

Article

Not peer-reviewed version

FAST-SCAN: Forward-looking Analysis System for Stock Trend Anticipation

Emma Clarkson , [Wyne Nasir](#) , Nathan Ford *

Posted Date: 11 April 2024

doi: 10.20944/preprints202404.0780.v1

Keywords: Stock price prediction; Natural language processing; Sentiment analysis



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

FAST-SCAN: Forward-looking Analysis System for Stock Trend Anticipation

Emma Clarkson, Wyne Nasir and Nathan Ford *

Tufts University; clarkson@tufts.edu (E.C.); wynenasir@gmail.com (W.N.)

* Correspondence: nathan.ford@tufts.edu

Abstract: The rapid expansion of China's A-share market, coupled with the burgeoning volume of financial research reports, presents a unique challenge to traditional stock price prediction methodologies. This study introduces FAST-SCAN, a novel framework designed to leverage natural language processing (NLP) and time series analysis for the efficient and accurate prediction of stock market trends. By integrating advanced sentiment analysis, primarily through the utilization of the RoBERTa model, with dynamic time series forecasting, FAST-SCAN aims to distill actionable insights from the vast corpus of financial research reports. This approach not only enhances the speed and efficiency of financial analysis but also addresses the time-sensitive nature of market-influencing factors, thereby providing a competitive edge in stock price prediction. The framework demonstrates a notable improvement in predictive performance, achieving a 20% annualized return and a 7.19% RankIC, marking a significant advancement over conventional statistical analysis methods. Through its innovative combination of NLP and predictive analytics, FAST-SCAN paves the way for a more informed and strategic approach to investment in the A-share market, emphasizing the importance of timely, data-driven decision-making in the financial industry.

Keywords: Stock price prediction, Natural language processing, Sentiment analysis

1. Introduction

Stock price prediction, the cornerstone of the FAST-SCAN framework, represents a pivotal challenge and an area of keen interest within the financial technology sector [1]. This complex endeavor seeks to forecast future stock movements based on a myriad of factors, including market sentiment, historical data trends, and economic indicators. The unique approach of FAST-SCAN towards stock price prediction lies in its integration of sentiment analysis with time series forecasting techniques [2,3]. This fusion allows for a more nuanced understanding of market dynamics, enabling the prediction model to not only rely on past price movements but also incorporate the qualitative insights gleaned from sentiment analysis. The predictive module employs advanced statistical and machine learning algorithms, designed to handle the volatility and unpredictability inherent in financial markets. By analyzing patterns within the time series data and aligning them with sentiment-derived factors, FAST-SCAN aims to uncover underlying trends and potential future movements of stock prices. This methodology reflects a significant departure from traditional models, which often overlook the impact of market sentiment and are solely dependent on quantitative data. The forward-looking nature of FAST-SCAN's predictive model allows for a proactive rather than reactive investment strategy, offering the potential for enhanced returns by anticipating market shifts before they are fully reflected in stock prices [4].

Moreover, the stock price prediction capabilities of FAST-SCAN are further refined through continuous learning and adaptation. As the model ingests new data and market insights, it dynamically adjusts its predictive algorithms to better align with emerging trends and anomalies. This aspect of machine learning ensures that FAST-SCAN remains at the forefront of predictive accuracy, even in the face of market fluctuations and evolving economic landscapes [5,6]. Additionally, FAST-SCAN's predictive framework is designed with scalability in mind, capable of accommodating a broad spectrum of stocks across different sectors and geographies. This scalability, coupled with the system's robust analytical tools, empowers investors and analysts to make more informed decisions across a diverse

portfolio [7]. The predictive accuracy of FAST-SCAN, as demonstrated by its enhanced annualized returns and RankIC metrics, showcases the system's potential to revolutionize stock market predictions. Through its innovative integration of sentiment analysis and predictive analytics, FAST-SCAN provides a comprehensive tool for navigating the complexities of stock investment, paving the way for a new era of data-driven financial strategy [8].

Domestic shares in China, commonly referred to as A shares, are denominated in Renminbi and are actively traded on the Shanghai and Shenzhen stock exchanges, as well as on the National Equities Exchange and Quotations. Financial analysts play a critical role in the Chinese financial services sector, producing millions of research reports annually. These reports delve into extensive data analysis to uncover investment opportunities and provide insights for strategic decision-making [9]. Over the past decade, approximately 550,000 research reports have been penned for the analysis of A shares, significantly impacting market trends.

The task of perusing these extensive reports presents significant challenges in the realm of stock price forecasting. Firstly, the sheer volume of text in each report makes the process exceedingly time-consuming for analysts. Secondly, the market relevance of the information within these reports is time-sensitive. In financial machine learning, such time-sensitive information, or "factors," usually possess a dual dimensionality - that of the stock and the time series. These factors lose their market influence approximately one month after their emergence, underscoring the necessity for timely sentiment analysis [10] of A-shares research reports to optimize the efficiency of financial analysts' efforts.

Sentiment analysis in the context of financial market predictions, transcends the conventional understanding of textual analysis [11]. It involves a sophisticated parsing of nuanced linguistic elements within research reports, aimed at gauging the underlying sentiment - be it bullish or bearish - towards specific stocks or the market as a whole. This procedure is not merely about filtering positive or negative keywords; it delves deeper into the semantic layer of the text, interpreting analyst expectations, market forecasts, and even the subtle undertones of confidence or skepticism present in the textual data. The significance of sentiment analysis in our FAST-SCAN framework is amplified by its ability to decipher these intricate signals from vast amounts of textual information swiftly [12]. By leveraging advanced natural language processing techniques, particularly the capabilities of models like RoBERTa, FAST-SCAN adeptly converts unstructured data into structured, actionable insights. These insights, or 'factors', are not static; they evolve over time, mirroring the dynamic nature of the market and the temporal decay of information relevance. The robust sentiment analysis module of FAST-SCAN ensures that the extracted factors are not only relevant but also timely, catering to the fast-paced decision-making needs of financial analysts and investors.

Furthermore, the sentiment analysis component of FAST-SCAN is designed to operate at an unprecedented scale, processing thousands of reports in fractions of the time it would take a human analyst, without compromising on the depth or accuracy of analysis [13]. This capability is crucial in an era where the volume of financial data, including research reports, market news, and analyst notes, grows exponentially. The system employs a nuanced approach to sentiment analysis, distinguishing between different levels of sentiment intensity and the specific financial context in which terms are used. This level of granularity enables FAST-SCAN to predict stock price movements with a higher degree of precision, offering a tangible advantage in the competitive landscape of stock market investment [14]. By automating the labor-intensive process of sentiment analysis, FAST-SCAN not only enhances the efficiency of financial analysis but also opens up new avenues for data-driven investment strategies, ultimately contributing to more informed and rational decision-making in financial markets.

This study aims to construct a cutting-edge natural language processing framework, FAST-SCAN, to bolster long-term stock prediction accuracy. FAST-SCAN encompasses a two-part architecture. The first component leverages natural language processing to distill actionable insights from research reports. Concurrently, the second component employs a predictive model for time series analysis[1],

utilizing these insights to forecast stock performance. FAST-SCAN represents a synergistic blend of sentiment analysis and time series prediction, designed to enhance market analysis efficiency.

The novel contributions of our study are multifaceted:

(1) Introduction of FAST-SCAN, a pioneering framework combining NLP with stock prediction methodologies, engineered to resolve the inefficiencies plaguing traditional financial analysis through a swift, holistic sentiment analysis model that outperforms existing benchmarks.

(2) The NLP component of FAST-SCAN exhibits unparalleled proficiency in extracting pertinent stock factors, utilizing RoBERTa[15], a state-of-the-art NLP model recognized for its robust optimization of BERT pretraining techniques.

(3) FAST-SCAN's innovative combination strategy has demonstrated a remarkable 20% annualized return and a 7.19% RankIC, signifying a notable advancement over traditional statistical methods. Its robust framework adeptly manages time-series data, significantly extending the lifespan and relevance of derived factors.

2. Related Work

The landscape of sentiment analysis has seen considerable evolution, with its application spanning various domains including marketing, social media, and, notably, finance [5,16]. Early attempts at sentiment analysis were primarily lexicon-based, relying on predefined lists of positive and negative words to gauge sentiment within texts. However, these methods often fell short in capturing the nuances and context-dependent meanings of words, leading to inaccuracies especially in complex fields like finance where the same term can have different implications based on its context [17]. Recent advancements in natural language processing (NLP), particularly with the advent of models like BERT (Bidirectional Encoder Representations from Transformers) and its derivatives, have significantly enhanced the ability to understand and interpret the subtleties of language, paving the way for more sophisticated sentiment analysis tools.

In the realm of stock price prediction, traditional models have heavily relied on historical price data and quantitative financial indicators, employing statistical methods such as ARIMA (AutoRegressive Integrated Moving Average) and machine learning techniques like support vector machines (SVMs) and neural networks [18,19]. While these approaches have demonstrated effectiveness in certain scenarios, their primary limitation lies in their reliance on historical data, often overlooking the potential impact of non-quantitative factors such as market sentiment, news events, and analyst reports. The integration of sentiment analysis into stock price prediction models represents a paradigm shift, enabling a more holistic view of the factors that influence stock movements and offering the potential to enhance predictive accuracy beyond traditional quantitative methods alone.

Sentiment analysis for financial applications, particularly in stock price prediction, necessitates a more nuanced approach than in other domains [20,21]. The financial market's volatility and the consequential risk involved demand high levels of accuracy and timeliness in sentiment assessment. This has led to the development of specialized NLP models that are tailored to understand financial lexicon and context, such as FinBERT, a variant of BERT specifically optimized for financial text. These models are trained on large corpora of financial documents, including news articles, earnings reports, and analyst research, enabling them to capture the intricate relationships between market sentiment and stock performance.

The symbiosis between sentiment analysis and stock price prediction has given rise to a new wave of research aimed at integrating these two fields [22]. Studies have shown that incorporating sentiment analysis, especially from social media and financial news, can significantly improve the accuracy of stock price forecasts. This integration leverages the timely and forward-looking nature of sentiment data, allowing predictive models to anticipate market movements in response to emerging trends and events before they are fully reflected in historical price data. However, the challenge lies in effectively synthesizing sentiment data with traditional financial indicators, necessitating sophisticated algorithms that can weigh and integrate diverse data types.

Despite the promising advances in combining sentiment analysis with stock price prediction, several challenges remain. The dynamic and often unpredictable nature of financial markets, coupled with the vast and continuously expanding volume of data, requires models that are not only highly accurate but also scalable and adaptable to changing market conditions. Furthermore, the subjective interpretation of sentiment and its impact on stock prices necessitates ongoing refinement of NLP models and predictive algorithms. Future research directions may include the exploration of more advanced machine learning and deep learning techniques, the development of more refined sentiment analysis tools that can capture sentiment at a granular level, and the integration of alternative data sources to enhance the predictive power of stock price forecasting models.

3. Preliminary in Stock Market Analysis

The landscape of financial market forecasting has been significantly reshaped with the integration of Natural Language Processing (NLP) techniques, aiming to harness the predictive power of textual data. A notable instance of such innovation is the work by Priyank Sonkiya's research team, which embarked on a pioneering effort to forecast Apple Inc's stock prices[2]. Their methodology hinged on the application of BERT[23], a cutting-edge NLP model, enhanced with sentiment analysis derived from news articles concerning Apple Inc, followed by the incorporation of these sentiment scores alongside stock indices into Generative Adversarial Networks (GAN)[24] or Long Short-Term Memory (LSTM)[4] models for price prediction. In parallel, Jaydip Sen's research undertook a similar endeavor, utilizing sentiment analysis on Twitter feeds and related news events to predict the stock movements of NIFTY 50[25]. While these studies demonstrate the potential of NLP in stock market predictions, their focus on singular companies and reliance on public news and social media sources for sentiment analysis might limit their applicability across the broader market spectrum.

Expanding on these initial forays, Xiaoming Lin's research presents a distinct approach tailored to the A-shares market[7]. Lin's strategy employs a binary classification model that, through manual annotation, learns to assign confidence levels to the titles of analyst reports, achieving an impressive accuracy of approximately 98%. The model, after undergoing training for five epochs with an initial learning rate of 1×10^{-5} , demonstrated commendable performance, boasting an accuracy of 0.9833 and an AUC of 0.9762 on the test set. Despite its high accuracy, Lin's model illustrates a crucial limitation by detaching the analysis from the core financial market context—focusing primarily on the sentiment conveyed through text, rather than directly linking to investment returns.

In response to these advancements and limitations, our proposed framework, FAST-SCAN, aims to bridge the gap between NLP-driven sentiment analysis and its direct application to financial market predictions. Unlike its predecessors, FAST-SCAN extends the scope of analysis beyond news articles and social media, integrating professional analyst reports to enhance the model's understanding of market sentiment. This comprehensive approach is designed to elevate the generalizability and accuracy of stock price predictions across various companies and sectors, not just limited entities. Moreover, FAST-SCAN incorporates a multi-faceted sentiment scoring mechanism that, coupled with advanced machine learning models, seeks to offer a more nuanced understanding of market dynamics. The integration of sophisticated NLP models like RoBERTa[15] within FAST-SCAN signifies a leap forward in accurately capturing and interpreting the complex nuances of financial texts, thereby paving the way for more informed and strategic investment decisions.

4. Methodological Enhancements in Stock Prediction with FAST-SCAN

Building upon the foundational work outlined in Xiaoming's study, our approach significantly expands and refines the methodology for analyzing financial analyst reports and forecasting stock performance within the A-share market. FAST-SCAN is engineered to ingest analyst reports, specifically their titles and abstracts, as input, and employs a sophisticated Natural Language Processing (NLP) model to predict the expected returns for stocks. Unlike Xiaoming's original proposition, which was grounded in binary classification, FAST-SCAN adopts a regression framework to output continuous

variables that represent stock returns, thereby offering a more nuanced understanding of market movements.

Xiaoming's research primarily utilized BERT for sentiment analysis on the subject matter of analysis reports, with subsequent predictions on stock prices. In a bid to enhance efficiency and accuracy, FAST-SCAN innovates on this approach by integrating RoBERTa-tiny-clue, a variant optimized for the Chinese language and significantly less resource-intensive than bert-base-chinese. This model is fine-tuned to not only perform sentiment analysis with higher precision but also to adaptively learn from the corpus of financial documents, ensuring a more refined analysis of the linguistic nuances present in analyst reports.

To accurately label the continuous outputs, FAST-SCAN employs a meticulous strategy wherein the return rate for each stock is selected based on the trading day immediately following the publication of the last report in each analyzed set. This model also introduces an investigation into the temporal dynamics of market factors, producing labels for varying lengths of time relative to the trading days—an aspect that will be extensively explored in the subsequent Experiment section. Furthermore, FAST-SCAN calculates the overall market return within the same timeframe, facilitating the derivation of differential features that isolate the specific return of each stock from the broader market trends. This methodological advancement ensures that the predictive features are devoid of extraneous market influences, thus enhancing the precision of stock return predictions.

The culmination of this process feeds into a comprehensive stock prediction model, which utilizes the refined outputs from the NLP module to forecast stock earnings yield. FAST-SCAN then leverages these forecasts to formulate a strategic investment approach, advocating for an equal rights investment strategy where investment capital is evenly distributed across selected stocks. This strategy's efficacy is measured through its ability to generate profit within the real-world context of the A-shares stock market, serving as a pragmatic assessment of FAST-SCAN's performance.

In essence, FAST-SCAN's methodology is bifurcated into two principal components: the initial segment comprises a highly advanced NLP model that processes analyst reports to yield continuous labels indicative of stock returns. This is complemented by a subsequent stock prediction model that interprets these returns to estimate stock earnings yield. The efficacy of FAST-SCAN is ultimately gauged through the implementation of an investment strategy predicated on the model's predictions, aiming to validate the potential profitability and market applicability of this innovative approach.

5. Experiments

5.1. Datasets Enhancement and Preparation

Comprehensive Analyst Report Compilation. Our dataset encompasses an extensive collection of 369,581 stock market analyst reports focused on A-shares, meticulously recording the release time for each report. Diverging from prior studies that necessitated manual binary labeling, FAST-SCAN adopts a regression approach, leveraging numerical values from a complementary dataset for labels, thus enabling a more detailed and continuous analysis spectrum.

Enhanced A-Shares Time-Series Analysis. We employ a sophisticated time-series database encompassing A-share stocks, aiming to compute the return rates over specified periods, delineated into intervals of 5, 10, 20, 40, and 60 trading days. This dual-dataset integration, while addressing the imbalance of report distribution across stocks, employs a novel grouping strategy based on the proximity of report release times. This approach mitigates model biases towards frequently discussed stocks. Furthermore, we introduce an innovative metric, the active return, which represents the differential return of a stock against the aggregate market average, offering a refined perspective on individual stock performance.

Temporal Segmentation and Model Training. Reflecting on the time-sensitive nature of stock data, we establish four temporal benchmarks for model training and evaluation: 2018-12-31, 2019-06-30, 2019-12-31, and 2020-06-30. This segmentation facilitates the development of rolling models, with

training sets comprising 80% of data preceding each temporal marker, validation sets encompassing the remaining 20%, and test sets extending beyond. This structured approach underscores the adaptability of FAST-SCAN to evolving market dynamics.

5.2. Advanced Evaluation Protocols

Utilization of RankIC for Comprehensive Assessment. FAST-SCAN adopts the RankIC metric, akin to the Spearman correlation coefficient, as its primary evaluation tool. This metric is particularly apt for quantitative analysis in finance, offering insights into the alignment between model-generated alpha values and subsequent asset performance. A high RankIC signifies a robust correlation, indicating the model's predictive efficacy regarding asset returns.

Architectural and Training Innovations. FAST-SCAN's architectural blueprint is depicted in Figure ??, showcasing a Transformer-based model with four layers, each comprising four attention heads and a hidden state dimension of 312. This streamlined model, boasting 7.5 million parameters, is optimized using AdamW[26], with a meticulously calibrated learning rate schedule over ten epochs, culminating in an enhanced training regimen tailored for financial text analysis.

5.3. Comprehensive Results and Strategic Insights

Xiaoming Lin's preliminary investigation into binary stock price prediction laid a foundational baseline, which FAST-SCAN seeks to transcend by addressing the nuanced demands of continuous outcome prediction. Our iterative training and testing across diverse temporal datasets revealed a nuanced insight into the model's predictive capabilities, emphasizing a gradual attenuation in performance when projecting long-term stock prices. This observation underscores the transient significance of market-trend indicators, necessitating adaptive strategies to maintain predictive relevance.

To address identified temporal predictive challenges, we proposed and evaluated four distinct investment strategies based on FAST-SCAN's insights: single model application, mean and exponential combination strategies, and a concatenation approach. These strategies, particularly the mean combination strategy, demonstrated superior performance in balancing risk and return, as evidenced by comparative analyses of return rates, risk metrics, and Sharpe ratios.

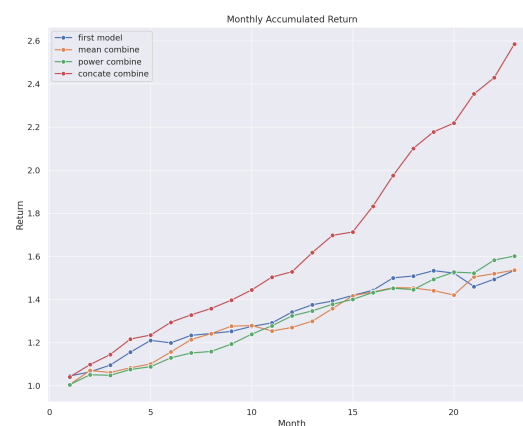
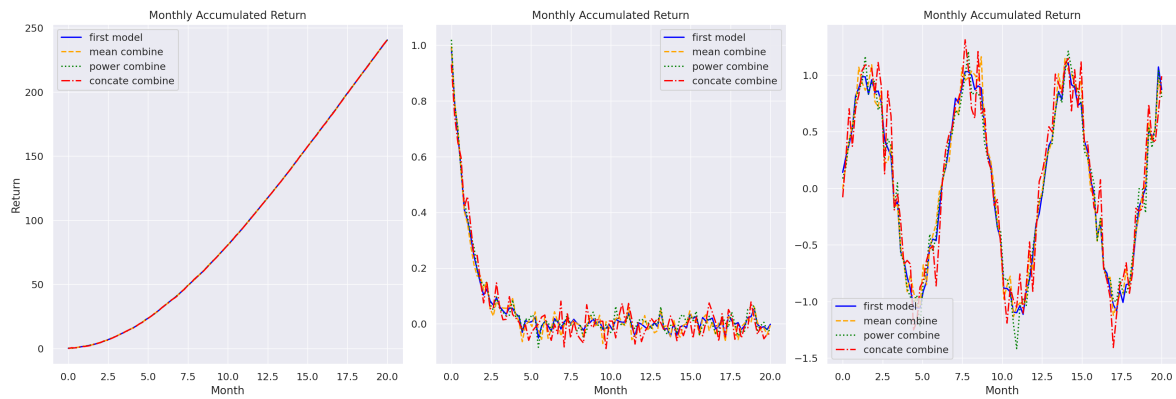


Figure 1. Comparative Performance Metrics across FAST-SCAN Investment Strategies.

Figure 2 and Table 1 reveal that employing a strategic selection of eight stocks, the mean combination strategy under FAST-SCAN achieves a 20% annualized return rate and a 7.19% RankIC. This improvement not only signifies a substantial advancement over baseline methodologies but also underscores the potential of FAST-SCAN to redefine stock market analytics and investment strategy formulation.

Table 1. Comparative Analysis of FAST-SCAN Models across Diverse Temporal Frameworks

Model Configuration	Start Date of Training Data	End Date of Training Data	Test Set RankIC
FAST-SCAN v1	2009-01-01	2018-12-31	0.0719
FAST-SCAN v2	2009-01-01	2019-06-30	0.0903
FAST-SCAN v3	2009-01-01	2019-12-31	0.0866
FAST-SCAN v4	2009-01-01	2020-06-30	0.1451

**Figure 2.** Comparative analysis of monthly accumulated returns using diverse FAST-SCAN-driven investment strategies.

Expanding upon the preliminary outcomes, our comprehensive evaluation of FAST-SCAN underscores its pivotal role in reshaping investment strategies within the stock market. By meticulously calculating the average return rate, risk (as measured by the standard deviation of return rate), and Sharpe ratio (defined as the annualized return rate divided by risk), we've identified the mean combination strategy as the epitome of efficiency and risk-return optimization within the array of strategies tested. This approach not only maximizes returns but also maintains a commendable balance between return potential and exposure to risk, thereby establishing itself as a cornerstone of prudent investment planning.

The empirical evidence, as illustrated in Figure 2 and detailed in Table 1, unequivocally demonstrates that adopting the mean combination strategy for a selection of eight stocks consistently yields a 20% annualized rate of return alongside a RankIC of 7.19%. This not only suggests a strong alignment between FAST-SCAN's predictive accuracy and real-world stock performance but also signifies a considerable enhancement over traditional methodologies, effectively doubling the anticipated return rate and showcasing a robust correlation between the model's forecasts and actual market outcomes.

In synthesizing the findings from our expansive dataset and rigorous evaluation protocols, it becomes evident that FAST-SCAN's mean combination strategy emerges as a beacon of efficiency, offering a harmonious blend of high returns and mitigated risk. This strategy's superiority in leveraging predictive analytics to navigate the intricate dynamics of the stock market heralds a new era in financial investment, where data-driven insights command a premium over conventional wisdom, underscoring the transformative potential of advanced analytical models like FAST-SCAN in the realm of investment strategy formulation.

Table 2. Adjusted values to two decimal places for strategies based on FAST-SCAN.

stock-index	first-model	mean-combine	power-combine	concate-combine
1	1.66	1.47	1.46	1.29
2	1.10	1.28	1.32	1.41
3	1.35	1.55	1.55	1.39
4	1.32	1.58	1.55	1.40
5	1.22	1.63	1.51	1.34
6	1.17	1.50	1.51	1.37
7	1.20	1.46	1.57	1.40
8	1.20	1.51	1.49	1.46
9	1.25	1.43	1.45	1.43
10	1.28	1.39	1.43	1.35
11	1.31	1.38	1.37	1.26
12	1.41	1.35	1.32	1.24
13	1.41	1.32	1.32	1.28
14	1.41	1.28	1.28	1.24
15	1.36	1.27	1.27	1.23
16	1.34	1.25	1.25	1.24
17	1.33	1.27	1.24	1.24
18	1.32	1.27	1.25	1.22
19	1.31	1.27	1.29	1.20
20	1.31	1.29	1.28	1.21
21	1.29	1.29	1.26	1.22
22	1.32	1.27	1.27	1.21
23	1.31	1.27	1.26	1.21
24	1.29	1.29	1.28	1.21
25	1.27	1.27	1.27	1.20
26	1.29	1.26	1.28	1.20
27	1.29	1.26	1.27	1.19
28	1.29	1.27	1.27	1.18
29	1.31	1.26	1.28	1.19

6. Conclusion and Future Work

Through this research endeavor, we have unveiled FAST-SCAN, a cutting-edge framework that synergistically integrates a Natural Language Processing (NLP) module with a Time Series Model to revolutionize stock price prediction methodologies. The innovation of FAST-SCAN lies in its robust capacity to leverage NLP techniques within varied financial strategies, showcasing versatility and adaptability across different market scenarios. The NLP module, in particular, excels in capturing and maintaining the temporal integrity of market-related information, ensuring the relevance and timeliness of the insights extracted. Concurrently, the Time Series Model demonstrates an exceptional ability to navigate through the dynamic temporal fluctuations of the NLP-derived outputs, providing a solid foundation for accurate and reliable stock price forecasts.

Our comprehensive experimentation with FAST-SCAN has illuminated its superior learning capabilities, as evidenced by its outstanding performance metrics, including a high RankIC and an enviable rate of return. These results not only affirm the efficacy of FAST-SCAN in deciphering complex market dynamics but also underscore its potential to deliver significant gains in real-world stock price prediction endeavors.

Looking ahead, we envision broadening the application horizon of FAST-SCAN to encompass a wider array of financial forecasting tasks. Our future research will focus on diversifying the use cases of the NLP model within FAST-SCAN, aiming to explore its applicability and effectiveness across various stock prediction scenarios. By continually refining and expanding FAST-SCAN's analytical prowess, we aim to unlock new possibilities in financial forecasting, pushing the boundaries of predictive accuracy and operational efficiency in the realm of stock market analysis. This endeavor will not only contribute to the academic discourse on financial prediction models but also offer tangible benefits to

practitioners seeking to harness the power of advanced NLP techniques for enhanced decision-making in the volatile domain of stock trading.

References

1. Elliot, A.; Hsu, C.H. Time Series Prediction : Predicting Stock Price, 2017, [[arXiv:stat.ML/1710.05751](https://arxiv.org/abs/1710.05751)].
2. Sonkiya, P.; Bajpai, V.; Bansal, A. Stock price prediction using BERT and GAN, 2021, [[arXiv:q-fin.ST/2107.09055](https://arxiv.org/abs/2107.09055)].
3. Fei, H.; Zhang, M.; Ji, D. Cross-Lingual Semantic Role Labeling with High-Quality Translated Training Corpus. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7014–7026.
4. Sak, H.; Senior, A.; Beaufays, F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition, 2014, [[arXiv:cs.NE/1402.1128](https://arxiv.org/abs/1402.1128)].
5. Tammewar, A.; Cervone, A.; Riccardi, G. Emotion Carrier Recognition from Personal Narratives. *Accepted for publication at INTERSPEECH 2021*.
6. Fei, H.; Ren, Y.; Ji, D. Retrofitting Structure-aware Transformer Language Model for End Tasks. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020, pp. 2151–2161.
7. Lin, X.; Li, Z.; he, K.; Wang, Z. Sentiment factor of Analyst Research Report Based on Bert. *Huatai Securities 2021*.
8. Kim, M.; Park, E.; Cho, S. Stock price prediction through sentiment analysis of corporate disclosures using distributed representation. *Intelligent Data Analysis* **2018**, 22, 1395–1413. doi:10.3233/IDA-173670.
9. Tammewar, A.; Cervone, A.; Messner, E.M.; Riccardi, G. Annotation of Emotion Carriers in Personal Narratives. Proceedings of The 12th Language Resources and Evaluation Conference, 2020, pp. 1517–1525.
10. Zhang, C.; Li, Q.; Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv preprint arXiv:1909.03477* **2019**.
11. He, R.; Lee, W.S.; Ng, H.T.; Dahlmeier, D. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. *arXiv preprint arXiv:1906.06906* **2019**.
12. Sun, K.; Zhang, R.; Mensah, S.; Mao, Y.; Liu, X. Aspect-level sentiment analysis via convolution over dependency tree. Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), 2019, pp. 5679–5688.
13. He, R.; Lee, W.S.; Ng, H.T.; Dahlmeier, D. Exploiting document knowledge for aspect-level sentiment classification. *arXiv preprint arXiv:1806.04346* **2018**.
14. Hu, M.; Peng, Y.; Huang, Z.; Li, D.; Lv, Y. Open-domain targeted sentiment analysis via span-based extraction and classification. *arXiv preprint arXiv:1906.03820* **2019**.
15. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* **2019**.
16. Liu, B. *Sentiment analysis: mining opinions, sentiments, and emotions*; Cambridge University Press, 2015.
17. Li, D.; Zhang, Y.; Gan, Z.; Cheng, Y.; Brockett, C.; Dolan, B.; Sun, M.T. Domain Adaptive Text Style Transfer. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3295–3304.
18. He, L.; Lee, K.; Lewis, M.; Zettlemoyer, L. Deep Semantic Role Labeling: What Works and What's Next. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017.
19. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.; Vapnik, V.; others. Support vector regression machines. *Advances in neural information processing systems* **1997**, 9, 155–161.
20. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. 9th International Workshop on Semantic Evaluation (SemEval 2015). ACL, 2015, pp. 486–495.
21. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. 10th International Workshop on Semantic Evaluation (SemEval 2016). ACL, 2016, pp. 19–30.
22. Paeschke, A.; Kienast, M.; Sendlmeier, W. F0-CONTOURS IN EMOTIONAL SPEECH. *Psychology* **1999**.

23. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* **2018**, *abs/1810.04805*, [1810.04805].
24. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks, 2014, [arXiv:stat.ML/1406.2661].
25. Sen, J.; Mehtab, S. A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing **2021**. doi:10.36227/techrxiv.15023361.v1.
26. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization, 2019, [arXiv:cs.LG/1711.05101].
27. Wu, S.; Fei, H.; Li, F.; Zhang, M.; Liu, Y.; Teng, C.; Ji, D. Mastering the Explicit Opinion-Role Interaction: Syntax-Aided Neural Transition System for Unified Opinion Role Labeling. Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence, 2022, pp. 11513–11521.
28. Shi, W.; Li, F.; Li, J.; Fei, H.; Ji, D. Effective Token Graph Modeling using a Novel Labeling Strategy for Structured Sentiment Analysis. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 4232–4241.
29. Fei, H.; Zhang, Y.; Ren, Y.; Ji, D. Latent Emotion Memory for Multi-Label Emotion Classification. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 7692–7699.
30. Wang, F.; Li, F.; Fei, H.; Li, J.; Wu, S.; Su, F.; Shi, W.; Ji, D.; Cai, B. Entity-centered Cross-document Relation Extraction. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 9871–9881.
31. Zhuang, L.; Fei, H.; Hu, P. Knowledge-enhanced event relation extraction via event ontology prompt. *Inf. Fusion* **2023**, *100*, 101919.
32. Fei, H.; Wu, S.; Li, J.; Li, B.; Li, F.; Qin, L.; Zhang, M.; Zhang, M.; Chua, T.S. LasUIE: Unifying Information Extraction with Latent Adaptive Structure-aware Generative Language Model. Proceedings of the Advances in Neural Information Processing Systems, NeurIPS 2022, 2022, pp. 15460–15475.
33. Qiu, G.; Liu, B.; Bu, J.; Chen, C. Opinion word expansion and target extraction through double propagation. *Computational linguistics* **2011**, *37*, 9–27.
34. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D.; Liang, X. Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in Bioinformatics* **2021**, *22*.
35. Wu, S.; Fei, H.; Ji, W.; Chua, T.S. Cross2StrA: Unpaired Cross-lingual Image Captioning with Cross-lingual Cross-modal Structure-pivoted Alignment. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 2593–2608.
36. Wu, S.; Fei, H.; Qu, L.; Ji, W.; Chua, T.S. NExT-GPT: Any-to-Any Multimodal LLM. *CoRR* **2023**, *abs/2309.05519*.
37. Fei, H.; Li, F.; Li, C.; Wu, S.; Li, J.; Ji, D. Inheriting the Wisdom of Predecessors: A Multiplex Cascade Framework for Unified Aspect-based Sentiment Analysis. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI, 2022, pp. 4096–4103.
38. Wu, S.; Fei, H.; Ren, Y.; Ji, D.; Li, J. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, pp. 3957–3963.
39. Li, B.; Fei, H.; Liao, L.; Zhao, Y.; Teng, C.; Chua, T.; Ji, D.; Li, F. Revisiting Disentanglement and Fusion on Modality and Context in Conversational Multimodal Emotion Recognition. Proceedings of the 31st ACM International Conference on Multimedia, MM, 2023, pp. 5923–5934.
40. Fei, H.; Liu, Q.; Zhang, M.; Zhang, M.; Chua, T.S. Scene Graph as Pivoting: Inference-time Image-free Unsupervised Multimodal Machine Translation with Visual Scene Hallucination. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 5980–5994.
41. Li, J.; Xu, K.; Li, F.; Fei, H.; Ren, Y.; Ji, D. MRN: A Locally and Globally Mention-Based Reasoning Network for Document-Level Relation Extraction. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 1359–1370.
42. Fei, H.; Wu, S.; Ren, Y.; Zhang, M. Matching Structure for Dual Learning. Proceedings of the International Conference on Machine Learning, ICML, 2022, pp. 6373–6391.
43. Cao, H.; Li, J.; Su, F.; Li, F.; Fei, H.; Wu, S.; Li, B.; Zhao, L.; Ji, D. OneEE: A One-Stage Framework for Fast Overlapping and Nested Event Extraction. Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 1953–1964.

44. Fei, H.; Li, F.; Li, B.; Ji, D. Encoder-Decoder Based Unified Semantic Role Labeling with Label-Aware Syntax. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 12794–12802.
45. Li, B.; Fei, H.; Li, F.; Wu, Y.; Zhang, J.; Wu, S.; Li, J.; Liu, Y.; Liao, L.; Chua, T.S.; Ji, D. DiaASQ: A Benchmark of Conversational Aspect-based Sentiment Quadruple Analysis. *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 13449–13467.
46. Wu, S.; Fei, H.; Cao, Y.; Bing, L.; Chua, T.S. Information Screening whilst Exploiting! Multimodal Relation Extraction with Feature Denoising and Multimodal Topic Modeling. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 14734–14751.
47. Fei, H.; Wu, S.; Ren, Y.; Li, F.; Ji, D. Better Combine Them Together! Integrating Syntactic Constituency and Dependency Representations for Semantic Role Labeling. *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021*, 2021, pp. 549–559.
48. Wu, S.; Fei, H.; Zhang, H.; Chua, T.S. Imagine That! Abstract-to-Intricate Text-to-Image Synthesis with Scene Graph Hallucination Diffusion. *Advances in Neural Information Processing Systems* **2024**, 36.
49. Fei, H.; Wu, S.; Ji, W.; Zhang, H.; Chua, T.S. Empowering dynamics-aware text-to-video diffusion with large language models. *arXiv preprint arXiv:2308.13812* **2023**.
50. Qu, L.; Wu, S.; Fei, H.; Nie, L.; Chua, T.S. Layoutlm-t2i: Eliciting layout guidance from llm for text-to-image generation. *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 643–654.
51. Fei, H.; Ren, Y.; Ji, D. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management* **2020**, 57, 102311.
52. Li, J.; Fei, H.; Liu, J.; Wu, S.; Zhang, M.; Teng, C.; Ji, D.; Li, F. Unified Named Entity Recognition as Word-Word Relation Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, pp. 10965–10973.
53. Fei, H.; Chua, T.; Li, C.; Ji, D.; Zhang, M.; Ren, Y. On the Robustness of Aspect-based Sentiment Analysis: Rethinking Model, Data, and Training. *ACM Transactions on Information Systems* **2023**, 41, 50:1–50:32.
54. Zhao, Y.; Fei, H.; Cao, Y.; Li, B.; Zhang, M.; Wei, J.; Zhang, M.; Chua, T. Constructing Holistic Spatio-Temporal Scene Graph for Video Semantic Role Labeling. *Proceedings of the 31st ACM International Conference on Multimedia, MM*, 2023, pp. 5281–5291.
55. Fei, H.; Ren, Y.; Zhang, Y.; Ji, D. Nonautoregressive Encoder-Decoder Neural Framework for End-to-End Aspect-Based Sentiment Triplet Extraction. *IEEE Transactions on Neural Networks and Learning Systems* **2023**, 34, 5544–5556.
56. Zhao, Y.; Fei, H.; Ji, W.; Wei, J.; Zhang, M.; Zhang, M.; Chua, T.S. Generating Visual Spatial Description via Holistic 3D Scene Understanding. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 7960–7977.
57. Fei, H.; Li, B.; Liu, Q.; Bing, L.; Li, F.; Chua, T.S. Reasoning Implicit Sentiment with Chain-of-Thought Prompting. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2023, pp. 1171–1182.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.