Cureus Journal of Business and Economics

Open Access Original Article

Received 04/03/2025
Review began 04/08/2025
Review ended 05/05/2025
Published 05/29/2025

DOI: https://doi.org/10.7759/s44404-025-04183-4

Optimizing Exchange-Traded Fund Predictions Using Autoregressive Recurrent Networks With Sentiment Scores and Temporal Keyword-Enhanced Integration

Waleed Soliman [1] , Zhiyuan Chen [1] , Colin Johnson [2] , Sabrina Wong [3]

1. School of Computer Science, University of Nottingham Malaysia, Semenyih, MYS 2. School of Computer Science, University of Nottingham UK, Nottingham, GBR 3. Management, Hy-Alpha Sdn Bhd, Kuala Lumpur, MYS

Corresponding authors: Waleed Soliman, hcxwm1@nottingham.edu.my, Zhiyuan Chen, zhiyuan.chen@nottingham.edu.my

Abstract

Exchange-traded funds (ETFs) have emerged as a favored investment vehicle due to their liquidity, affordability, and diversified exposure across asset classes. However, their price dynamics, influenced by macroeconomic trends, market sentiment, and real-time news, challenge traditional forecasting methods. This study advances ETF price prediction by integrating the probabilistic forecasting with autoregressive recurrent networks (DeepAR) with normalized sentiment scores (NSS) and a novel temporal keyword-enhanced sentiment framework. Temporal keyword-enhanced sentiment combines trending keywords from Google Trends with user-defined macroeconomic terms to enrich news sentiment capture, while NSS consolidates conflicting sentiments into a daily signal. Applied to eight U.S. ETFs spanning bonds, equities, and commodities, the enhanced DeepAR-NSS model consistently outperformed the baseline DeepAR, reducing mean absolute percentage error by 22-49.7% and root mean square error by up to 40.6%. Notable improvements were observed in bond ETFs (e.g., GOVT) and volatile assets (e.g., GLD), though high-risk ETFs like JNK exhibited increased root mean square error variability. These findings highlight the efficacy of sentiment integration in financial forecasting and underscore the need for asset-specific model refinement.

Categories: Banking and financial services, Fintech and Cryptocurrencies, Risk Management
Keywords: financial forecasting, fintech, machine learning, sentiment analysis, fintech applications, deepar, macroeconomic indicators, finbert

Introduction

Exchange-traded funds (ETFs) have gained prominence among investors due to their liquidity, cost-effectiveness, and adaptability, providing diversified exposure across asset classes such as bonds, equities, and commodities (Lettau and Madhavan, 2018). Probabilistic forecasting with autoregressive recurrent networks (DeepAR) is a probabilistic deep learning model designed for time-series forecasting, leveraging recurrent neural networks to capture temporal dependencies and predict future values based on historical data (Salinas et al., 2017). Normalized sentiment scores (NSS) refer to quantified measures of sentiment derived from news articles, typically ranging from negative to positive, which reflect market perceptions and investor reactions to events. The temporal keyword-enhanced sentiment (TKES) framework is a novel approach introduced in this study, integrating trending keywords from public search data with macroeconomic terms to enhance the relevance and timeliness of sentiment analysis. These keywords and tools are pivotal in understanding and predicting ETF price movements, as they bridge structured financial data with unstructured textual information.

The intricate dynamics of ETF price movements are driven by macroeconomic trends (Soliman et al., 2023) and (Chen et al., 1986), market sentiment, and real-time news (Tetlock, 2007), presenting substantial challenges to conventional forecasting approaches (Poon and Granger, 2003). Recent advancements in deep learning, notably time-series forecasting models like DeepAR (Salinas et al., 2017), have demonstrated potential in modeling temporal dependencies and incorporating exogenous variables within financial datasets. However, these models often struggle to effectively utilize unstructured data, such as news sentiment (Bollen et al., 2011), or to accommodate the diverse characteristics of ETFs spanning various asset classes and risk profiles (Fama and French, 1993).

Building on our prior research, which explored ETF price prediction using sentiment scores extracted from news article summaries via financial bidirectional encoder representations from transformers (FinBERT) sentiment analysis, we previously integrated these quantified sentiment metrics into machine learning predictive frameworks. That study employed DeepAR as a probabilistic forecasting model and benchmarked its performance against long short-term memory (LSTM), revealing DeepAR's superior

How to cite this article
Soliman W, Chen Z, Johnson C, et al. (May 29, 2025) Optimizing Exchange-Traded Fund Predictions Using Autoregressive Recurrent Networks With Sentiment Scores and Temporal Keyword-Enhanced Integration. Cureus J Bus Econ 2 : es44404-025-04183-4. DOI https://doi.org/10.7759/s44404-025-04183-4

efficacy and the notable enhancement in prediction accuracy achieved through sentiment score integration (Soliman et al., 2025). However, a key research gap identified from past studies is the limited incorporation of dynamic, real-time keyword trends, and the lack of a robust mechanism to consolidate conflicting news sentiments into a unified signal for diverse ETF types. While prior work has utilized static sentiment inputs or ticker-specific keywords (Tetlock, 2007) and (Bollen et al., 2011), it often overlooks broader market narratives and temporal shifts in public interest that could enhance forecasting accuracy across asset classes.

The objective of this study is to address this gap by enhancing the DeepAR model with NSS and the TKES framework, aiming to improve ETF price prediction accuracy by capturing a more comprehensive and temporally relevant sentiment signal. This approach seeks to outperform baseline models by integrating real-time trending keywords and aggregated sentiment scores, tailored to the unique dynamics of ETFs across bonds, equities, and commodities.

Related work

The role of news in financial forecasting is well-established, with Tetlock (Tetlock, 2007) demonstrating that media content, such as the Wall Street Journal's "Abreast of the Market" column, can predict market movements by reflecting investor sentiment. This paper, published in 2007, uses textual analysis to quantify sentiment, showing a significant correlation with stock market returns, which are relevant for ETFs given their underlying assets. Bollen et al. (Bollen et al., 2011), in their 2011 study, extend this to social media, using X posts to predict stock market trends based on mood indicators like "calm" or "alert," highlighting the timeliness of real-time data for forecasting.

Keywords are crucial for filtering relevant news, and Orăştean et al. (Orăştean et al., 2024) demonstrate that Google Trends data, which captures public interest through search queries, can be used to improve the prediction of Bitcoin and Ethereum prices when combined with other financial metrics. This highlights the influence of public interest on market dynamics. Our TKES technique similarly leverages Google Trends to identify trending topics, thereby enhancing the capture of news sentiment for ETF forecasting and addressing the limitations of relying solely on specific ETF tickers. Huang et al. (Huang et al., 2020) further support this by using Google search volume to predict stock movements, suggesting that keyword-based approaches can enhance model relevance.

Sentiment analysis is facilitated by tools like FinBert, introduced by (Araci, 2019), which fine-tunes bidirectional encoder representations from transformers for financial text to classify news sentiment, providing scores from -1 to 1. This is complemented by Shapiro et al. (Shapiro et al., 2022), which discusses sentiment-scoring model for sentiment aggregation, ensuring models can handle overlapping news with conflicting sentiments. Our NSS approach aggregates news into a single daily score by applying weights, providing a consolidated sentiment signal that reduces noise and enhances interpretability for time-series models.

Research Method

This section presents the methodological framework for enhancing ETF price prediction through advanced machine learning techniques. It introduces the TKES technique, which integrates trending keywords from Google Trends with user-defined macroeconomic terms to improve news sentiment capture. The NSS is developed to aggregate overlapping sentiments into a robust daily signal. Data sources include the Yahoo Finance application programming interface (API) for ETF price data (Yahoo Finance, 2025), the Federal Reserve Economic Data (FRED) API for macroeconomic indicators (Federal Reserve Bank of St. Louis, 2025), and the Gnews API for news articles (Gnews, 2025).

Temporal keyword-enhanced sentiment

The TKES technique leverages a broader set of macroeconomic and thematic keywords (e.g., "AI," "Inflation," "Interest Rate") to capture news records that are used for sentiment enrichment, offering several advantages over relying solely on ticker-specific keywords (e.g., "VOO").

In this research, we introduce a novel TKES technique by integrating the latest trending keywords from Google Trends API with user-defined manual keywords. This approach aims to improve the capture of news sentiment for financial forecasting, particularly for ETFs. As illustrated in Figure *1* The flowchart begins with a user initiating the process, which branches into two primary data sources:

• Google Trends API: This retrieves trending data categorized into top topics, top queries, rising topics, and rising queries.

• Manual Keywords: A predefined list of keywords (e.g., inflation, employment, interest rate, recession, AI) is provided by the user based on domain knowledge.

These sources converge to form a set of selected keywords, which are then used to refresh news data for sentiment analysis within the TKES framework.

Google Trends API provides access to real-time data on trending topics and queries (e.g., top topics, rising queries), which reflect current public interest and market discourse. For instance, if "AI" emerges as a Top Topic due to recent technological advancements, the TKES process can incorporate this keyword to capture sentiment that may influence ETF prices, such as VOO, which is sensitive to tech sector developments. This dynamic inclusion ensures that TKES remains responsive to evolving market narratives that ticker-specific keywords like "VOO" might miss.

The manual keywords (e.g., inflation, employment, interest rate) are predefined based on the user's understanding of macroeconomic factors that impact ETFs. These keywords, as shown in the flowchart, ensure that critical economic indicators are consistently monitored, even if they are not currently trending on Google. By including these keywords, TKES ensures a comprehensive sentiment profile that captures both persistent economic drivers and transient trends, addressing gaps that a purely data-driven approach might overlook. Finally, the process involves selecting keywords from both Google Trends (top topics, rising queries) and the manual list to form a refined set of "Selected Keywords." This step, as depicted in the flowchart, allows for a balanced approach where trending topics (e.g., "AI" or "Magnificent SEVEN") are combined with user-defined keywords (e.g., "Recession"). This ensures that the TKES process captures both emerging market buzz and established economic indicators, enhancing the relevance of the sentiment data.

To showcase TKES mechanism, Figures *2* and 3 offer insights into the application of TKES technique for capturing news sentiment related to VOO ETF, focusing on the month of January 2024. The first plot illustrates the timeline of record counts by keyword, including "AI," "Artificial Intelligence," "Bond Yields," "Employment," "Inflation," "Interest Rate," "Magnificent SEVEN," "Recession," and "VOO," while the second plot displays the total daily count of records for the same period.

A noticeable surge occurs around January 17-25, with multiple keywords peaking simultaneously. "AI" records the highest count, reaching 10 on January 29, followed by "Employment" and "Inflation" with counts of 6 on January 25. "Bond Yields," "Interest Rate," and "Recession" also show increased activity, with counts of 3-4 during this period. In contrast, "VOO" remains consistently low, with counts rarely exceeding 1.

On the other hand, the total daily count plot shows a steady baseline of 2-3 records per day from January 1 to January 17, with minor fluctuations. A significant spike occurs around January 21-25, peaking at 10 records on January 25, followed by a sharp increase to 17 on January 29, driven primarily by the "AI" keyword. This surge aligns with the keyword-specific plot, indicating that the increase in total records is largely attributable to heightened discussion around macroeconomic themes, rather than ticker-specific news like "VOO."

The sparse activity for "VOO" highlights the inadequacy of ticker-specific keywords in capturing sufficient sentiment data, whereas the TKES approach fills these gaps by leveraging a diverse set of keywords, providing a more comprehensive and temporally relevant sentiment signal. This method not only addresses data sparsity but also enhances the model's ability to capture broader market dynamics, making it a superior choice for sentiment-driven forecasting in financial markets.

Normalized sentiment score (NSS)

Overlapping news (e.g., multiple articles on "Inflation" and "AI") may contain conflicting sentiments (e.g., positive economic outlook vs. negative tech concerns). The NSS aggregates these into a single daily score by applying weights, providing a consolidated sentiment signal that reduces noise and enhances interpretability for time-series models like DeepAR.

The NSS calculation can be formalized using the following mathematical expressions.

Weighted sentiment score (WSS) for a single record is represented by Equation 1:

$$WSS_i = S_i \times W_L \qquad (1)$$

where S_i is the SentimentScore for the i-th news article (e.g., a value between -1 and 1 from FinBERT); W_L is the weight assigned to the sentiment label L (e.g., W_{Negative} = 1, W_{Neutral} = 2, W_{Positive} = 3); i indexes each news article on a given day.

Daily NSS is represented by Equation 2:

$$NSS_d = \frac{(\Sigma_{i=1}^{n_d} WSS_i)}{(\Sigma_{i=1}^{n_d} |S_i|)} \qquad (2)$$

where NSS_d is the normalized sentiment score for day d, n_d is the number of news articles published on day d, $\Sigma_{i=1}^{n_d} WSS_i$ is the sum of weighted sentiment scores for all articles on day d, and $\Sigma_{i=1}^{n_d} |S_i|$ is the sum of the absolute values of sentiment scores, serving as a normalization factor to account for the magnitude of sentiment intensity, ensuring the score remains bounded (e.g., between 0 and 3, depending on weights).

Normalization rationale

The denominator $\Sigma_{i=1}^{n_d} |S_i|$ normalizes the weighted sum by the total sentiment intensity, preventing bias from days with a high volume of news or extreme scores. This ensures that NSS_d reflects a balanced sentiment trend rather than raw frequency or amplitude.

FinBERT outputs a sentiment score (e.g., 0 to 1) and a sentiment label (negative, neutral, positive). We assigned sentiment weight classes (1, 2, 3) to amplify these scores, allowing the NSS to reflect the perceived market impact. For instance, a FinBERT score of 0.8 with a "positive" label becomes 0.8 × 3 = 2.4, significantly influencing the daily NSS.

Table *1* presents empirical evidence from news data spanning December 23 to December 28, 2024, supporting the use of weights (1, 2, 3) by demonstrating their ability to enhance the NSS's sensitivity to sentiment variations. On days with mixed sentiments, such as December 23 (5 negative, 1 neutral, 2 positive articles), the weighted NSS of 1.56 reflects the positive articles' amplified impact (weight = 3), contrasting with an unweighted NSS of 1.00 that fails to differentiate sentiment classes. Similarly, on December 26 (4 negative, 1 neutral, 4 positive articles), the weighted NSS of 2.12 captures a positive-leaning sentiment driven by headlines like “China raises its estimate for the size of its economy in 2023” while the unweighted NSS remains 1.00, offering no insight into sentiment dynamics. This shows that the weights produce a more nuanced and market-relevant signal compared to an unweighted approach.

Furthermore, the weights improve interpretability and reduce noise for time-series models like DeepAR. On December 28, with all neutral articles, the weighted NSS of 2.00 clearly indicates neutrality, while on December 27 (5 negative, 1 neutral, 2 positive), the NSS of 1.51 balances negative dominance with positive contributions, unlike the unweighted NSS's constant 1.00. Testing alternative weights (e.g., 1, 1.5, 2) on December 26 yielded a lower NSS of 1.51, underemphasizing positive sentiment's impact. These results justify the weights (1, 2, 3) as they align NSS with market dynamics, enhance trend detection, and provide a robust signal for financial analysis.

Data

The dataset encompasses eight distinct U.S. ETFs, each characterized by its own inception date, as detailed in Table *2*. These ETFs represent a diverse range of asset classes. Specifically, BND and GOVT are Bond ETFs, comprising fixed-income securities; GLD is a Gold ETF, designed to track the price movements of the precious metal; IVOO and VIOO are Equity ETFs, capturing mid-cap and small-cap U.S. companies, respectively; JNK is a High-Yield Corporate Bond ETF, indicative of elevated risk and return potential; and SPY and VOO are Large-Cap Equity ETFs, mirroring the S&P 500 Index to offer broad market exposure.

Furthermore, the risk profiles of these ETFs exhibit considerable variation. For instance, BND and GOVT are categorized as lower-risk investments, attributable to their focus on government or high-quality bonds. In contrast, JNK entails higher risk due to its allocation to lower-rated, high-yield corporate bonds. Meanwhile, IVOO, VIOO, SPY, and VOO fall within a moderate-to-high risk spectrum, reflecting their exposure to equity investments and inherent market volatility.

The eight ETF tickers share a common set of foundational attributes, including "Open," "Close," "High," "Low," "Volume," and "Adjusted Price." Data for these attributes were sourced from the Yahoo Finance API, with the temporal coverage determined by each ticker's respective inception date. The availability of historical data for these ETFs also differs significantly. SPY, for example, has data extending back to 1994, providing a long-term perspective on its performance. Conversely, the historical data for BND, GLD, GOVT, IVOO, JNK, VIOO, and VOO commence from 2005 onward, offering a more contemporary context for evaluating their performance, as illustrated in Table *2*.

To improve the model's responsiveness to shifts in economic policy decisions, a curated set of U.S. macroeconomic and monetary indicators was incorporated into the base attributes, resulting in the creation of a “merged dataset.” This dataset includes variables such as the Federal Reserve Fund Rate, the Consumer Price Index, and Federal Reserve Total Assets, as outlined in Table *3*. The selection of these indicators was informed by correlation analyses conducted across the various tickers, complemented by insights from domain experts. The data for these features were primarily obtained from the Federal

Reserve Economic Data API.

The news data employed in this study were obtained from the Gnews API. The collection process entailed real-time queries of news articles, with the resulting data organized on a daily basis, beginning from the inception date of the corresponding ETF under analysis.

Experiment

This study adhered to a consistent process for formulating the base dataset, drawing on three distinct APIs - Yahoo Finance, FRED, and Gnews - as primary data sources. The initial stages of this process involved data cleaning and preprocessing to ensure quality and usability. In addition, calculated features such as the "3-Day Moving Average," "10-Day Moving Average," "Buy-Sell on Close," "Buy-Sell on Open," "Day of the Year," and "Month of the Year" were integrated into the foundational dataset to capture additional trends and patterns within the data.

Following the incorporation of macroeconomic indicators and base attributes, data standardization was performed using the Standard Scaler and Min-Max Scaler techniques to normalize the dataset. A subsequent correlation matrix analysis was conducted to detect and remove highly correlated features, culminating in the creation of the "Merged Dataset". This unified dataset served as the basis for model evaluation across various experiments.

The DeepAR model was evaluated directly using the “Merged Dataset”, establishing it as the baseline model. In contrast, the DeepAR-NSS model required further dataset enhancement. To examine the impact of integrating news sentiment on model performance, an “Enriched Dataset” was developed. This process involved retrieving ETF-related news articles via the Gnews API, followed by data cleaning and tokenization. The news data were then analyzed using "FinBERT" to assign sentiment classifications - namely "Positive", "Negative", and "Neutral” with each news record receiving a value ranging from 0 to 1. Given that multiple news records could correspond to a single day, a NSS score was computed to determine the prevailing sentiment class for each daily entry. For time steps lacking news records, the TKES method was applied to identify and incorporate relevant news content for those periods.

For both the DeepAR and DeepAR-NSS models, the dataset was divided, with 20% designated for testing and the remaining 80% utilized for training. During the training phase, a sliding window approach was adopted, wherein the models were trained on 30 time-steps to forecast the subsequent 15 time-steps. The model architecture consisted of a fully connected RNN with seven hidden layers and employed a rectified linear unit (ReLU) activation function. Training was conducted using the Adam optimization algorithm, configured with a learning rate of 0.001, over a total of 700 epochs, performance evaluations were carried out at both 50% and 90% confidence intervals.

In the present study, the mean absolute percentage error (MAPE) was utilized to assess the accuracy of the time-series models in comparison to actual values. MAPE is defined as the average absolute percent error of each time period, calculated as the ratio of the absolute difference between actual and predicted values and the actual values (Scikit-learn, 2022). Additionally, the root mean square error (RMSE) was employed to evaluate the quality of the predictions. RMSE provides information on the deviation of the predictions from the true values, as determined through the Euclidean distance (RMSE, 2022).

Table *1* displays financial news articles from December 23 to 28, 2024, used to evaluate the effectiveness of discrete sentiment weights (1 for negative, 2 for neutral, 3 for positive) in calculating NSS. Columns include date, headline, sentiment class, sentiment score (0 to 1 from FinBERT), and daily weighted NSS. The NSS formula amplifies positive and neutral sentiments, producing varied scores (1.00 to 2.12) that reflect market trends.

Date	News Headline	Sentiment Class	Sentiment Score	Weighted NSS
2024-12-23	Lego brick sets take new shape, keep toy industry afloat	Negative	0.95	1.56
2024-12-23	Rigetti Computing Launches 84-Qubit System	Positive	0.85	1.56
2024-12-23	An analyst looks ahead to how the US economy might fare under Trump	Positive	0.81	1.56
2024-12-23	The Container Store files for Chapter 11 bankruptcy protection	Negative	0.97	1.56
2024-	Trump's 25% tariffs an existential threat to Canada's auto industry	Negative	0.94	1.56

12-23				
2024-12-23	Biden launches new Chinese chips trade probe, will hand off to Trump	Neutral	0.49	1.56
2024-12-23	'Santa Rally' MIA as Bitcoin Falls to Lowest Price in a Month	Negative	0.96	1.56
2024-12-23	Amazon, Starbucks and the future of worker organizing	Negative	0.86	1.56
2024-12-24	Party City Fired Everyone Without Severance Right Before Christmas	Negative	0.94	1.00
2024-12-24	Huge twist in abrupt closure of 800-store national chain as bankruptcy was triggered by its own experts	Negative	0.70	1.00
2024-12-24	Big banks sue Federal Reserve over minimum cash requirements	Negative	0.51	1.00
2024-12-25	Will President-Elect Donald Trump's Plan to Implement Tariffs Cause Stocks to Plunge? Here's What History Tells Us.	Negative	0.93	1.00
2024-12-26	China raises its estimate for the size of its economy in 2023	Positive	0.90	2.12
2024-12-26	21 States To Raise Minimum Wage For 2025	Positive	0.88	2.12
2024-12-26	US applications for unemployment benefits hold steady, but continuing claims rise to 3-year high	Positive	0.91	2.12
2024-12-26	Stock market today: Stocks edge lower after a holiday pause for US markets	Negative	0.98	2.12
2024-12-26	Turkey's central bank lowers key interest rate to 47.5%	Negative	0.96	2.12
2024-12-26	Recurring jobless claims rise to three-year high	Negative	0.93	2.12
2024-12-26	Unemployed workers having trouble finding a job, latest claims data suggest	Negative	0.56	2.12
2024-12-26	Average rate on a 30-year mortgage climbs to 6.85%, highest since July, as bond yields keep rising	Positive	0.94	2.12
2024-12-26	Richard Parsons, Former Time Warner CEO, Dies at 76	Neutral	0.70	2.12
2024-12-26	Holiday shoppers increased spending by 3.8% despite higher prices	Positive	0.94	2.12
2024-12-27	China's industrial profits extend decline to a fourth straight month, dropping 7.3% in November	Negative	0.98	1.51
2024-12-27	Why is crypto down today? Looking at Bitcoin, FED, and past week's sell-off effects!	Neutral	0.61	1.51
2024-12-27	Major changes coming to Social Security regardless of Trump’s presence	Negative	0.80	1.51
2024-12-27	The Trump Economy Begins: 4 Money Moves Boomers Shouldn't Make Until After Inauguration Day	Positive	0.65	1.51
2024-12-27	Florida condo owners look at higher costs as new regulations take effect in the new year	Positive	0.70	1.51
2024-12-27	Markets stumble as Wall Street sells off Big Tech	Negative	0.96	1.51
2024-12-27	Stock market today: Wall Street slips as the 'Magnificent 7' weighs down the market	Negative	0.97	1.51
2024-				

12-27	Mixed reactions to Park City Mountain patrol strike from skiers	Negative	0.86	1.51
2024-12-28	10 Stocks For 2025	Neutral	0.75	2.00
2024-12-28	Is the Stock Market Going to Crash in 2025 2 Historically Flawless Indicators Paint a Clear Picture.	Neutral	0.79	2.00
2024-12-28	Donald Trump's tariff plan is really just a tax on Americans	Neutral	0.92	2.00
2024-12-28	Here's What Could Happen to Your Money in Trump's First 40 Days in Office	Neutral	0.90	2.00

TABLE 1: Daily News Sentiment Records With Weighted NSS Scores (December 23–28, 2024).

Source: (Gnews, 2025).

NSS, Normalized Sentiment Score

Table 2 shows eight ETFs with their tickers, codes, and data start dates, ranging from 1994 to 2012, covering bonds, gold, and various stock market indices.

No.	Code	Name	Data Start Date
1	BND	Vanguard Total Bond Market Index Fund	01/05/2007
2	GLD	SPDR Gold Shares	01/01/2005
3	GOVT	iShares U.S. Treasury Bond	01/03/2012
4	IVOO	Vanguard S&P Mid-Cap 400 Index Fund	01/10/2010
5	JNK	SPDR Bloomberg High Yield Bond	01/01/2008
6	VIOO	Vanguard S&P Small-Cap 600 Index Fund	01/10/2010
7	SPY	SPDR S&P 500 ETF Trust	01/01/1994
8	VOO	S&P 500 ETF	01/10/2010

TABLE 2: Selected ETF Tickers.

Source: (Yahoo Finance, 2025).

ETF, Exchange-Traded Fund

Table 3 lists 16 macroeconomic indicators with their codes and names, covering key economic metrics such as interest rates, inflation, currency, GDP, employment, Fed assets, bond yields, and market volatility.

Code	Name
FEDFUNDS	Federal Reserve Fund Rate
CPALTT01USM661S	Consumer Price Index
DTWEXBGS	Nominal Broad U.S. Dollar Index
WALCL	Federal Reserve Total Assets (QE)
GDPC1	Real Gross Domestic Product
DGS3MO	Market Yield on U.S. Treasury Securities at 3-Month Constant Maturity
EMVOVERALLEMV	Overall Equity Market Volatility Tracker
RBUSBIS	Real Broad Effective Exchange Rate
PAYEMS	Employment Level
IRLTLT01USM156N	Long-Term Government Bond Yields: 10-year
DGS2	Market Yield on U.S. Treasury Securities at 2-Year Constant Maturity
DGS10	Market Yield on Treasury Securities at 10-Year
DFII10	Market Yield on Treasury Securities at 10-Year - inflation
VIXCLS	CBOE Volatility Index
EMVMACROINTEREST	Interest Rates Equity Market Volatility Tracker
EMVMACROBUS	Investment Sentiment Equity Market Volatility Tracker

TABLE 3: Selected Macroeconomic Indicators.

Source: (Federal Reserve Bank of St. Louis, 2025).

Figure *1* illustrates a process for selecting keywords using Google Trends API. A user retrieves data on top topics, top queries, rising topics, and rising queries from the API. They also define manual keywords to complement the API data. The user then selects relevant keywords from both sources, combining them into a set of "Selected Keywords," which are used to refresh news content.

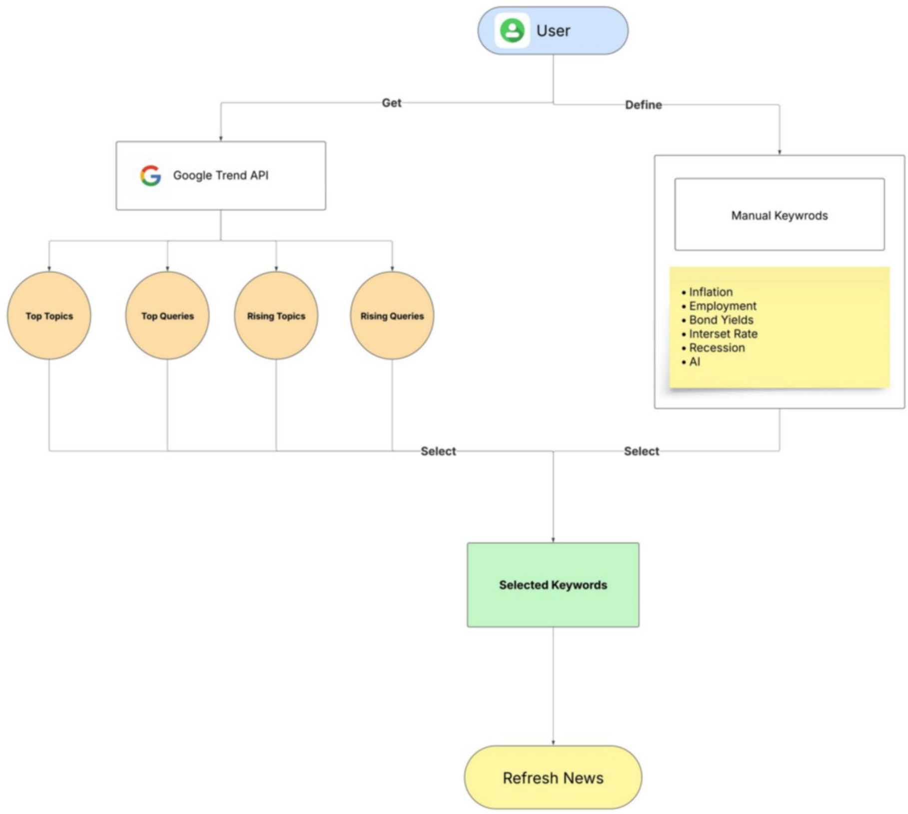

FIGURE 1: Flowchart of Temporal Keyword-Enhanced Sentiment Technique.

Source: Author.

Figure 2 tracks eight keywords over January 2024. VOO maintains a consistently low count, fluctuating between 0 and 1 throughout the month. In comparison, most other keywords also show low counts (0-2) but with more variability: "AI" spikes dramatically to 10 on January 29, "Employment" and "Inflation" peak at 4-5 around January 21-25, and "Bond Yields" reaches 2 early in the month. Overall, VOO has the least activity compared to the more volatile trends of AI, Employment, and Inflation.

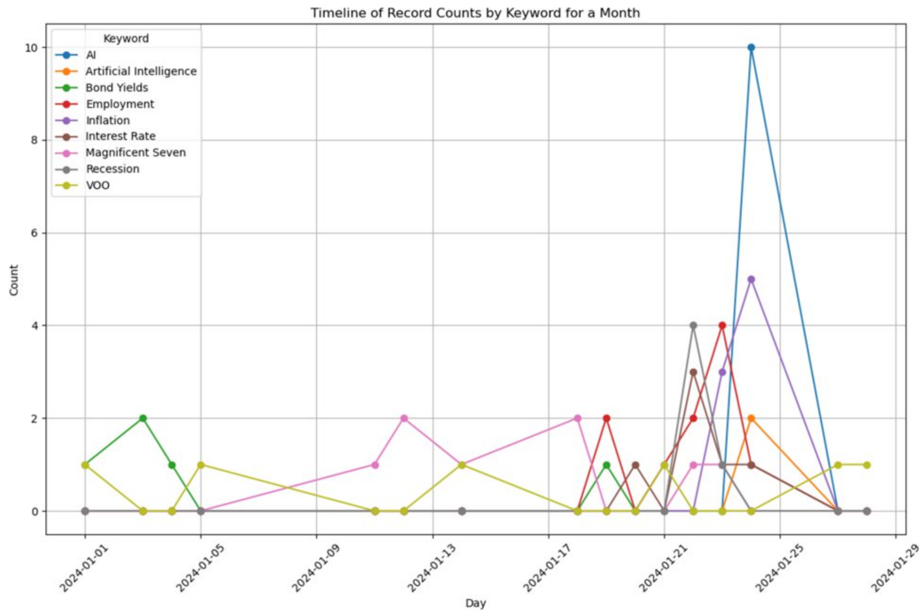

FIGURE 2: The Timeline of Record Counts by Keyword.

Source: Author.

Figure 3 shows the total daily count of news records per day from January 1 to January 17, with minor fluctuations. A significant spike occurs around January 21-25 driven primarily by the "AI" keyword indicating that the increase in total records is largely attributable to heightened discussion around macroeconomic themes, rather than ticker-specific news like "VOO."

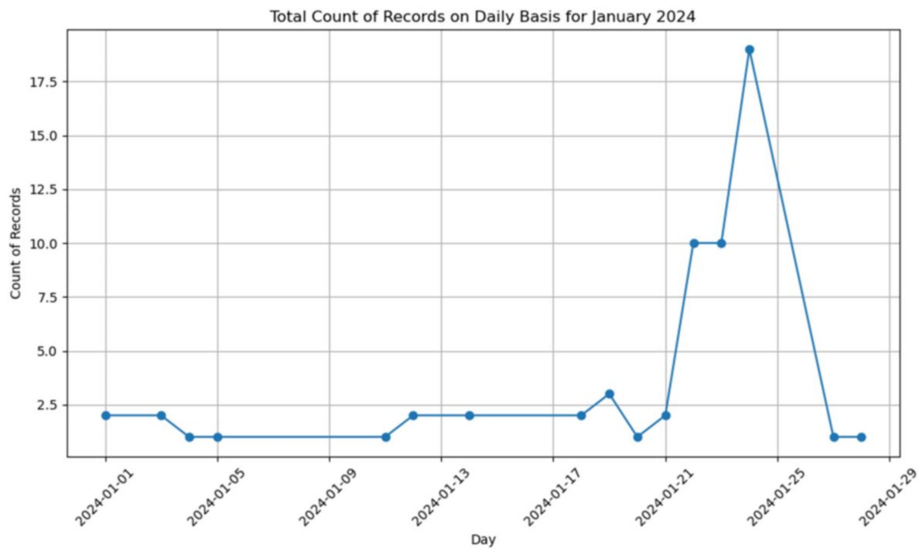

FIGURE 3: The Total Daily Count of News Records.

Source: Author.

Results

The results, as reported in Table 4, demonstrate that the DeepAR-NSS model, enriched with news sentiment data, consistently outperformed the baseline DeepAR model across all eight ETFs in terms of MAPE, indicating enhanced predictive accuracy. Notably, the most substantial MAPE improvement occurred for GOVT (from 0.0167 to 0.0084, a 49.7% reduction), suggesting that news sentiment integration significantly enhances forecasting for bond ETFs with lower risk profiles. GLD also exhibited a marked improvement (from 0.0506 to 0.0377, a 25.5% reduction), reflecting the influence of sentiment on gold price volatility. For equity ETFs like SPY, VOO, IVOO, and VIOO, DeepAR-NSS consistently reduced MAPE by 22-39%, underscoring the value of sentiment data in capturing market dynamics. Regarding RMSE, DeepAR-NSS generally improved performance, with notable reductions for VOO (32.4%), SPY (27.4%), IVOO (27.3%), GOVT (40.6%), GLD (26.1%), and BND (34.9%). However, an exception emerged with JNK, where RMSE slightly increased from 0.3355 to 0.3930 despite a lower MAPE, possibly indicating that news sentiment introduced additional variability in predictions for this high-yield bond ETF. This anomaly warrants further investigation into the interplay between sentiment and high-risk asset classes.

Key findings highlight that integrating news sentiment via the DeepAR-NSS model enhances forecasting accuracy, particularly for ETFs sensitive to macroeconomic shifts (e.g., GOVT, BND) and volatile assets (e.g., GLD). The consistent MAPE reductions across all tickers affirm the efficacy of sentiment enrichment, while the varied RMSE outcomes suggest that the impact of sentiment may differ by asset type, with potential trade-offs in precision for certain high-risk ETFs like JNK. These results underscore the importance of tailoring model enhancements to specific ETF characteristics.

The same superior performance can be seen in Figure *4* which compares the predictive performance of the DeepAR and DeepAR-NSS models for different tickers. Across all ETFs, the DeepAR-NSS model (left subplots) consistently demonstrates predictions that more closely align with the actual values compared to the DeepAR model (right subplots). This visual observation aligns with the previously discussed numerical results, where DeepAR-NSS exhibited lower MAPE values for all tickers. For instance, in the case of GOVT, the DeepAR-NSS predictions closely follow the actual price trajectory, with narrower confidence intervals, whereas the DeepAR model shows wider deviations, particularly during periods of fluctuation (e.g., late 2023). This supports the earlier finding of a 49.7% MAPE reduction for GOVT with DeepAR-NSS.

The confidence intervals for DeepAR-NSS are generally narrower than those for DeepAR, indicating greater predictive certainty. For example, for GLD, the 90% CI in the DeepAR-NSS plot is more constrained, reflecting the 25.5% MAPE improvement and 26.1% RMSE reduction reported earlier. However, for JNK, the DeepAR-NSS confidence intervals appear slightly wider during certain periods (e.g., early 2024), which may explain the observed increase in RMSE (from 0.3355 to 0.3930) despite a lower MAPE. This suggests that while sentiment integration improved point prediction accuracy, it introduced some uncertainty in the prediction range for this high-yield bond ETF.

For bond ETFs (BND, GOVT, JNK), DeepAR-NSS shows a clear improvement in tracking price movements, particularly during stable periods. For instance, BND’s DeepAR-NSS plot captures subtle upward trends more accurately than DeepAR, consistent with the 34.9% RMSE reduction. Gold ETF (GLD) exhibits higher volatility, and DeepAR-NSS better captures these fluctuations, as seen in the sharper peaks and troughs aligning with actual values. This aligns with the significant MAPE reduction (25.5%) for GLD. Equity ETFs (IVOO, SPY, VIOO, VOO) display more pronounced improvements with DeepAR-NSS, particularly during market dips (e.g., late 2023 for SPY). The tighter alignment of predictions with actual values reflects the 22-39% MAPE reductions noted in the numerical results.

Finally, Figure *4* cover a period of market dynamics, with visible trends and corrections (e.g., a dip in late 2023 across most ETFs, followed by a recovery in early 2024). DeepAR-NSS appears more adept at capturing these shifts, likely due to the incorporation of news sentiment, which may reflect real-time market reactions to events. For VOO, the DeepAR-NSS model accurately predicts the recovery phase in early 2024, whereas DeepAR underestimates the upward trend, consistent with the 32.4% RMSE improvement for DeepAR-NSS.

In general, DeepAR-NSS outperforms DeepAR across all ETFs, particularly in capturing short-term fluctuations and reducing prediction uncertainty, as evidenced by narrower confidence intervals. The most notable improvements are observed for bond ETFs (e.g., GOVT, BND) and volatile assets like GLD, where sentiment data likely enhances the model’s ability to reflect market sentiment-driven price movements. However, for JNK, the wider confidence intervals in DeepAR-NSS suggest that sentiment integration, while improving MAPE, may introduce variability in prediction ranges for high-risk assets, warranting further investigation.

Table *4* presents forecasting performance for different ETFs using two models: DeepAR and DeepAR-NSS. Performance is evaluated using MAPE and RMSE metrics. Across all experiments, the DeepAR-NSS model consistently outperforms the standard DeepAR, achieving lower MAPE and RMSE values, indicating better prediction accuracy.

No.	Experiment	MAPE	RMSE
1	VOO_DeepAR	0.0167	10.764
2	VOO_DeepAR_NSS	0.0132	7.2718
3	VIOO_DeepAR	0.0136	1.8077
4	VIOO_DeepAR_NSS	0.0083	1.5708
5	SPY_DeepAR	0.0122	9.3612
6	SPY_DeepAR_NSS	0.0095	6.7956
7	JNK_DeepAR	0.0032	0.3355
8	JNK_DeepAR_NSS	0.0024	0.3930
9	IVOO_DeepAR	0.0091	1.3720
10	IVOO_DeepAR_NSS	0.0077	0.9976
11	GOVT_DeepAR	0.0167	0.3730
12	GOVT_DeepAR_NSS	0.0084	0.2214
13	GLD_DeepAR	0.0506	14.5228
14	GLD_DeepAR_NSS	0.0377	10.7387
15	BND_DeepAR	0.0133	0.9852
16	BND_DeepAR_NSS	0.0086	0.6413

TABLE 4: DeepAR and DeepAR-NSS Models Results.

Source: Author.

MAPE, Mean Absolute Percentage Error; NSS, Normalized Sentiment Score; RMSE, Root Mean Square Error

Figure *4* shows a visual comparison between DeepAR and DeepAR-NSS across multiple ETFs, showing that the DeepAR-NSS model consistently delivers more accurate and stable forecasts. The predicted trends from DeepAR-NSS align more closely with actual price movements and exhibit narrower confidence intervals, indicating higher certainty and less volatility.

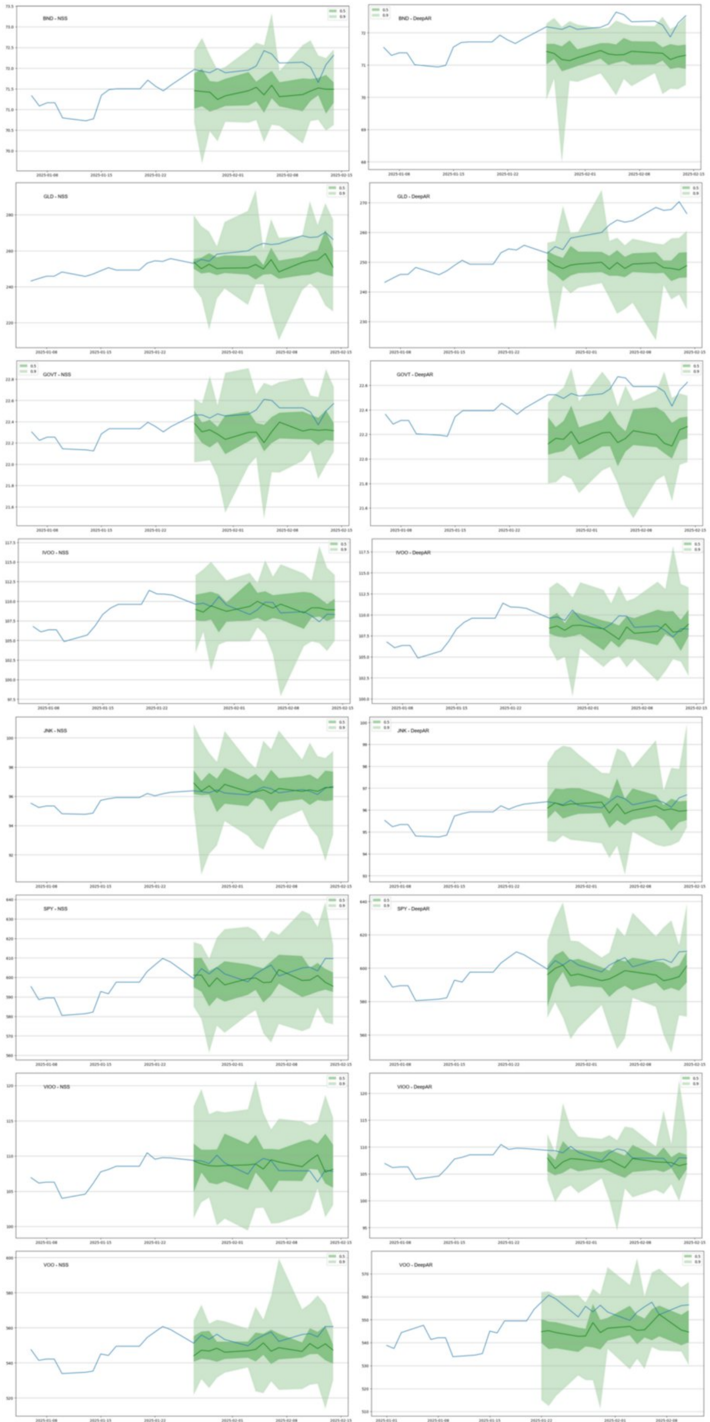

FIGURE 4: DeepAR and DeepAR-NSS Results for Different Tickers.

The X-axis in each subplot represents time, whereas the Y-axis represents the close price of the ticker.

Source: Author.

NSS, Normalized Sentiment Score

Discussion

The results align with the study's objective to enhance ETF price prediction accuracy by integrating NSS and TKES into the DeepAR model. The significant MAPE reductions (22-49.7%) across all ETFs demonstrate that incorporating news sentiment and dynamic keywords effectively captures market dynamics, addressing the research gap of limited real-time sentiment integration identified in prior studies (Tetlock, 2007) and (Bollen et al., 2011). For instance, the 49.7% MAPE reduction for GOVT suggests that bond ETFs, which are sensitive to macroeconomic shifts, benefit greatly from sentiment data, corroborating (Tetlock, 2007) findings on the predictive power of news sentiment for stable assets. Similarly, the 25.5% MAPE improvement for GLD aligns with (Orăștean et al., 2024), where public interest (via Google Trends) enhanced forecasting for volatile assets like cryptocurrencies, extending this insight to gold.

Compared to our previous work (Soliman et al., 2025), which showed DeepAR outperforming LSTM with static sentiment scores, the current DeepAR-NSS model further improves accuracy by leveraging TKES's dynamic keyword selection. This advancement addresses the limitation of ticker-specific approaches, as TKES captures broader market narratives (e.g., "AI" surges in January 2024) missed by ETF tickers alone. However, the increased RMSE for JNK (0.3355 to 0.3930) contrasts with the general trend, suggesting that sentiment integration may introduce variability for high-risk assets. This finding partially aligns with (Shapiro et al., 2022), where sentiment aggregation improved accuracy but occasionally amplified noise in complex datasets.

Differing behaviour was observed in bond ETFs. While government bond ETFs such as GOVT and BND benefit from sentiment-driven models due to their direct sensitivity to macroeconomic indicators, JNK - composed of high-yield, lower-credit-quality bonds - reacts more strongly to speculative sentiment and credit risk perceptions. These dynamics make JNK more susceptible to noise introduced by general sentiment, which may not be directly tied to its underlying fundamentals. As a result, while point accuracy (e.g., MAPE) improves, the model's uncertainty increases, as reflected in the wider prediction intervals and higher RMSE. This emphasizes the need for asset-specific sentiment modelling, particularly for riskier bond instruments.

Limitations include the potential overfitting of TKES to trending topics, which may not always correlate with ETF prices, and the reliance on Gnews API, which may miss niche financial news sources. The anomaly with JNK highlights a need for asset-specific sentiment weighting, as high-yield bonds may react differently to sentiment than equities or government bonds. Practically, these findings imply that financial analysts could leverage DeepAR-NSS for more accurate ETF forecasts, particularly for low-to-moderate risk assets, enhancing portfolio management. This study advances sentiment-based forecasting by demonstrating the value of hybrid keyword approaches.

Conclusions

This research demonstrates the significant potential of augmenting the DeepAR model with NSS and the TKES framework to enhance ETF price predictions. By leveraging a hybrid approach that integrates trending keywords from Google Trends with expertly curated macroeconomic terms, the TKES methodology effectively captures a broader and more dynamic sentiment landscape, addressing limitations of ticker-specific approaches. The NSS further refines this process by consolidating overlapping and conflicting news sentiments into a robust daily signal, thereby improving the interpretability and predictive power of the DeepAR-NSS model. Empirical results across eight diverse U.S. ETFs reveal consistent improvements in forecasting accuracy, with MAPE reductions ranging from 22% to 49.7% and RMSE decreases of up to 40.6%, particularly for bond ETFs (e.g., GOVT, BND) and volatile assets (e.g., GLD). However, the increased RMSE observed for the high-yield bond ETF JNK suggests that sentiment integration, while enhancing point prediction accuracy, may introduce variability in prediction ranges for high-risk assets, necessitating further exploration. These findings affirm the value of incorporating unstructured news data into time-series forecasting and highlight the importance of tailoring sentiment-based enhancements to the unique risk profiles and asset characteristics of ETFs. Future research could investigate the optimization of sentiment weighting schemes and the application of these techniques to other financial instruments, further advancing the precision of predictive models in dynamic market environments.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Waleed Soliman, Zhiyuan Chen, Colin Johnson, Sabrina Wong

Acquisition, analysis, or interpretation of data: Waleed Soliman, Zhiyuan Chen

Drafting of the manuscript: Waleed Soliman

Critical review of the manuscript for important intellectual content: Waleed Soliman, Zhiyuan Chen, Colin Johnson, Sabrina Wong

Supervision: Zhiyuan Chen, Colin Johnson

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** This article was made possible through the support of a collaborative research program between Hy-Alpha Sdn Bhd and the University of Nottingham Malaysia Campus (Project NVHT0006). **Financial relationships:** Sabrina Wong declare(s) a grant from Hy-Alpha Sdn Bhd. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The data of this study were collected from different websites, notably those of the Yahoo Finance "https://finance.yahoo.com/", FRED "https://fred.stlouisfed.org/", and Gnews "https://gnews.io/". The datasets are available.

References

1. Araci D: FinBERT: Financial sentiment analysis with pre-trained language models. arXiv:1908.10063. 2019, 10.48550/arXiv.1908.10063
2. Bollen J, Mao H, Zeng X: Twitter mood predicts the stock market. Journal of Computational Science. 2011, 2:1-8. 10.1016/j.jocs.2010.12.007
3. Chen NF, Roll R, Ross SA: Economic forces and the stock market. The Journal of Business. 1986, 59:383-403. 10.1086/296344
4. Fama EF, French KR: Common risk factors in the returns on stocks and bonds. Journal of Financial Economics. 1993, 33:3-56. 10.1016/0304-405x(93)90023-5
5. Federal Reserve Bank of St. Louis. (2025). https://fred.stlouisfed.org/.
6. Gnews. (2025). https://gnews.io/.
7. Huang MY, Rojas RR, Convery PD: Forecasting stock market movements using Google Trend searches. Empirical Economics. 2020, 59:2821-2839. 10.1007/s00181-019-01725-1
8. Lettau M, Madhavan A: Exchange-traded funds 101 for economists. The Journal of Economic Perspectives. 2018, 32:135-154. 10.1257/jep.32.1.135
9. Orăştean R, Mărginean SC, Sava R: Exploring the relationship between Google Trends and cryptocurrency metrics. Studies in Business and Economics. 2024, 19:368-379. 10.2478/sbe-2024-0020
10. Poon SH, Granger CWJ: Forecasting volatility in financial markets: A review. Journal of Economic Literature. 2003, 41:478-539. 10.1257/002205103765762743
11. RMSE. (2022). Accessed: January 15, 2023: https://c3.ai/glossary/data-science/root-mean-square-error-rmse/.
12. Salinas D, Flunkert V, Gasthaus J: DeepAR: Probabilistic forecasting with autoregressive recurrent networks. arXiv:1704.04110. 2017, 10.48550/arXiv.1704.04110
13. Scikit-learn. (2022). Accessed: January 10, 2023: https://scikit-learn.org/stable/modules/model_evaluation.html#mean-absolute-percentage-error.
14. Shapiro AH, Sudhof M, Wilson DJ: Measuring news sentiment. Journal of Econometrics. 2022, 228:221-243. 10.1016/j.jeconom.2020.07.053
15. Soliman WM, Chen Z, Johnson C, Wong S: ETF markets' prediction & assets management platform using probabilistic autoregressive recurrent networks. The Eurasia Proceedings of Science Technology Engineering and Mathematics. 2023, 23:485-494. 10.55549/epstem.1372067
16. Soliman W, Chen Z, Johnson C, Wong S: Improving ETF prediction through sentiment analysis: A DeepAR and FinBERT approach with controlled seed sampling. Intelligent Systems in Accounting Finance and Management. 2025, 32:e70004. 10.1002/isaf.70004
17. Tetlock PC: Giving content to investor sentiment: the role of media in the stock market. The Journal of Finance. 2007, 62:1139-1168. 10.1111/j.1540-6261.2007.01232.x
18. Yahoo Finance. (2025). https://finance.yahoo.com/.