Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ30008

A Lightweight Approach to News-Driven Stock Movement Prediction with Limited Data

Liqin Fei

School of Future Technology, South China University of Technology, Guangzhou, China
202364870631@mail.scut.edu.cn

Abstract. Stock movement prediction is an important topic in the financial domain. However, it remains challenging due to high data noise and the limited availability of high-quality samples. To address this issue, this paper proposes a lightweight framework for news-driven stock movement prediction. In this framework, features are extracted from news texts using the Term Frequency–Inverse Document Frequency (TF-IDF) method. Dimension reduction is applied via Truncated Singular Value Decomposition (SVD) to compress the feature space. Finally, a shallow Multilayer Perceptron (MLP) is used to perform classification. Experimental results indicate that, under limited-data conditions, the combination of minimal word granularity, lower SVD dimensions, and a shallower network structure achieves the best performance and stability. Increased feature complexity incurs higher computational costs and may lead to performance degradation. Further comparative analysis shows that the proposed framework outperforms the traditional Random Forest baseline. This result validates the practicality of a lightweight deep learning approach. This study provides a reproducible and robust baseline method for news-driven stock price prediction in limited-data scenarios. Moreover, it offers a foundation for future research integrating multiple texts, multimodal data, and richer feature representations.

Keywords: Stock Movement Prediction, News Text Analysis, Term Frequency–Inverse Document Frequency, Truncated Singular Value Decomposition, Shallow Multilayer Perceptron

1. Introduction

Stock movement prediction has long drawn attention in financial research. Its influence extends from individual investment choices to the stability of the entire market. Early study including Random Walk Hypothesis (RWH) suggested that short-term stock price movements are inherently unpredictable [1]. Gradually, advances in text feature extraction made such prediction possible. Emotional signals in news and media texts provide predictive signals for stock price movements [2]. Together, these developments have laid a foundation for text-based stock price prediction.

Subsequently, with the rise of deep learning, researchers have increasingly explored more advanced semantic modeling methods for news analysis. BERT introduced and popularized the pre-training–fine-tuning paradigm based on bidirectional Transformers [3]. This paradigm provided a methodological foundation for learning general-purpose semantic representations. Its financial-

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ30008

domain extension, FinBERT, performs domain-adaptive pre-training on large-scale financial corpora [4]. It demonstrates strong performance in analytical tasks and has become a widely recognized baseline in financial text analysis. Building on these developments, later studies have utilized representations derived from pre-trained models for stock price prediction tasks, integrating them into downstream predictive networks and achieving improved performance [5]. PEN combines news texts with historical stock price features and further enhances prediction performance through multimodal information fusion [6]. CausalStock, on the other hand, introduces an end-to-end causal discovery framework that uncovers causal relationships between news and the movements of multiple stocks, thereby improving the interpretability of prediction [7].

However, financial news data contain substantial noise [7]. High-quality large-scale datasets remain limited. Investigating stock price prediction tasks on limited-data sets becomes particularly important. Under limited-sample conditions, higher model complexity increases the risk of overfitting [8]. This leads to degraded generalization performance. Therefore, in data-constrained scenarios, lightweight models tend to be more robust than deep models.

Based on this consideration, this paper proposes a lightweight prediction method that combines TF-IDF feature extraction, SVD-based dimensionality reduction, and a shallow MLP. This study systematically compares the effects of different text features, dimensionality reduction levels, and MLP network depths on prediction performance. Traditional machine learning methods are further contrasted with the proposed approach. Through these experiments, the effectiveness of news-based stock movement prediction are validated on a limited-size labeled dataset.

2. Method

2.1. Datasets and preprocessing

This study employs three publicly available datasets obtained from Kaggle, covering trading days from August 8, 2008, to July 1, 2016. The Combined_News_DJIA dataset contains 25 daily news headlines along with the corresponding DJIA movement labels (0 for decrease and 1 for increase). The RedditNews dataset starts slightly earlier (June 8, 2008), while only the overlapping period with Combined_News_DJIA is utilized in this study. It also provides 25 daily news headlines collected from the r/worldnews subreddit. The upload_DJIA_table dataset includes daily DJIA market information, such as opening price, closing price, highest and lowest prices, trading volume, and adjusted closing price.

In the data preprocessing stage, the news headlines were first subjected to standard text cleaning, including the removal of punctuation marks and the conversion of uppercase letters to lowercase. Subsequently, all news headlines from Combined_News_DJIA and RedditNews were merged, resulting in 50 headlines per trading day. The market data from upload_DJIA_table were then used to calculate the daily return, and the corresponding labels were generated accordingly. The data were divided chronologically into training and testing sets. In the training set, trading days with minor price changes were excluded to reduce noise, whereas no such filtering was applied to the testing set. This approach is motivated by related studies, which indicate that trading days with small price fluctuations often provide limited predictive signals and may lead to decreased classification performance [5]. After the aforementioned processing, the number of news headline samples per trading day remains limited, consistent with the small-data setting emphasized in this study.

DOI: 10.54254/2755-2721/2026.TJ30008

2.2. Feature extraction

To convert news headlines into numerical features suitable for model input, TF-IDF was first applied for vectorization. TF-IDF measures the importance of terms while reducing the weight of common words with limited predictive value [9]. Five n-gram configurations were employed in feature construction: 1-gram, 1–2-gram, 2-gram, 2–3-gram, and 3-gram. These configurations were used to compare the effects of different token granularities on model prediction performance.

The statistics indicate that, for certain n-gram configurations, the TF-IDF feature dimension is relatively high. Considering the limited sample size, a high-dimensional feature space tends to produce a sparse feature matrix, which increases computational cost and may adversely affect model training. To address this issue, Truncated SVD was applied to the feature matrix, retaining only the top k singular values and their corresponding feature vectors [10]. The value of k was set between 16 and 1024 to evaluate how the degree of dimension reduction affects prediction performance.

2.3. Model architecture

To address the stock price prediction task under small-sample conditions, this study designed a shallow MLP-based classification model [11]. Compared with sequence models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers, the MLP is better suited for processing fixed-dimensional feature vectors extracted by TF-IDF and SVD. Its smaller parameter count is also better suited for training on limited datasets. To investigate the effect of hidden layer depth on model performance, this paper constructed shallow MLPs with 3 to 9 layers. Layer Normalization and Dropout were applied to each hidden layer to stabilize training and improve generalization [12, 13]. Finally, a binary classification output layer was employed to predict stock price movements.

2.4. Model training and evaluation

The experiments were conducted on a GPU-equipped environment with an NVIDIA GeForce RTX 4060. The model takes feature vectors extracted by TF-IDF and reduced via Truncated SVD as input, and outputs binary labels indicating stock price movements. Training was performed using the cross-entropy loss function and the AdamW optimizer, with early stopping applied to enhance stability and prevent overfitting [14, 15].

To ensure reproducibility and assess model stability, each combination of n-gram settings, SVD dimension, and MLP depth was trained under the same five fixed seeds. In the testing phase, softmax was applied to the model outputs, followed by a threshold of 0.55 to convert probabilities into class labels. The model performance was evaluated using Accuracy and macro-F1 metrics, and the results were reported as the mean ± standard deviation.

3. Experimental results

In this study, the stock price prediction task was explored as a three-class classification problem, which led to a substantial decrease in accuracy. This result aligns with previous findings: on the E1 test set, the binary classification task achieved an Acc@1% of 74.5%, whereas the three-class task reached only 61.0% [5]. This indicates that adding a “neutral” category increases the difficulty of prediction. Therefore, the study ultimately adopts a binary classification setup.

Table 1 presents the optimal SVD dimensions, MLP layers, and corresponding performance under five n-gram configurations. The best performance occurs with 1-gram features (Accuracy

DOI: 10.54254/2755-2721/2026.TJ30008

0.833 ± 0.008, F1 0.831 ± 0.008). In this case, the SVD dimension is 32 and the network consists of three layers. As the n-gram range increases, the optimal configurations change in a consistent way. The SVD dimension and the number of MLP layers needed for peak performance gradually increase. The overall accuracy and F1 score decline slightly. These results suggest that, with limited data, increasing feature complexity requires a larger model and more computational resources. It also slightly reduces performance, leading to lower overall results.

Table 1. Optimal SVD dimensions, MLP layers, and performance under five N-gram configurations

n-gram	SVD dimension	Optimal Number of MLP Layers	Accuracy ± std	F1 ± std
1,1	32	3	0.833 ± 0.008	0.831 ± 0.008
1,2	256	5	0.820 ± 0.002	0.812 ± 0.002
2,2	512	6	0.815 ± 0.019	0.810 ± 0.020
3,3	512	6	0.815 ± 0.007	0.809 ± 0.004
2,3	512	6	0.810 ± 0.014	0.802 ± 0.014

Table 2. Performance of different MLP layers under the optimal feature combination (1-gram + SVD 32)

MLP Layer Count	Accuracy ± std	F1 ± std
3	0.833 ± 0.008	0.831 ± 0.008
4	0.832 ± 0.010	0.831 ± 0.010
5	0.634 ± 0.172	0.536 ± 0.251
6	0.510 ± 0.013	0.337 ± 0.006
7	0.510 ± 0.013	0.337 ± 0.006
8	0.503 ± 0.016	0.335 ± 0.007
9	0.510 ± 0.013	0.337 ± 0.006

Table 2 illustrates the impact of different MLP layer counts on model performance under the optimal feature combination (1-gram + SVD 32). The results indicate that the three-layer network achieves the highest performance (Accuracy 0.833 ± 0.008, F1 0.831 ± 0.008), while the four-layer network yields comparable results. Performance drops significantly when the network depth increases to five layers (Accuracy 0.634 ± 0.172, F1 0.536 ± 0.251). For six layers or more, the model essentially loses predictive capability, with performance approaching random levels (Accuracy ≈ 0.51, F1 ≈ 0.33). These results indicate that, under limited data conditions, a moderate network depth is sufficient to achieve effective performance, whereas excessive depth makes it difficult for the model to converge properly, leading to a substantial decline in performance.

DOI: 10.54254/2755-2721/2026.TJ30008

Table 3. Performance of different SVD dimensions under the optimal n-gram (1,1) and MLP with three layers

SVD dimension	Accuracy ± std	F1 ± std
16	0.717 ± 0.015	0.717 ± 0.015
24	0.822 ± 0.007	0.822 ± 0.007
32	0.833 ± 0.008	0.831 ± 0.008
64	0.812 ± 0.013	0.809 ± 0.013
128	0.821 ± 0.004	0.818 ± 0.004
256	0.811 ± 0.009	0.809 ± 0.008
512	0.803 ± 0.022	0.798 ± 0.018
1024	0.804 ± 0.008	0.802 ± 0.008

Table 3 shows the effect of different SVD dimensions on model performance under the best configuration of n-gram (1,1) features and MLP with three layers. The results indicate that excessively low dimensions (16) lead to a significant drop in performance due to insufficient information. Moderate dimensions (24, 32) achieve almost best performance. As the dimensions increase further, performance remains at a relatively high level, but the downward trend is evident. This trend is consistent with the results in Table 1, indicating that under limited data conditions, moderate feature representations are more conducive to model learning. Excessively low or high feature complexity can impair predictive performance.

In addition, to compare with traditional machine learning methods, this study also evaluated the Random Forest [16]. When the SVD dimension is relatively low (e.g., 32), the Random Forest achieves an accuracy of approximately 0.68. As the dimension increases, performance gradually improves, reaching around 0.82 at SVD = 128, and shows little change with further increases in dimension. In contrast, the MLP achieves optimal performance (0.832) at a low dimension (32), whereas the Random Forest requires higher dimensions to achieve stable performance. This indicates that a shallow MLP has advantages under lightweight feature conditions and is more suitable for small-sample scenarios.

4. Conclusion

For predicting news-driven stock movements under limited datasets, this paper presents a lightweight prediction framework encompassing text feature representation, dimensionality reduction, deep model training, and performance evaluation. The framework is built on TF-IDF features, combined with Truncated SVD for dimensionality reduction, and employs a shallow multilayer perceptron for classification; model performance and stability metrics are used to validate its effectiveness, resulting in a prediction method that is both parameter-efficient and structurally simple.

Experimental results indicate that under limited data conditions, a lightweight combination of the smallest word granularity, lower SVD dimensionality, and a shallow MLP demonstrates superior classification performance and experimental stability. Increasing feature complexity not only raises the demands on model architecture and computational resources, but also leads to a decline in performance. To further assess the effectiveness of the prediction framework under limited sample conditions, it was compared with a traditional random forest. The results show that the framework demonstrates superior overall performance.

DOI: 10.54254/2755-2721/2026.TJ30008

Despite demonstrating strong classification performance and stability under limited data conditions, the proposed lightweight framework still has certain limitations. First, this study only utilizes news headlines as input, lacking support from longer texts or multimodal information, which may limit the model's ability to capture richer market signals. In addition, the adopted feature representation (TF-IDF combined with Truncated SVD) and shallow network structure, while lightweight and efficient, may have limitations in handling more complex linguistic relationships or nonlinear market patterns. Future work may consider incorporating additional text sources (e.g., full news articles, financial reports, earnings call transcripts) and multimodal data, while exploring deeper or semantically enhanced feature representations to further improve the model's predictive capability and generalizability.

References

[1] Fama, E. F. (1965) The Behavior of Stock-Market Prices. The Journal of Business, 38, 34-105.
[2] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of finance, 62(3), 1139-1168.
[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).
[4] Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv: 1908.10063.
[5] Chen, Q. (2021). Stock movement prediction with financial news using contextualized embedding from bert. arXiv preprint arXiv: 2107.08721.
[6] Li, S., Liao, W., Chen, Y., & Yan, R. (2023, June). PEN: prediction-explanation network to forecast stock price movement with better explainability. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 4, pp. 5187-5194).
[7] Li, S., Sun, Y., Lin, Y., Gao, X., Shang, S., & Yan, R. (2024). CausalStock: Deep end-to-end causal discovery for news-driven multi-stock movement prediction. Advances in Neural Information Processing Systems, 37, 47432-47454.
[8] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning.
[9] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information processing & management, 24(5), 513-523.
[10] Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM review, 53(2), 217-288.
[11] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.
[12] Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. arXiv preprint arXiv: 1607.06450.
[13] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.
[14] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv: 1711.05101.
[15] Prechelt, L. (2002). Early stopping-but when? In Neural Networks: Tricks of the trade (pp. 55-69). Berlin, Heidelberg: Springer Berlin Heidelberg.
[16] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.