Received: Jun. 21, 2024
Revised: Jul. 21, 2024
Accepted: Jul. 23, 2024

Journal of Management Science
Dhonburi Rajabhat University
วารสารวิทยาการจัดการ
มหาวิทยาลัยราชภัฏธนบุรี

Intraday Multi-timeframe Prediction of the Thailand Stock Market Index Futures

Ekkalak Sukumalpongkul[1]
Thapana Boonchoo[2]

Abstract

This quantitative research study investigates the intraday multi-timeframe prediction of SET50 Index Futures prices in the Thailand stock market, employing advanced machine learning techniques within the context of management science. The research objectives are to enhance predictive accuracy and improve strategic decision-making for futures trading by integrating technical indicators across multiple timeframes. The study focuses on the SET50 Index Futures, initiated by the Thailand Futures Exchange (TFEX) in 2006, recognizing these contracts as critical tools for price discovery and risk management. The research scope encompasses the application and comparison of two machine learning models: eXtreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM). These models are utilized to analyze technical indicators derived from multiple intraday timeframes of the SET50 Index Futures data. The study employs a comprehensive set of technical indicators and conducts extensive experiments to evaluate the models' effectiveness in various timeframe configurations. The research results reveal that XGBoost consistently outperforms LSTM, particularly in multi-timeframe configurations. This finding underscores the importance of multi-timeframe analysis in effective risk management and strategic planning for futures trading. The superior performance of XGBoost in processing complex, multi-dimensional data offers valuable insights for managers and researchers in optimizing futures trading strategies and improving operational efficiency in the context of the Thailand stock market.

Keywords: Intraday Multi-timeframe Prediction, Thailand Stock Market, Index Futures, Machine Learning, XGBoost, LSTM

[1] Master Student, Faculty of Science and Technology, Department of Computer Science, Thammasat University
e-mail: ekkalak.suk@dome.tu.ac.th

[2] Assistant Professor Dr. Faculty of Science and Technology, Department of Computer Science, Thammasat University
e-mail: thapanab@tu.ac.th

Introduction

The Thailand Futures Exchange (TFEX), established in 2004, initiated trading of the SET50 Index Futures in 2006. Futures contracts, agreements between buyers and sellers for future asset delivery at a predetermined price, are traded on futures markets and settled daily, with losses and gains transferred between parties. Leverage, a key feature of futures markets, allows investors to control substantial contract values with minimal capital, potentially amplifying returns (Gottesman, 2016). However, this leverage also escalates risks, underscoring the need for accurate market forecasting and analysis to enhance risk management strategies (Chen & Guestrin, 2016).

The significance of futures markets extends beyond financial gains. In the realm of management science, futures markets serve as a critical tool for price discovery and risk hedging. By providing a platform for trading contracts based on future expectations, these markets contribute to efficient resource allocation and informed decision-making in various industries. For instance, agricultural futures enable farmers and producers to manage price risks associated with their crops, while commodity futures aid manufacturers in securing stable input prices. The insights derived from futures market analysis can inform strategic management decisions, production planning, and inventory management, ultimately enhancing operational efficiency and profitability (Peterson, 2018).

Previous research has explored diverse techniques for analyzing and predicting futures market movements, encompassing statistical models and advanced deep learning networks (Gandhmal & Kumar, 2019). These studies have collectively contributed to the evolving field of financial market prediction, offering valuable insights and predictive capabilities. Decision tree-based models like XGBoost are renowned for their effectiveness in pattern recognition and classification tasks (Chen & Guestrin, 2016). In the realm of deep learning, techniques like Long Short-Term Memory (LSTM) networks are highlighted for their ability to capture temporal dependencies in time-series data, a crucial factor in price analysis (Yu et al., 2019).

Furthermore, predicting stock returns from historical data, especially over longer time horizons like daily, weekly, and monthly periods, presents a significant challenge (Reboredo et al., 2012). The integration of multi-timeframe strategies, combining short-term data for optimizing entry points with longer-term trend analysis, has gained prominence. This approach recognizes that trend identification is often more accurate over extended periods, while exploiting specific entry points necessitates swift action. Consequently, a strategic blend of multiple timeframes, each tailored to specific objectives, is deemed beneficial. Accurate trend identification empowers traders to identify short-term trading opportunities with a higher probability of profitability (Kaufman, 2019). Moreover, high-frequency trading may offer investors a wider array of opportunities compared to lower-frequency alternatives, potentially mitigating the risks associated with long-term investments (Savani, 2012; Loveless et al., 2013).

In this research, we propose a forecasting framework for SET50 Index Futures prices based on the SET50 Index in the Stock Exchange of Thailand (SET). We investigate the combination of features from multiple timeframes extracted from the collected data using various technical indicators, aiming to capture the dynamic behavior of the market. We focus on two machine learning models for incorporation into the system: eXtreme Gradient Boosting (XGBoost) (Nobre & Neves, 2019; Kumar et al., 2022) and Long Short-Term Memory (LSTM) (Liu, Liao & Ding, 2018; Zou & Qu, 2020), chosen for their demonstrated superiority in past research. Finally, we conducted extensive experiments to examine the performance of feature combinations in each individual model and their comparative effectiveness.

Research Objectives

1. To investigate the comparative effectiveness of utilizing multiple timeframes in predicting SET50 Index Futures prices.

2. To explore the integration of technical indicators derived from various intraday timeframes for enhancing the accuracy of SET50 Index Futures price prediction models.

3. To identify optimal strategies for combining multiple timeframes to improve the efficacy of machine learning models (XGBoost and LSTM) in forecasting SET50 Index Futures prices.

Literature Review

The quest to understand and predict stock market trends has spurred the evolution of analytical methodologies, transitioning from traditional statistical models to cutting-edge machine learning and deep learning techniques.

Futures and Their Role in Trading Strategies

Futures contracts, particularly index futures like the SET50 Index Futures (S50IF) in Thailand, have become pivotal instruments in trading strategies due to their unique characteristics and advantages.

Primarily, futures facilitate hedging, empowering investors and fund managers to mitigate price volatility risks by securing future prices. This mechanism acts as insurance against adverse market movements, safeguarding portfolio values. Furthermore, futures enable speculation, wherein traders capitalize on anticipated price fluctuations by predicting market trends. By accurately forecasting movements, traders can generate substantial profits through strategic buying or selling of futures contracts. Another compelling feature of futures is leverage. This financial instrument requires a comparatively smaller initial investment than acquiring the underlying assets directly, potentially amplifying returns. However, it is imperative to acknowledge that leverage also magnifies potential losses, necessitating prudent risk management. Moreover, futures broaden market accessibility. Investors can engage with diverse and high-value markets, transcending geographical boundaries and accessing international indices. This accessibility enhances portfolio diversification and expands investment opportunities (Chance & Brooks, 2009).

For hedging purposes, Cao et al. explore using reinforcement learning (RL) to optimize futures-based hedging strategies by accounting for transaction costs. Traditional delta hedging, which involves frequent adjustments to neutralize risk, incurs significant trading costs, thereby reducing profitability. The RL approach dynamically adjusts hedging positions, sometimes under-hedging or over-hedging, to minimize these costs (Cao et al., 2021).

In using futures as instruments for speculation with greater access to diverse markets, numerous studies primarily focus on predicting the direction of trends or futures prices, especially for commodities such as steel, silver, gold, oil, and bitcoin, as well as agricultural products like soybean, coffee, and corn (Wang & Gao, 2018; Chen et al., 2018; Gong & Lin, 2021; Chatziantoniou et al., 2019; Czudaj, 2019; Liu et al., 2023).

Advancements in Machine Learning Techniques

The eXtreme Gradient Boosting (XGBoost) algorithm has been recognized for its superior performance in various predictive modeling tasks. Based on decision trees, specifically Classification and Regression Trees (CART), XGBoost demonstrates enhanced accuracy compared to traditional statistical approaches. Research by Chen and Guestrin highlights XGBoost's ability to effectively model complex non-linear patterns within data. In addition to XGBoost's general advantages, Yun et al. propose a hybrid GA-XGBoost prediction system incorporating feature set expansion, data preparation, and feature selection. Their findings demonstrate that expanding the feature set with 67 technical indicators significantly improved prediction accuracy, surpassing benchmark models. Importantly, they achieved a parsimonious feature set using GA-XGBoost, leading to comparable performance with fewer features. This underscores the critical interplay between feature engineering and model selection in financial prediction, highlighting the need to manage the trade-offs between dimensionality and predictive power (Yun et al., 2021).

The Emergence of Deep Learning

Deep learning, specifically through architectures like Long Short-Term Memory (LSTM) networks, has revolutionized stock price prediction. LSTM models outperform traditional approaches in handling volatile and unpredictable financial time series data (Sanboon et al., 2019). Moghar et al. leverage an LSTM-based Recurrent Neural Network (RNN) to forecast future opening prices of GOOGL and NKE assets. While initial results show promise, they will focus on optimizing data length and training epochs for these assets (Moghar & Hamiche, 2020). Ding and Qin introduce their "Associated Net," an LSTM-based architecture designed for simultaneous multi-price prediction. This model outperforms standard and deep recurrent LSTM benchmarks while maintaining over 95% individual price point accuracy. Overall, these studies point to the importance of model customization, whether through architecture design, hyperparameter tuning, or tailored optimization for greater accuracy in financial forecasting (Deng et al., 2023).

Multi-Timeframe Strategy in Financial Markets

Multi-timescale feature integration has emerged as a promising approach in predictive modeling. By capturing insights across different temporal granularities, from short-term fluctuations to long-term trends, this approach has the potential to enhance predictive accuracy. Deng et al. apply such a strategy to high-frequency apple futures trading, utilizing XGBoost models with features derived from five, ten, and twenty-period timescales. Particle Swarm Optimization (PSO) refines trading decisions within a rule-based framework. This multi-timescale approach outperforms benchmarks in direction prediction accuracy, transaction returns, and return/risk ratio (Deng et al., 2020). Similarly, Hao and Gao focus on multi-timescale features in stock index data. Their innovative hybrid neural network employs a single CNN for simultaneous short-term, medium-term, and long-term feature extraction. These features are then modeled by separate LSTMs to capture timescale-specific temporal dependencies. This work demonstrates significant performance gains over single-scale methods (SVM, CNN, LSTM) and more complex multi-network models (Hao & Gao, 2020). Guang et al. also targets the limitations of single-scale stock trend prediction (STP). Their Multi-scale Two-way Deep Neural Network (MTDNN) combines wavelet-based (XGBoost) and down sampling-based (RCNN) techniques for diverse timescale insight extraction. MTDNN achieves state-of-the-art performance on FI-2010 and underscores the value of multi-timescale information through validation on CSI-2016 dataset (Liu et al., 2020). Deng et al.'s study on high-frequency forecasting of crude oil futures prices using multiple timeframe predictions and trading simulation represents a significant contribution to the literature on financial prediction and trading strategy optimization. The proposed MTXGBoost-NSGA-II-SHAP model effectively combines Multiple Timeframe eXtreme Gradient Boosting (MTXGBoost), Non-dominated Sorting Genetic Algorithm-II (NSGA-II), and Shapley Additive exPlanation (SHAP) to enhance prediction accuracy and trading profitability while managing risk. This research addresses gaps in existing literature by integrating predictions across multiple timeframes and optimizing trading strategies through multi-objective algorithms. The empirical results demonstrate superior performance in terms of direction prediction accuracy, accumulated return, and maximum drawdown, thereby underscoring the model's potential as a robust decision support system for market participants. This study's innovative approach provides a valuable framework for future research in high-frequency trading and financial market prediction, suggesting broader applications across various financial instruments and markets (Deng et al., 2023).

The literature review reveals a significant shift towards advanced machine learning techniques in financial market prediction, with XGBoost and LSTM emerging as powerful tools for capturing complex market dynamics. Multi-timeframe analysis has shown promise in enhancing prediction accuracy and trading performance across various financial instruments, allowing for the capture of both short-term fluctuations and long-term trends. The importance of feature engineering and selection is evident, with studies demonstrating improved performance through the use of technical indicators and optimal feature sets. There is also a growing trend towards developing hybrid models that combine multiple techniques. However, a notable gap exists in the application of these sophisticated approaches specifically to the SET50 Index Futures market in Thailand. While the effectiveness of multi-timeframe analysis and hybrid models has been demonstrated in various markets, their performance in emerging markets like Thailand remains largely unexplored. This presents an opportunity for the current study to contribute valuable insights into the effectiveness of multi-timeframe, hybrid machine learning approaches in predicting SET50 Index Futures prices, potentially enhancing trading strategies and risk management in this specific market context.

Methodology

The research methodology presented in the diagram aims to forecast intraday trends in the SET50 Index Futures (S50IF) by utilizing a multi-timeframe approach and various machine learning models. The methodology is divided into three main phases: data preprocessing and feature engineering, model training, and model evaluation.

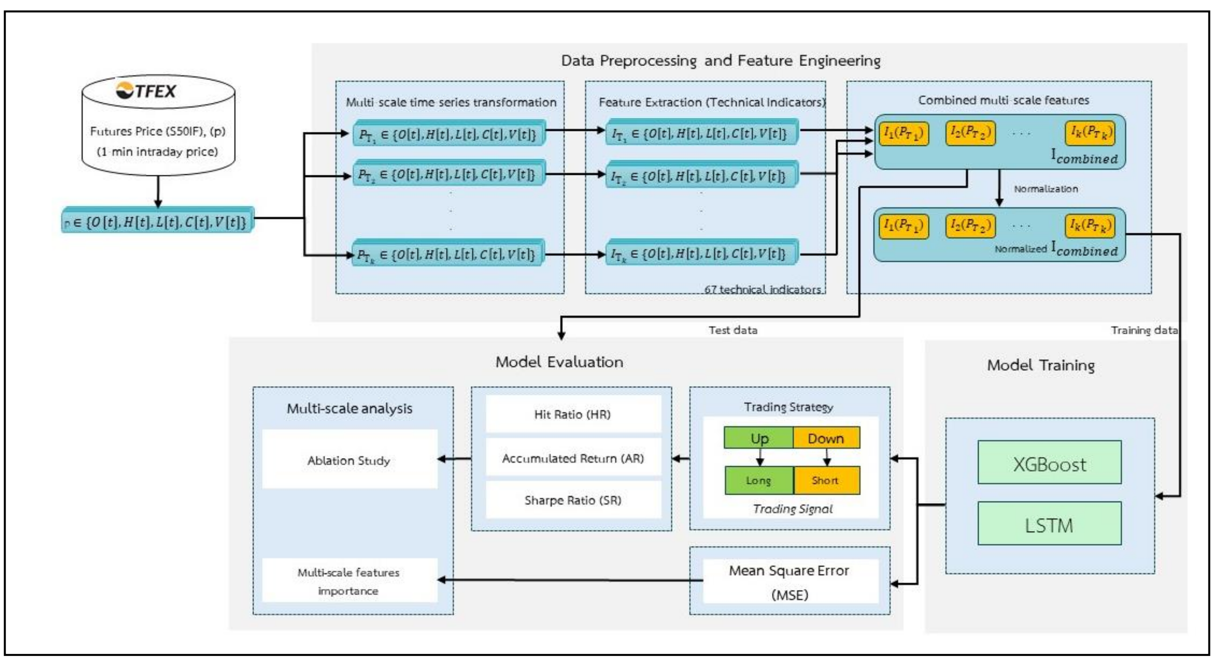

Figure 1 The overall framework

Data Preprocessing and Feature Engineering

The first phase involves transforming the raw 1-minute intraday price data of S50IF, denoted as $P_T = \{O[t], H[t], L[t], C[t], V[t]\}$ into multiple time frames (e.g., 15-minute, 30-minute, 1-hour) through resampling. This results in a set of multi-scale time series representations:

$$P_{\mathrm{T}_k} = \{O[t], H[t], L[t], C[t], V[t]\}$$

where (k) represents the time frame (e.g., 15 minutes, 30 minutes, etc.).

Next, a comprehensive set of 67 technical indicators, as shown in Table 1, is calculated for each time frame (k), denoted as (I_{T_k}). These indicators are categorized into different groups, such as overlap studies, momentum indicators, volatility indicators, and volume indicators. These indicators are categorized into different groups, such as overlap studies, momentum indicators, volatility indicators, and volume indicators.

The extracted features from all time frames are then combined into a single feature vector:

$$\mathrm{I}_{combined} = [I_1(P_{\mathrm{T}_1}), I_2(P_{\mathrm{T}_2}), \ldots, I_j(P_{\mathrm{T}_k})]$$

where (k) is the total number of timeframes considered. To ensure that all features are on a similar scale, the combined feature vector is normalized:

$$\mathrm{I}_{normalized} = Normalize(I_{combined})$$

Finally, the normalized data is split into training and test sets to prepare for model training and evaluation.

Table 1 List of groups of technical indicators used in the research

Category	Technical Indicators	Description
Category 1: Overlap Studies Indicators	1BBUPPER	Bollinger Bands Upper
	1BBMIDDLE	Bollinger Bands Middle
	1BBLOWER	Bollinger Bands Lower
	1DEMA	Double Exponential Moving Average
	1MIDPOINT	Midpoint over period
	1MIDPRICE	Midpoint Price over period
	1SMA	Simple Moving Average
	1T3	Triple Exponential Moving Average (T3)
	1TEMA	Triple Exponential Moving Average
	1TRIMA	Triangular Moving Average
	1WMA	Weighted Moving Average

Table 1 (cont.)

Category	Technical Indicators	Description
Category 2: Momentum Indicators	2ADX	Average Directional Movement Index
	2ADXR	Average Directional Movement Index Rating
	2APO	Absolute Price Oscillator
	2AROONDOWN	Aroon Down
	2AROONUP	Aroon Up
	2AROONOSC	Aroon Oscillator
	2BOP	Balance of Power
	2CCI	Commodity Channel Index
	2CMO	Chande Momentum Oscillator
	2DX	Directional Movement Index
	2MACD	Moving Average Convergence/Divergence
	2MACDSIG	Moving Average Convergence/Divergence
	2MACDHIST	Moving Average Convergence/Divergence
	2MACDEXT	MACD with controllable MA type
	2MACDEXTSIG	MACD with controllable MA type
	2MACDEXTHIST	MACD with controllable MA type
	2MACDFIX	Moving Average Convergence/Divergence Fix 12/26
	2MACDFIXSIG	Moving Average Convergence/Divergence Fix 12/26
	2MACDFIXHIST	Moving Average Convergence/Divergence Fix 12/26
	2MFI	Money Flow Index
	2MINUS_DI	Minus Directional Indicator
	2MINUS_DM	Minus Directional Movement
	2MOM	Momentum
	2PLUS_DI	Plus Directional Indicator
	2PLUS_DM	Plus Directional Movement
	2PPO	Percentage Price Oscillator
	2ROC	Rate of change: ((price/prevPrice)-1)*100
	2ROCP	Rate of change Percentage: (price-prevPrice)/prevPrice
	2ROCR	Rate of change ratio: (price/prevPrice)
	2ROCR100	Rate of change ratio 100 scale: (price/prevPrice)*100
	2RSI	Relative Strength Index
	2SLOWK	Stochastic %K High
	2SLOWD	Stochastic %K Low
	2STOCHFH	Stochastic Fast High
	2STOCHFL	Stochastic Fast Low

Table 1 (cont.)

Category	Technical Indicators	Description
	2STOK	Stochastic Relative Strength Index High
	2STOD	Stochastic Relative Strength Index Low
	2TRIX	1-day Rate-Of-Change (ROC) of a Triple Smooth EMA
	2ULTOSC	Ultimate Oscillator
	2WILLR	Williams' %R
Category 3: Volatility Indicators	3ATR	Average True Range
	3NATR	Normalized Average True Range
	3TRANGE	True Range
Category 4: Volume Indicators	4AD	Chaikin A/D Line
	4DOSC	Chaikin A/D Oscillator
Category 5: Price Transform	5AVGPRICE	Average Price
	5MEDPRICE	Median Price
	5TYPPRICE	Typical Price
	5WCLPRICE	Weighted Close Price
Category 6: Cycle Indicators	6HT_DCPERIOD	Hilbert Transform - Dominant Cycle Period
	6HT_DCPHASE	Hilbert Transform - Dominant Cycle Phase
	6INPHASE	Hilbert Transform - Phasor Components Inpahse
	6QUADRATURE	Hilbert Transform - Phasor Components Quadrature
	6SINE	Hilbert Transform - SineWave
	6LEADSINE	Hilbert Transform - LeadSineWave
	6HT_TRENDMODE	Hilbert Transform - Trend vs Cycle Mode

Model Training

Two different machine learning models, XGBoost (eXtreme Gradient Boosting) and LSTM (Long Short-Term Memory), are trained using the training data. These models represent diverse approaches to time series forecasting, capturing different aspects of the data.

XGBoost (eXtreme Gradient Boosting) is a powerful machine learning algorithm for regression and classification, distinguished by its objective function. This function has two key parts: a differentiable loss function, $L(y_i, \hat{y}_i)$, measuring the difference between predicted values $(\hat{y}_i)$ and true values (y_i), and a regularization term, $\Omega(f_k)$, penalizing model complexity to prevent overfitting. Its regularization term is particularly notable:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

This design highlights XGBoost's core principle: it seeks to balance predictive accuracy on training data with model simplicity, enhancing generalization and avoiding overfitting. The algorithm's flexibility in handling missing values and diverse loss functions further contributes to its adaptability across various predictive problems.

Long Short-Term Memory (LSTM) networks mitigate the vanishing gradient problem of RNNs through their cell state and three gates: input, forget, and output. Key equations governing LSTM behavior include:

Forget Gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f$

Input gate layer: $i_t = \sigma(W_\mathrm{i} \cdot [h_{t-1}, x_t] + b_i)$, $\widetilde{C_t} = tanh(W_c \cdot [h_{t-1}, x_t] + b_c$

Output gate: $O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

Where W and b represent weights and biases, respectively.

LSTM training typically uses backpropagation through time and gradient descent-based optimization. Its ability to retain and selectively modify information over long sequences makes it powerful for time-series prediction and other complex sequence problems.

Model Evaluation

Trading Strategy: Market trend determination is achieved through a comparative analysis of previous and current predicted values. An upward trend ("Up") is identified when the current predicted value ($\boldsymbol{Predicted_t}$) exceeds the previous predicted value ($\boldsymbol{Predicted_{t-1}}$). Conversely, a downward trend ("Down") is established when the opposite condition holds. Formally, this is expressed as:

$$Trend_i = \begin{cases} Up & if\ Predicted_t > Predicted_{t-1} \\ Down & if\ Predicted_t < Predicted_{t-1} \end{cases}$$

Subsequently, Trading signals are generated based on the identified trend. A shift to an upward trend ("Up") triggers a "Long" signal, indicating an anticipated price increase. The maintenance of a downward trend ("Down") or transition to a downward trend extracts a "Short" signal, signifying an expected price decline. This decision process is formalized as:

$$Trading\ Signal_t = \begin{cases} Long & if\ Trend_t = Up \\ Short & otherwise \end{cases}$$

Hit ratio (HR): The Hit Ratio (HR) is a metric commonly used in financial markets for evaluating the accuracy of a predictor's direction forecast. The hit ratio is calculated using the formula given by:

$$HR = \frac{CU + CD}{N}$$

CU denotes the number of correct "Up" trend predictions, while CD signifies the number of correct "Down" trend predictions. These values, CU and CD, reflect the frequency of accurate forecasts made by the model for long and short-selling transactions, respectively. Finally, N represents the total number of transactions executed based on the model's predicted market direction.

Accumulated Return (AR): Accumulated return offers a measure of profitability across a specified timeframe (e.g., quarterly or within 15-minute intervals). It represents the net profit or loss over these defined periods, calculated as:

$$AR = \sum_{i=1}^{n} (P_i - C_i)$$

where: P_i represents the profit for the i-th period, C_i represents the transaction cost for the i-th period (set to 0\% in this simulation). Accumulated return provides a holistic view of strategy effectiveness over time.

Sharpe Ratio (SR): The Sharpe ratio is a risk-adjusted performance metric used to measure the efficiency of an investment by comparing the return of an investment with its risk. It's a mathematical expression of the insight that excess returns over a specified duration may signify more volatility and risk, rather than investing skill. It's calculated by subtracting the risk-free rate of return from the average return of the investment and then dividing the result by the investment's standard deviation, which represents the investment's volatility or risk. The Sharpe Ratio (SR) is calculated as follows:

$$SR = \frac{E[AR - R_f]}{\sigma}$$

Where: $E[AR - R_f]$ is the expected value of the excess returns on the asset, which means the average return of the asset above the risk-free rate (R_f). σ (sigma) denotes the standard deviation of the excess returns, which measures the volatility of the excess returns.

Experimental results

We employ a historical intraday data of SET50 Index Futures (S50IF) prices, collected from Thailand Futures Exchange (TFEX). The dataset comprises one-minute interval price data and spans a substantial period from April 28, 2006, to June 17, 2023. This reflects the fine-grained price movements of a portfolio comprised 50 highly liquid, large-capitalization securities listed on the Stock Exchange of Thailand (SET). The total number of observations is 1,288,776 points in which we allocated 70% of data for model training, 15% for validation, and the rest 15% for testing. Utilizing a combination of multiple time frames allows technical analysis instruments to gain a more comprehensive understanding of market behavior concerning historical data. As long-term charts often provide clearer trends in price movements, shorter time frames enable timely adjustments to capitalize on favorable entry points. Filtering short-term trades based on these longer trends can increase the likelihood of profitable outcomes.

Ablation Study Results: The Impact of Timeframes on Model Performance

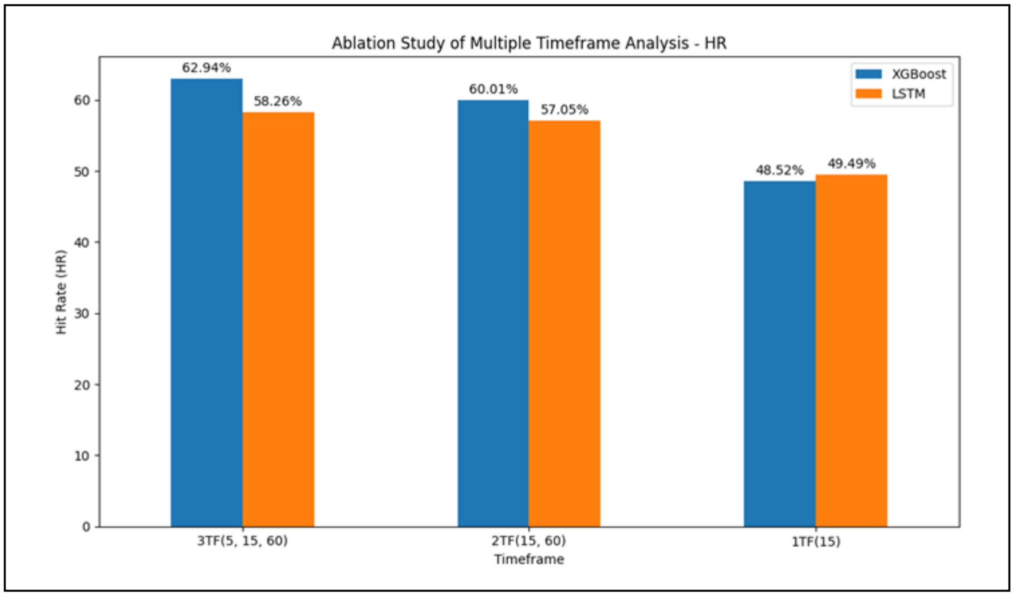

Figure 2 Hit Ratio (HR)

The Hit Ratio (HR) analysis demonstrates the significant impact of multi-timeframe approaches on predictive model accuracy. Utilizing three timeframes: 3TF(5, 15, 60) consistently resulted in the highest HR percentages across all models. The XGBoost model achieved an HR of 62.94% with the 3TF configuration. Reducing the number of timeframes led to a noticeable decrease in HR. This highlights the importance of incorporating multiple timeframes for enhanced forecasting accuracy within financial markets, as shown in Figure 2.

Accumulated Return (AR) Analysis

Figure 3 highlights the impact of multi-timeframe analysis on trading strategy profitability, as measured by Accumulated Return (AR). The XGboost model achieved the highest AR (132.95%) with the 2TF configuration, outperforming other models. Reducing the number of timeframes led to lower profitability across all models. This is particularly evident in the XGboost model, where the AR dropped substantially to 1.00% in the 1TF scenario. These findings underscore the importance of integrating multiple timeframes for maximizing financial returns using forecasting models.

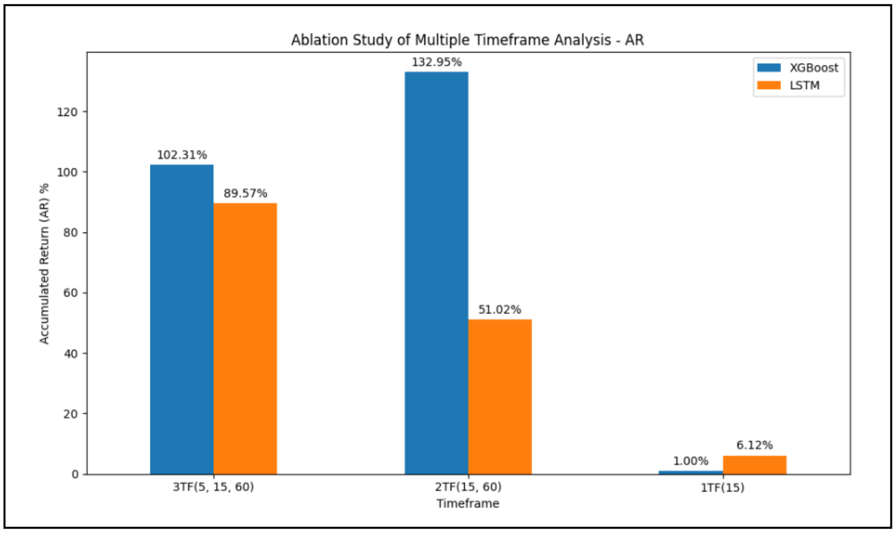

Figure 3 Accumulated return

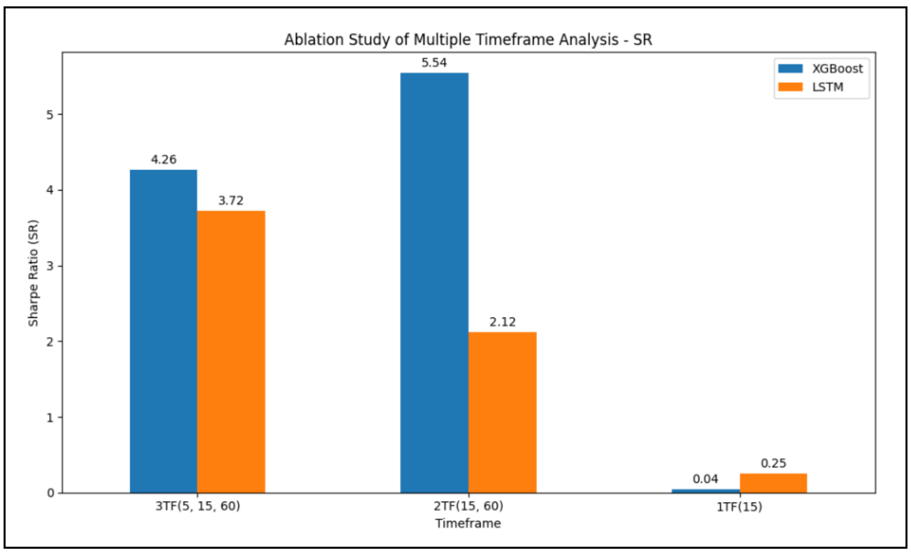

Figure 4 Sharpe Ratio (SR)

Sharpe Ratio (SR) Analysis, employed as a risk-adjusted performance metric, revealed that XGBoost generally outperformed LSTM across all assessed timeframe combinations (Figure 4). The highest Sharpe Ratio of 5.54 was observed for XGBoost in the two-timeframe (2TF) (15, 60)-minute combination, indicating superior risk-adjusted returns. Notably, a significant performance decline was evident for both models when utilizing only a single 15-minute timeframe, underscoring the importance of multiple timeframe analysis in enhancing model performance. This peak performance by XGBoost in the 2TF setting could be attributed to its resilience against market volatility and consistent return generation over time. As timeframes condensed, Sharpe Ratios generally decreased, although LSTM and SARIMAX models maintained relatively higher ratios compared to others, suggesting their effectiveness in risk management across diverse timeframes.

Feature Importance Analysis

To gain insights into the factors driving model predictions, it's essential to assess the relative importance of individual features.

Permutation Feature Importance (PFI) provides a computationally efficient and intuitive method for assessing the influence of individual features on model performance. By measuring the change in model error after randomizing a feature's values, PFI allows researchers to identify the features that bolster a model's decision-making. While initially introduced for random forests, PFI has evolved into a model-agnostic approach. This wider applicability, along with its conceptual simplicity, make PFI a valuable tool within the broader field of feature importance analysis (Breiman, 2001), (Fisher et al., 2019). To mathematically express the process, Mean Squared Error (MSE) was employed as the performance metric for consistency. Initially, a baseline MSE was calculated using the original dataset:

$$MSE_{baseline} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where N is the number of observations, y_i is the actual value of the i-th observation, and $\hat{y}_i$ is the predicted value of the i-th observation by the model.

Permutation Importance for Feature j:

Permute the values of feature j in the validation/test set, denoted as X_j', where the permuted dataset is X' with all features the same as X except for X_j'. Calculate the MSE with the permuted feature j:

$$MSE_{permuted_j} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}'_{i,j}\right)^2$$

Where $\hat{y}'_{i,j}$ is the predicted value of the i-th observation by the model using the permuted dataset X'. The importance of feature j is then the change in MSE due to permutation:

$$Importance_j = MSE_{baseline} - MSE_{permuted_j}$$

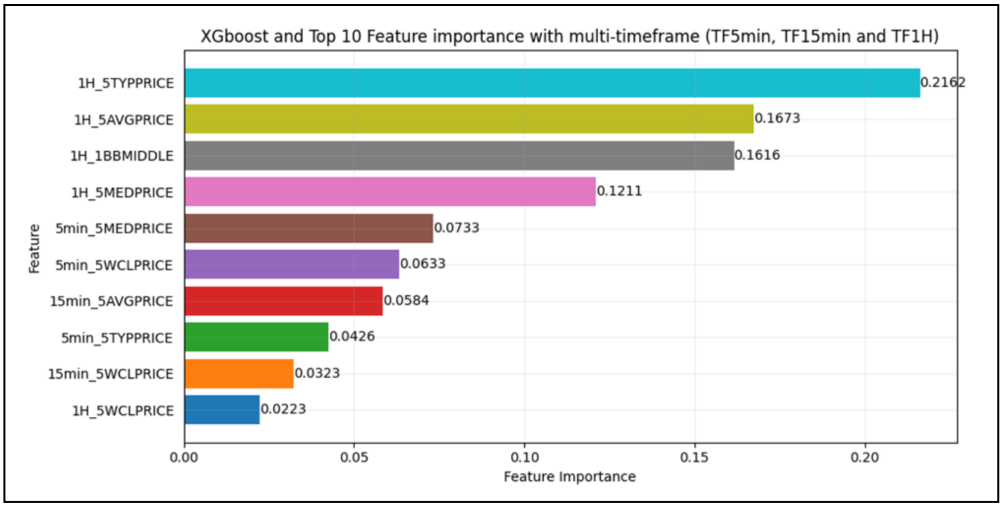

Figure 5 XGboost and Top 10 feature importance (5min, 15min, 1H)

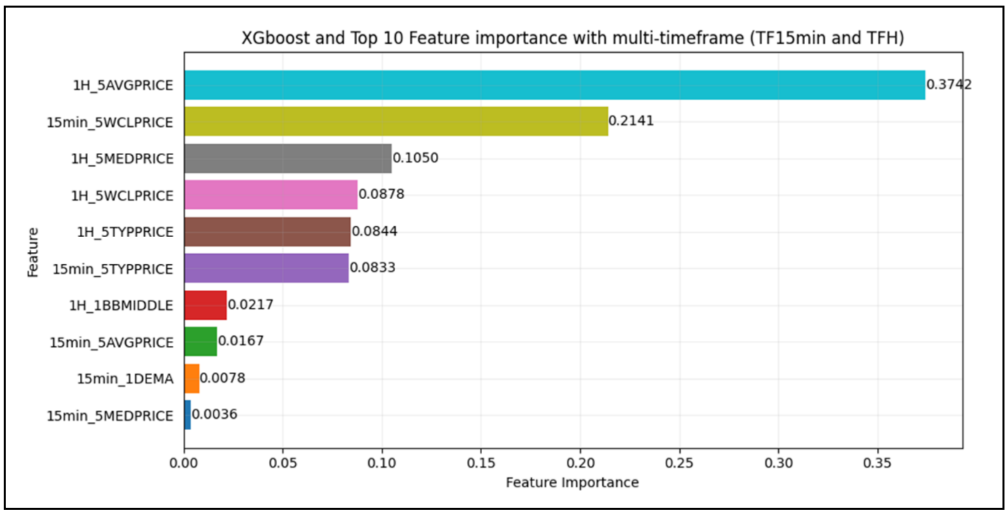

Figure 6 XGboost and Top 10 feature importance (15min, 1H)

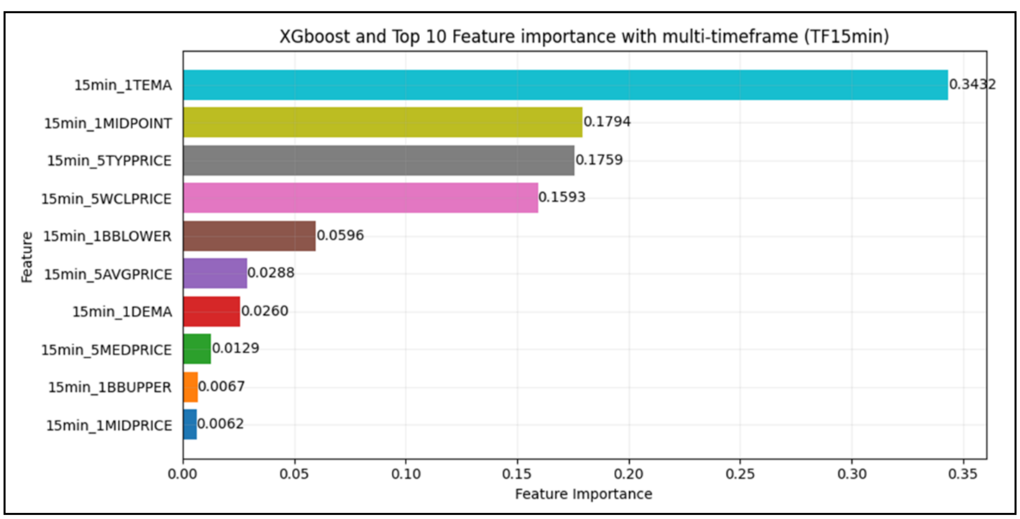

Figure 7 XGboost and Top 10 feature importance (15min)

The results depicted in Figures 5-7 illuminate the key features influencing the XGBoost model's prediction accuracy. Two categories emerge as particularly significant: Category 5 (Price Transform) and Category 1 (Overlap Studies Indicators). Within Category 5, features such as Average Price and Typical Price across various timeframes (including two and three eras) hold substantial weight in the model's predictions. However, the most influential factor appears to be the Triple Exponential Moving Average from Category 1. This feature significantly impacts the model's performance, especially for 15-minute timeframes. When combined with trading strategy results, the XGBoost model achieves its best performance using two or three timeframes (specifically 15-minute/1-hour and 5-minute/15-minute/1-hour combinations). Notably, elements from Category 5 and Category 1 remain consistently relevant across all timeframes analyzed.

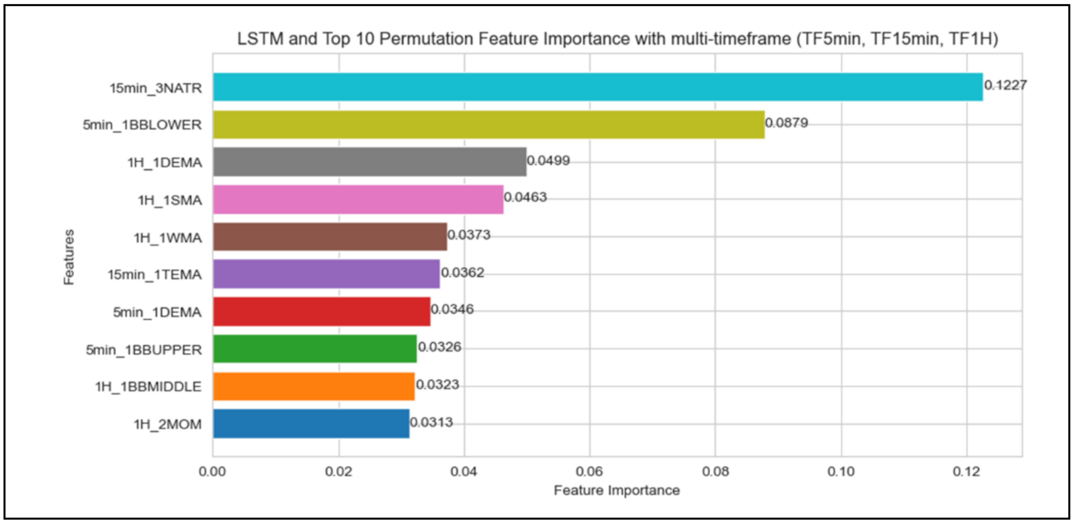

Figure 8 LSTM and Top 10 feature importance (5min, 15min, 1H)

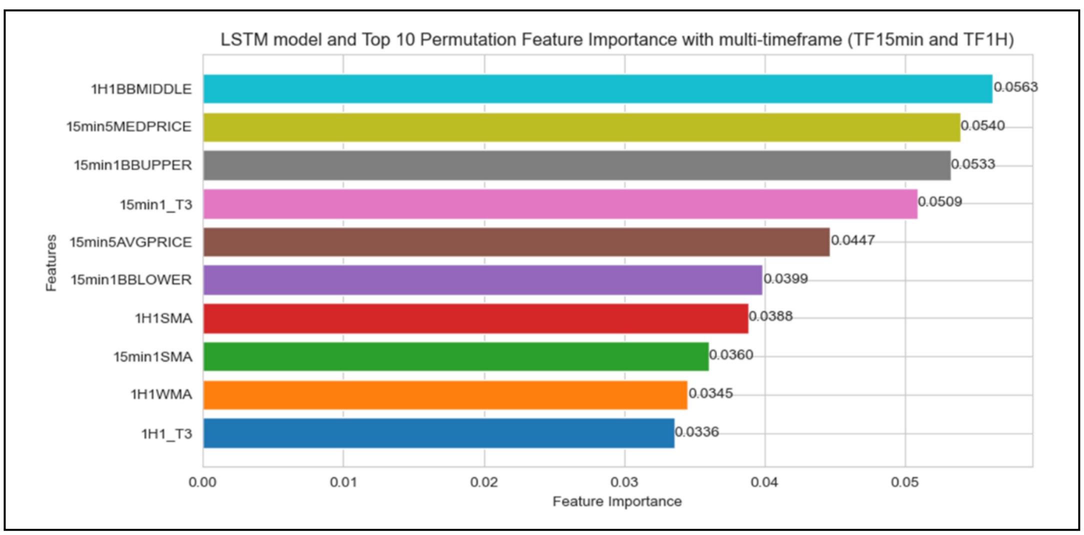

Figure 9 LSTM and Top 10 feature importance (15min, 1H)

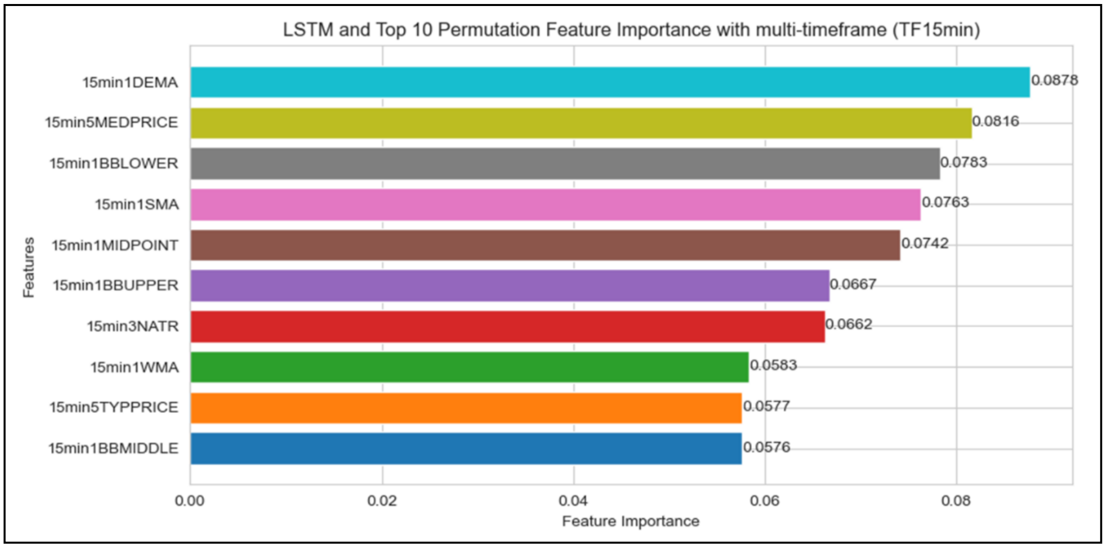

Figure 10 LSTM and Top 10 feature importance (15min)

Figures 8-10, in conjunction with trading strategy analysis, reveals the influence of timeframe selection on feature importance within LSTM models. Two high-performing timeframe configurations emerge: a three-timeframe combination (5-minute/15-minute/1-hour) and a two-timeframe combination (15-minute/1-hour). When using three timeframes, Category 3 (Volatility Indicators), specifically Average True Range (ATR) features, exert the strongest influence on predictions. Category 1 (Overlap Studies), including Bollinger Bands, is also important but less dominant. Feature influence shifts when analyzing two-timeframe combinations. With just 15-minute and 1-hour intervals, Category 1 (Overlap Studies) and Category 5 (Price Transform) features, like Median Price, exhibit the most impact.

The experimental results demonstrate that employing multiple timeframes, particularly the 3TF(5, 15, 60) configuration, significantly enhances the accuracy of SET50 Index Futures price predictions. The XGBoost model achieved the highest Hit Ratio (HR) of 62.94% with the 3TF configuration. The Accumulated Return (AR) analysis shows that the XGBoost model using technical indicators from the 2TF (15, 60) configuration achieved the highest AR of 132.95%, outperforming other models. A reduction in the number of timeframes resulted in a noticeable decline in profitability across all models. The Sharpe Ratio (SR) analysis reveals that the XGBoost model with the 2TF (15, 60) configuration achieved the highest SR of 5.54, indicating superior risk-adjusted returns. Both models experienced a significant performance decline when utilizing only a single 15-minute timeframe, emphasizing the importance of multiple timeframe analysis.

Discussion

The findings highlight that the use of multiple timeframes significantly enhances the prediction accuracy of SET50 Index Futures prices. Specifically, the 2TF and 3TF configurations proved to be the most effective. Integrating various technical indicators improved model accuracy, with indicators from Category 1 (Overlap Studies Indicators) and Category 5 (Price Transform) being particularly influential. This discussion underscores the importance of selecting diverse timeframes and appropriate hyperparameter tuning in the prediction process. It also points to the potential application of machine learning models in improving trading strategies and risk management in Thailand's financial markets. The results indicate that both XGBoost and LSTM models can significantly enhance returns and reduce risks when multiple timeframes are utilized.

Conclusion

This research provides a detailed analysis of forecasting SET50 Index Futures prices using XGBoost and LSTM models within a management science framework, emphasizing the importance of multiple timeframes. The study demonstrates that incorporating multiple timeframes significantly enhances prediction accuracy and the profitability of trading strategies. XGBoost shows superior performance across various timeframe configurations, particularly with the two-timeframe (15-minute/1-hour) combination, achieving the highest Sharpe Ratio. Key features influencing model performance include the Triple Exponential Moving Average and other technical indicators. This study highlights the critical role of multi-timeframe analysis in capturing market dynamics, thereby supporting informed decision-making and strategic planning in financial management. Future research should explore the integration of additional technical indicators and advanced machine learning techniques to further improve prediction accuracy and risk management, enhancing the effectiveness of management practices in financial markets.

Future work

In future research, the study will be extended to investigate other asset classes, such as stocks, gold, international markets, and Bitcoin.

Reference

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32. http://dx.doi.org/10.1023/A:1010933404324

Cao, J., Chen, J., Hull, J., & Poulos, Z. (2020). Deep hedging of derivatives using reinforcement learning. *The Journal of Financial Data Science*, *3*(1), 10–27. https://doi.org/10.3905/jfds.2020.1.052

Chance, D. M., & Brooks, R. (2009). *An introduction to derivatives and risk management*. South-Western Pub.

Chen, L., Qiao, Z., Wang, M., Wang, C., Du, R., & Stanley, H. E. (2018). Which artificial intelligence algorithm better predicts the Chinese stock market?. *IEEE Access*, *6*, 48625-48633.

Chen, R., & Pan, B. (2016). Chinese Stock Index Futures Price Fluctuation Analysis and Prediction Based on Complementary Ensemble Empirical Mode Decomposition. *Mathematical Problems in Engineering*, *2016*, 1–13. https://doi.org/10.1155/2016/3791504

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). https://doi.org/10.1145/2939672.2939785

Chatziantoniou, I., Degiannakis, S., & Filis, G. (2019). Futures-based forecasts: How useful are they for oil price volatility forecasting?. *Energy Economics*, *81*, 639-649.

Czudaj, R. L. (2019). Crude oil futures trading and uncertainty. *Energy Economics*, *80*, 793-811.

Deng, S., Zhu, Y., Huang, X., Duan, S., & Fu, Z. (2022). High-frequency direction forecasting of the futures market using a machine-learning-based method. *Future Internet*, *14*(6), 180.

Deng, S., Huang, X., Wang, J., Qin, Z., Fu, Z., Wang, A., & Yang, T. (2020). A decision support system for trading in apple futures market using predictions fusion. *IEEE Access*, *9*, 1271-1285.

Deng, S., Zhu, Y., Duan, S., Yu, Y., Fu, Z., Liu, J., Yang, X., & Liu, Z. (2023). High-frequency forecasting of the crude oil futures price with multiple timeframe predictions fusion. Expert Systems With Applications, 217, 119580. https://doi.org/10.1016/j.eswa.2023.119580

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1-81.

Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, *34*, 100190. https://doi.org/10.1016/j.cosrev.2019.08.001

Gong, X., & Lin, B. (2021). Effects of structural changes on the prediction of downside volatility in futures markets. *Journal of Futures Markets*, *41*(7), 1124-1153.

Gottesman, A. (2016). *Derivatives essentials: An Introduction to Forwards, Futures, Options and Swaps*. US: Wiley.

Hao, Y., & Gao, Q. (2020). Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning. *Applied Sciences*, *10*(11), 3961.

Kaufman, P. J. (2019). *Trading Systems and Methods*. 6th Edition. US: Wiley.

Kumar, D. S., Thiruvarangan, B. C., Vishnu, A., Devi, A. S., & Kavitha, D. (2022). Analysis and prediction of stock price using hybridization of sarima and xgboost. In *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)* (pp. 1-4). IEEE.

Liu, J., Zhang, B., Zhang, T., & Wang, J. (2023). Soybean Futures Price Prediction Model Based on EEMD-NAGU. *IEEE Access*.

Liu, G., Mao, Y., Sun, Q., Huang, H., Gao, W., Li, X., & Wang, X. (2020). Multi-scale Two-way Deep Neural Network for Stock Trend Prediction. In *IJCAI* (pp. 4555-4561).

Liu, S., Liao, G., & Ding, Y. (2018). Stock transaction prediction modeling and analysis based on LSTM. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 2787-2790). IEEE.

Loveless, J., Stoikov, S., & Waeber, R. (2013). Online algorithms in high-frequency trading. *Communications of the ACM*, *56*(10), 50–56. https://doi.org/10.1145/2507771.2507780

Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, *170*, 1168-1173.

Nobre, J., & Neves, R. F. (2019). Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets. *Expert Systems with Applications*, *125*, 181-194.

Peterson, P. E. (2018). *Commodity derivatives: A Guide for Future Practitioners*. New York: Routledge.

Reboredo, J. C., Matías, J. M., & Garcia-Rubio, R. (2012). Nonlinearity in forecasting of high-frequency stock returns. Computational Economics, *40*(3), 245-264.

Sanboon, T., Keatruangkamala, K., & Jaiyen, S. (2019, February). A deep learning model for predicting buy and sell recommendations in stock exchange of thailand using long short-term memory. In *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)* (pp. 757-760). IEEE.

Savani, R. (2012). High-Frequency trading: the faster, the better? *IEEE Intelligent Systems*, *27*(4), 70–73. https://doi.org/10.1109/mis.2012.75

Wang, C., & Gao, Q. (2018). High and low prices prediction of soybean futures with LSTM neural network. In *2018 IEEE 9th international conference on software engineering and service science (ICSESS)* (pp. 140-143). IEEE.

Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications*, *186*, 115716.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, *31*(7), 1235-1270.

Zou, Z., & Qu, Z. (2020). Using LSTM in stock prediction and quantitative trading. *CS230: Deep Learning, Winter*, 1-6.