive and combined assessment of the domain-specific features of the LSTM-based model and the domain-general features of the BERT-based model in determining the most likely class for the given input data.

$$r = \lambda\ s_{LSTM} + (1-\lambda)\ s_{BERT} \quad [3]$$

$$R = \mathrm{argmax}(r) \quad [4]$$

RESULTS AND DISCUSSION

A total of 24,388 samples were extracted, each corresponding to a specific company mentioned in the news headlines. For instance, the headline "Zhulian 3Q profit jumps 138%" was associated with Zhulian. The data extraction would produce two samples if a headline contained two different companies. The two samples consisted of the same headline with different target companies. Based on these news headlines, the model was trained to predict stock price trends for multiple companies. In 24388 data, it was split according to the date and time into a training set (17072), test set (5739), and validation set (4877), as shown in Table 5. The distributions of the data were as follows. The train set consisted of 4723 downtrends (0.2767), 6146 flat trends (0.3600), and 6203 uptrends (0.3633). On the other hand, the test set consisted of 645 downtrends (0.2644), 4438 flat trends (0.4666), and 656 uptrends (0.2690). The validation set consisted of 1434 downtrends (0.2940), 1984 flat trends (0.4068), and 1459 uptrends (0.2992).

Table 5
Count and distribution of news headlines dataset according to stock price trend

Trend	Train	Test	Validation	Total
Downtrend	4723 (0.2767)	645 (0.2644)	1434 (0.2940)	6802
Flat	6146 (0.3600)	1138 (0.4666)	1984 (0.4068)	9268
Uptrend	6203 (0.3633)	656 (0.2690)	1459 (0.2992)	8318
Total	17072	5739	4877	24388

Baseline Machine Learning Models

Table 6 presents the results of stock trend predictions using conventional machine learning algorithms, including Decision Trees, Naïve Bayes, and Support Vector Machine. The metrics used to evaluate the models were accuracy, precision, recall, and F1. Notably, the

Support Vector Machine model outshone its counterparts, boasting the highest accuracy at 0.4596 and F1 score of 0.4171. In contrast, the Naïve Bayes model exhibited the lowest accuracy and F1 score, standing at 0.3055 and 0.3621, respectively. Although Naïve Bayes could be computationally efficient and straightforward to implement, its performance was hampered by the assumption of feature independence and its inability to capture more intricate relationships within the data. The Decision Tree model, with an accuracy of 0.3916, demonstrated a balance between precision (0.3682) and recall (0.3688), yielding an F1 score of 0.3675. Decision Trees could be prone to overfitting, particularly in the context of complex and variable data such as text data from news headlines, which might result in a lower overall predictive accuracy. Overall, the Support Vector Machine emerged as the most effective conventional machine learning algorithm for stock trend prediction, offering a well-balanced performance across multiple evaluation criteria, as it could handle high-dimensional data, and its robustness to outliers contributed to its superior performance.

Table 6
Results of conventional machine learning models on stock trend prediction

Models	Accuracy	Precision	Recall	F1
Decision Tree	0.3916	0.3682	0.3688	0.3685
Naïve Bayes	0.3055	0.3729	0.3519	0.3621
Support Vector Machine	**0.4596**	**0.4265**	**0.4082**	**0.4171**

Deep Neural Networks

Next, we evaluated three deep neural networks: the CNN-based model, the LSTM-based model with attention, and the BERT-based model for stock trend prediction using news headlines and the target company as the input. Both the CNN-based model and LSTM-based model with attention use word embedding vectors trained using GloVe (Souma et al., 2019). We trained the domain-specific GloVe word embedding vectors using the news articles from The Edge Markets. About 16 MB of text data was used, and the vocabulary size was about 40 thousand.

The CNN-based model consisted of two CNN encoders that receive word embedding vectors as input, one from the news headline and another from the target company. The output from the encoders was joined using a dot product to form an attention matrix; then, the attention matrix was concatenated. The attention matrix is concatenated and passed through a dense layer, followed by a convolutional neural network (CNN). Each CNN comprised a convolutional layer with a filter size of 32 and a kernel size of 3, followed by max pooling with a pool size of 2. The flattened layer reshaped the tensor into a 1D vector, removing spatial structure while retaining sequential information. A dense layer processed the flattened tensor to capture complex relationships and patterns, and the output was fed into a Softmax layer for stock trend label classification. The model used a learning rate 0.0001 and trained for 8 epochs with early stopping to prevent overfitting.

The LSTM-based model consisted of the left branch of the proposed model in Figure 2, 2a. Two LSTM models, $LSTM_h$ and $LSTM_c$, were utilized to process news headlines vectors (v_h) and targeted company names (V_c). $LSTM_h$ took v_h as input, while $LSTM_c$ received v_c as input. This operation involved combining the encoder outputs and decoder outputs to create joined outputs using dot products to form an attention matrix. The attention matrix was concatenated and passed through a dense layer, followed by a convolutional neural network (CNN). Each CNN comprised a convolutional layer with a filter size of 32 and a kernel size of 3, followed by max pooling with a pool size of 2. The flattened layer reshaped the tensor into a 1D vector, removing spatial structure while retaining sequential information. A dense layer processed the flattened tensor to capture complex relationships and patterns, and the output was fed into a Softmax layer for label classification (downtrend, flat, uptrend). The model used a learning rate 0.0001 and trained for 8 epochs with early stopping to prevent overfitting.

While the BERT-based model consisted of the right branch of the proposed model in Figure 2, 2b, the encoded data, which consists of the company name and headline, are transformed into tensors, encompassing input IDs, attention masks, and corresponding labels for classification. Leveraging the Hugging Face Transformers library, the "bert-base-uncased" variant was chosen for its versatility in handling various NLP tasks. The chosen model architecture for this study was BertForSequenceClassification, a variant of BERT specifically designed for sequence classification tasks. It pre-trained on large language corpora, allowing it to capture intricate patterns and relationships within text. For efficient model training, the dataset tensors were organized into DataLoader objects with a batch size 8. The AdamW optimizer, specifically designed for transformer models, was employed with a learning rate of 0.00001. The training process was executed for a single epoch, considering the specific requirements of the task and computational constraints. The output was classified into three classes: downtrend, flat trend, and uptrend.

Table 7 presents the experiment results of deep neural networks for stock trend prediction using news headlines. All the deep neural networks performed better than conventional machine learning models in terms of accuracy and F1 score. The BERT-based model achieved the highest accuracy at 0.4903, surpassing the accuracy of the LSTM-based model with attention and the CNN-based model at 0.4777 and 0.4641, respectively. Nevertheless, the LSTM-based model with attention produced a higher F1 score at 0.4472 compared to the BERT-based model at 0.4396 and the CNN-based model at 0.4290.

The CNN-based model demonstrated the lowest performance in terms of both accuracy and F1 score compared to other models. In contrast, the LSTM-based model with attention achieved the highest F1 score of 0.4472, while the BERT-based model attained the highest accuracy at 0.4904. CNN models were well-suited for capturing spatial and local patterns

Table 7
Results of deep neural networks on stock trend prediction using news headlines

Deep Neural Network	Accuracy	Precision	Recall	F1
CNN-based model	0.4641	0.4667	0.3970	0.4290
LSTM-based model with attention	0.4777	0.4453	0.4492	**0.4472**
BERT-based model	**0.4904**	0.4545	0.4256	0.4396

in data, such as images or short text sequences. However, their ability to capture semantics in long sequential data, like news headlines, was limited compared to LSTM and BERT models. This likely explained CNN's weaker performance in this task. Both the LSTM-based attention and CNN-based models were trained using word embeddings from a small set of domain-specific data. The LSTM model's strength also stems from the use of pre-trained GloVe domain-specific word embeddings, which effectively capture the semantics within the domain. Also, the attention to company names on the news headlines contributed to better performance due to the deeper context and relation between company names and news headlines being captured. This might account for the LSTM-based attention model's higher F1 score compared to the BERT-based model. On the other hand, the BERT-based model, pre-trained on more than 3 billion words from diverse domains, demonstrated better generalization ability, which results in higher accuracy compared to the LSTM-based model.

Proposed Multi-attention Network

Finally, we evaluated our proposed multi-attention network. We applied identical hyperparameter settings, consistent with those previously detailed for both the LSTM-based model with attention and the BERT-based model in our proposed multi-attention network. During the experiment, we varied the normalization weight, denoted as λ, on the validation set to determine the optimal configuration for achieving the highest accuracy.

Table 8 shows the accuracy of the validation set when different λ was applied. During the experiment, we assessed a range of λ from 1 to 0, with intervals of 0.1. This enabled us to identify the most appropriate weightage that yielded the highest accuracy. The most optimum λ was 0.2, as it performed with the highest accuracy, 0.4693.

Table 8
The accuracies of the validation set with different λ

λ	Validation Set Accuracy
1	0.4380
0.9	0.4468
0.8	0.4521
0.7	0.4593
0.6	0.4618
0.5	0.4630
0.4	0.4650
0.3	0.4673
0.2	**0.4693**
0.1	0.4622
0	0.4581

In the experiment, we analyzed that the normalization weight λ of the multi-attention network was 0.2 on the test set and obtained an accuracy of 0.5068, precision of 0.4373, recall of 0.4657, and F1 of 0.4510. The result showed a higher accuracy and F1 score compared to the LSTM-based model with attention and the BERT-based model. The study showed that the proposed multi-attention network was able to take advantage of both domain-general features of the BERT-based model and the domain-specific features of the LSTM-based model with attention to perform better.

Bias and Limitations

We identified several limitations in this study. The first was the use of end-of-day (EOD) stock price data to calculate stock trends, particularly when a news event occurs during stock market trading hours. Real-time stock price data would be more accurate in this context, as it would allow for the tracking of stock prices immediately before a news event is published. However, due to the unavailability of real-time Bursa Malaysia stock data, we used the opening price from the EOD stock data as a substitute.

The second limitation was the assumption that only company-specific news affects stock trends. This study exclusively examined the impact of company-related news on stock prices, though macroeconomic factors and government policy announcements can also significantly influence stock movements. Additionally, only one news event per instance was modeled to predict stock trends, primarily due to the limited availability of data. Analyzing the contribution of multiple news items on stock price movements would require a much larger dataset. However, the study's approach could easily be extended in future research to incorporate various types of news and the simultaneous influence of multiple news events when sufficient data is available.

The third limitation was the reliance on news headlines for modeling and prediction. It was assumed that the headline accurately captures the key points and essence of the news article. While this assumption held in most cases, there were instances where the headline did not fully convey the content of the article.

Lastly, the study focused solely on news headlines to predict stock trends. As discussed earlier, stock trends could also be influenced by stock prices, technical indicators, chart patterns, and other latent variables. While the experimental results are promising, they also suggest that there is significant room for improvement in stock price trend prediction by incorporating a wider range of data.

CONCLUSION

In this study, we collected a news headline dataset and the stock price trend for Bursa Malaysia. We proposed an approach to automatically annotate the stock price trend based on the publication time of news and show that news can make changes in the stock price of the target companies. Furthermore, the multi-attention network models domain-specific and domain-general features from news headlines and stock price trends of the targeted companies. The experiment showed that the proposed model performed the highest accuracy at 0.5068 and F1 score at 0.4510 compared to baseline deep learning approaches and conventional machine learning approaches. The performance of the LSTM WITH ATTENTION +BERT model was evaluated against several traditional methods. The LSTM WITH ATTENTION +BERT model achieved an accuracy of 50.68%, demonstrating a significant improvement over another model. Specifically, it outperformed the Decision Tree by 11.26%, the Naïve Bayes model by 20.1%, and the Support Vector Machine by 5.12%. Compared to the CNN, LSTM and BERT models, the proposed model is 4.27%, 2.91% and 1.64% higher, respectively, in terms of accuracy. These results highlight the advantages of integrating BERT's domain-general features with LSTM's sequential modeling capabilities.

Future studies could address some of the limitations identified in this research to further enhance stock trend prediction. First, researchers could explore news summarization techniques, which may provide a more comprehensive representation of a news article's content rather than relying solely on headlines. Summarizations could capture nuanced details that are often missed in short headlines, potentially leading to more accurate stock trend predictions. Second, it would be valuable to investigate the influence of other types of news, such as macroeconomic announcements, global political developments, and sector-specific events, on stock price movements. Understanding how these broader factors affect market trends could provide a more holistic approach to stock prediction. Third, future studies could focus on incorporating multiple news items within a specific timeframe to predict stock trends rather than analyzing isolated news events. This would better reflect real-world scenarios where multiple news stories can impact investor sentiment and market behavior simultaneously. Fourth, improving stock trend prediction models could involve integrating additional data such as real-time stock prices, technical indicators, financial reports, and market sentiment from social media or other sources. A more multi-faceted approach would allow models to capture a broader range of factors that influence stock movements, potentially improving predictive accuracy.

ACKNOWLEDGEMENT

This research is supported by the Fundamental Research Grant Scheme (FRGS/1/2021/ICT02/USM/02/4) of the Ministry of Higher Education, Malaysia.

REFERENCES

Agarwal, H., Jariwala, G., & Shah, A. (2020). Analysis and prediction of stock market trends using deep learning. In *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)* (pp. 521-531). Springer. https://doi.org/10.1007/978-981-15-3369-3_39

Akita, R., Toshihara, A., Matsubara, T., & Ushara, K. (2016). Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)* (pp. 1-6). IEEE Publishing. https://doi.org/ 10.1109/ICIS.2016.7550882

Azhikodan, A. R., Jadhav, M., & Bhat, A. (2019). Stock trading bot using deep reinforcement learning. In *Innovations in Computer Science and Engineering: Proceedings of the Fifth ICICSE 2017* (pp. 41-49). Springer. https://doi.org/10.1007/978-981-10-8201-6_5

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. arXiv Preprint. https://doi.org/10.48550/arXiv.2010.11929

Hu, Z., Hu, W., Bian, J., Liu, X., & Liu, T. Y. (2018). Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the eleventh ACM International Conference on Web Search and Data Mining* (pp. 261-269). ACM Publishing. https://doi.org/10.1145/3159652.3159690

Jariwala, G., Agarwal, H., & Jadhav, V. (2020). Sentimental analysis of news headlines for the stock market. In *2020 IEEE International Conference for Innovation in Technology (INOCON)* (pp. 1-5). IEEE Publishing. https://doi.org/10.1109/INOCON50539.2020.9298333

Jurafsky, D., & Martin, J. H. (2022). *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed.). ResearchGate.

Kabbani, T., & Duman, E. (2022). Deep reinforcement learning approach for trading automation in the stock market. *IEEE Access, 10*, 93564-93574. https://doi.org/ 10.1109/ACCESS.2022.3203697

Karpagavalli, S., & Chandra, E. (2016). A review of automatic speech recognition architecture and approaches. *International Journal of Signal Processing, Image Processing and Pattern Recognition, 9*(4), 393–404. https://doi.org/10.14257/ijsip.2016.9.4.34

Kaur, S. (2017). *Algorithmic Trading using Sentiment Analysis and Reinforcement Learning*. Stanford University. https://api.semanticscholar.org/CorpusID:28323724

Liu, Y., & Carrio, A. (2018). Stock price movement from financial news with deep learning and knowledge graph embedding. In *Knowledge Management and Acquisition for Intelligent Systems: 15th Pacific Rim Knowledge Acquisition Workshop, PKAW 2018* (pp. 102-113). Springer International Publishing. https://doi.org/10.1007/978-3-319-97289-3_8

Nguyen, T., & Yoon, S. (2019). A novel approach to short-term stock price movement prediction using transfer learning. *Applied Sciences, 9*(22), Article 4745. https://doi.org/ 10.3390/app9224745

Prachyachuwong, K., & Vateekul, P. (2021). Stock trend prediction using deep learning approach on technical indicator and industrial specific information. *Information, 12*(6), Article 250. https://doi.org/10.3390/info12060250

Santur, Y. (2022). Candlestick chart based trading system using ensemble learning for financial assets. *Sigma Journal of Engineering and Natural Sciences, 40*(2), 370-379. https://doi.org/10.14744/sigma.2022.00039

Souma, W., Vodenska, I., & Aoyama, H. (2019). Enhanced news sentiment analysis using deep learning methods. *Journal of Computational Social Science, 2*(1), 33-46. https://doi.org/ 10.1007/s42001-019-00035-x

Tipirisetty, A. (2018). *Stock Price Prediction using Deep Learning*. [Master Project]. San Jose State University, USA. https://doi.org/10.31979/etd.bzmm-36m7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (pp. 6000–6010). NeurIPS Proceedings. https://dl.acm.org/doi/10.5555/3295222.3295349