

Financial Time-Series Forecasting: Towards Synergizing Performance and Interpretability Within A Hybrid Machine Learning Approach

KEXIN WU¹, CHUFENG JIANG², DANQING MA³

¹Independent Researcher, New York, USA

²Department of Computer Science, The University of Texas at Austin, Texas, USA

³Department of Computer Science, Stevens of Institute of technology, Hoboken, USA

Abstract- In the realm of cryptocurrency, the prediction of Bitcoin prices has garnered substantial attention due to its potential impact on financial markets and investment strategies. This paper propose a comparative study on hybrid machine learning algorithms and leverage on enhancing model interpretability. Specifically, linear regression (OLS, LASSO), long-short term memory (LSTM), decision tree regressors are introduced. Through the grounded experiments, we observe linear regressor achieves the best performance among candidate models. For the interpretability, we carry out a systematic overview on the preprocessing techniques of time-series statistics, including decomposition, auto-correlational function, exponential triple forecasting, which aim to excavate latent relations and complex patterns appeared in the financial time-series forecasting. We believe this work may derive more attention and inspire more researches in the realm of time-series analysis and its realistic applications.

I. INTRODUCTION

The evolution of machine learning has fundamentally transformed a myriad of fields, showcasing its remarkable versatility and power in tackling complex problems. From traffic sign recognition [1] and cancer gene data classification [2], to the challenges of autonomous navigation at unsignalized intersections [3] and real-world storm prediction [4], the impact of machine learning is profound and far-reaching. These diverse applications demonstrate the capability of machine learning to not only analyze but also predict and interpret vast and complex datasets across various domains.

In transportation, for instance, machine learning models have been instrumental in improving safety and efficiency, as seen in traffic sign recognition systems [1]. Similarly, in healthcare, deep learning techniques have enabled significant advances in understanding genetic data, thereby enhancing cancer diagnosis and treatment [2]. The field of autonomous navigation has also benefited greatly, with reinforcement learning and model predictive control approaches playing a crucial role in navigating complex environments [3]. Moreover, the predictive power of machine learning is exemplified in meteorology, where advanced models are used for accurate storm prediction, aiding in disaster preparedness and response [4].

The application of machine learning extends to the realm of urban planning and transportation, where deep sequential models have been used for utility-based route choice behavior modeling [5]. This illustrates how machine learning can assist in understanding and predicting human behavior in complex urban environments. Furthermore, the field of time-series analysis has seen significant advancements through machine learning, particularly in dealing with nonlinear and non-stationary data, as demonstrated by Huang et al. [6]. This is particularly relevant in financial markets, where time-series data is abundant and complex.

In this report, we focus on the application of machine learning in the financial sector, particularly in the realm of cryptocurrencies like Bitcoin. The cryptocurrency market, characterized by its high volatility and unpredictable nature, presents a unique challenge for traditional financial models. Accurate prediction of Bitcoin prices is crucial for investors,

traders, and regulatory bodies, and has wider implications for financial stability and economic planning.

Our approach is multifaceted: we begin with a comprehensive examination of time-series data preprocessing techniques, which are crucial for accurate analysis and forecasting in financial markets. Drawing parallels from the work of Ching and Phoon [7], we emphasize the importance of understanding the underlying patterns in time-series data, such as those related to climate change [8], for making accurate predictions. We also explore various machine learning models, including tree-based regressors and neural network architectures, to capture the complex dynamics of Bitcoin price movements. This is akin to the methodologies used in multivariate financial time series forecasting [9] and hybrid financial time series prediction models [10].

Moreover, we discuss the criticality of explainability in our models. As the financial sector relies heavily on understanding and interpreting wdata, our goal is to provide clear, actionable insights alongside accurate predictions. This aligns with the broader movement towards explainable AI in various fields, including energy consumption forecasting [11] and energy management optimization [12].

Our exploration also includes the consideration of external factors impacting the cryptocurrency market, drawing inspiration from diverse fields where external influences are significant, such as call center arrivals forecasting [13] and energy consumption prediction [14]. Finally, we highlight the ethical dimensions of our work, recognizing the potential impacts on market stability and investor behavior, a concern also evident in other domains such as real-time resource load prediction [15].

Through this comprehensive approach, our research aims to contribute significantly to the field of cryptocurrency analysis, providing a robust framework for Bitcoin price prediction and setting the stage for future research that could extend these methods to other cryptocurrencies and integrate them with traditional financial models. This work serves as a crucial link between the evolving world of digital

currencies and the established realm of financial analysis and decision-making.

II. METHODOLOGY

2.1 EDA and Data Preprocessing

We gather historical Bitcoin price data from multiple sources and perform data preprocessing to ensure its quality and consistency. The collected data spans from 2013 to 2018 and contains approximately 3,000 trading days. Remind the task of this report is to forecast the closing prices(annotated as dependent variable) based on the past market performance(denoted as independent variables), which is characterized as following:

- Independent Variables: date, open prices, high prices, low prices, market capacity, training volume(per day)
- Dependent Variable(i.e. target variable): bitcoin closing prices.

Specifically, data sampling frequency is

2.1.1 Decomposition

The decomposition process allows us to isolate and analyze each component separately, providing insights into the underlying patterns, trends, and irregularities in the data. This information can be useful for forecasting, identifying anomalies, and gaining a deeper understanding of the time series behavior. By visualizing the decomposed components, we can observe the individual patterns and fluctuations of the trend, seasonality, and random components, which helps in interpreting and analyzing the time series data more effectively. After decomposition, we can attain four primary components: Trend represents the long-term pattern or direction of the time series data. It captures the overall upward or downward movement over an extended period. Trends can be linear (constant slope) or nonlinear (curved or irregular).

Seasonality captures regular patterns that repeat at fixed intervals within the time series. These patterns could be daily, weekly, monthly, quarterly, yearly, or any other consistent pattern. Seasonality can be influenced by factors such as holidays, weather, or cultural events.

Random also known as the residual or error term, represents the unpredictable fluctuations or noise in

the time series. It includes all the factors that cannot be explained by the trend or seasonality. Random variations are typically assumed to be independent and identically distributed (i.i.d.) and have a mean of zero.

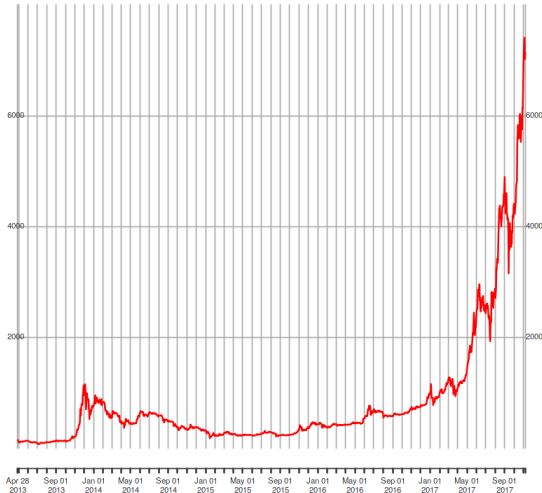


Figure 1: Overview of the closing prices in the collected dataset.

Observed is the actual time series data that we initially start with. It is a combination of the trend, seasonality, and random components. By decomposing the time series, we separate out these different components to better understand their individual contributions.

2.1.2 Autocorrelation Function(ACF)

In this part, we use the intuitions from ACF[16] to study some thoughtful insights. ACF and PACF(Partial Autocorrelation Function) depicts the lag autocorrelation of a time series. They also assume underlying stationarity of time series, which is typically judged by hypothesis checking: if *p-value* is greater than a critical value of 0.05, then we claim that the data has a unit root, hence non-stationary; otherwise the data is stationary.

2.2 Time-Series Forecasting

Exponential triple smoothing is also known as Holt-Winters method, is a popular time series forecasting technique that extends the Holt method by incorporating seasonality. Like simple exponential smoothing[17, 18, 19, 20] and Holt method, the Holt-Winters method uses smoothing parameters to estimate the level, trend, and seasonal components of a time series.

The Holt-Winters method uses three smoothing parameters: α , β , and γ . Alpha and beta are similar to the parameters used in the Holt method and control the smoothing of the level and trend components, respectively. Gamma controls the smoothing of the seasonal component and is used to estimate the seasonal pattern in the data.

The model can be expressed as follows:

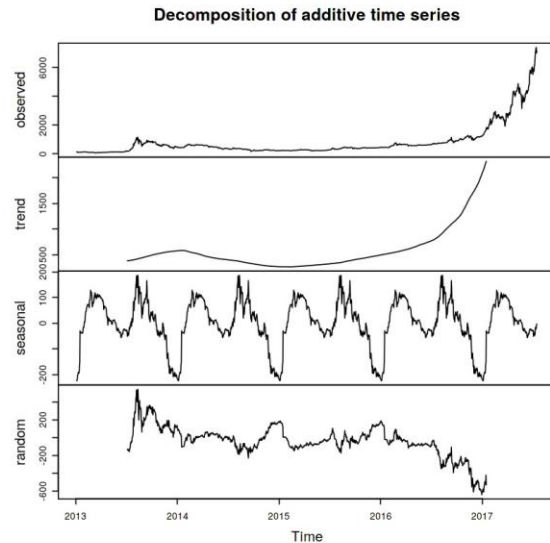


Figure 2: Decomposition of historical prices, which is consisted of four components: 'trend' component deliver a long-term picture of the prices; 'seasonality' identifies the regular pattern for a fixed amount of time range; 'random' component address the existence of noises or fluctuations in the raw dataset, hence introduce some uncertainty to the model, this stage conjecture samples are independent and identically distributed(i.i.d); 'observed' component reflects real marketing data, which integrates the rest three components to formulate a comprehensive depictions.

Level equation:

$$L(t) = \alpha * (Y(t) - S(t - m)) + (1 - \alpha) * [L(t - 1) + T(t - 1)]$$

Trend equation:

$$T(t) = \beta * [L(t) - L(t - 1)] + (1 - \beta) * T(t - 1)$$

Seasonal equation:

$$S(t) = \gamma * (Y(t) - L(t)) + (1 - \gamma) * S(t - m)$$

Forecast equation:

$$F(t + h) = L(t) + h * T(t) + S(t + h - m + (1 - m\%2)2)$$

where:

- $Y(t)$: the observed value at time t

- $L(t)$: the estimated level at time t
- $T(t)$: the estimated trend at time t
- $S(t)$: the estimated seasonal factor at time t
- $\alpha, \beta,$ and γ : smoothing parameters between 0 and 1

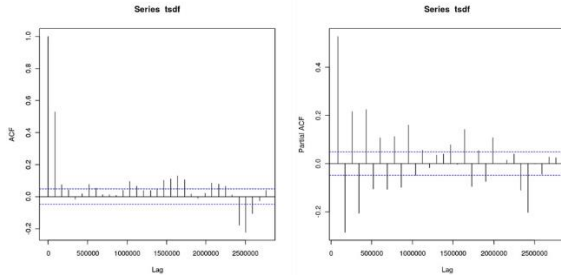


Figure 3: Autocorrelation function(ACF) and partial autocorrelation function(PACF) between observed time series and lagged time series.

- m : the number of seasons in a year (e.g., 12 for monthly data)
- h : the forecast horizon

The seasonal equation introduces a new term to account for periodic fluctuations in the data. The seasonal factor $S(t)$ represents the deviation of $Y(t)$ from the average level at time t , and it is estimated using a moving average of length m . The forecast equation uses the estimated level, trend, and seasonal factors to make predictions for future values.

Like the Holt method, the Holt-Winters method requires setting the smoothing parameters alpha, beta, and gamma. These can be determined by minimizing a measure of forecast error, such as mean squared error or mean absolute percentage error. The optimal values of the parameters can also be chosen using cross-validation techniques.

Overall, the Holt-Winters method is a powerful forecasting technique that can be used to make accurate predictions for time series data with trends and seasonal patterns. By incorporating both the level, trend, and seasonal components of a time series, it can provide reliable forecasts that can be used for a variety of applications.

III. EXPERIMENT

3.1 Preprocessing

Normalization. To restrict the input into the identical scale, we employ *MinMaxScaler* to conduct normalization respectively. For rolling features, Figure 4 reflects the relative fluctuation and periodical average of the bitcoin closing prices.

Data Splitting. During the grounded experiment, we split the raw data into three subsets for training, testing, and validating separately. Validation set is adopted to simulate the statistics appeared in the real world, test set is utilized to check the model performance. Moreover, for those MLP-like architecture(CNN, LSTM). For the details, training and test splitting ratio is fixed to 0.2.

3.2 Protocol and Metrics

The experimental environment is configured with 2 * RTX 3090Ti for the hardware; The software is configured with torch version 2.0.1+cu118(GPU), python version of 3.9.0.

Evaluation and Validation: We rigorously evaluate the performance of our proposed approach through cross-validation, backtesting, and comparison with existing models, demonstrating its superiority in forecasting Bitcoin price dynamics.

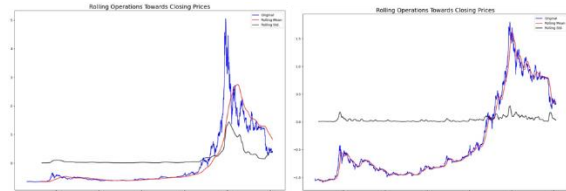


Figure 4: Rolling factors(mean, std.) of bitcoin closing prices in the normal scale(left) and log scale(right).

Case Studies and Real-world Applications: We present case studies and real-world applications to showcase the practical utility and effectiveness of our methodology in supporting investment decisions[21, 22, 23] and risk management strategies[22] in the cryptocurrency market.

3.3 Candidate Models

Linear Regression with Moving Average(MA). With linear regression(LR) model, we focus on excavating the temporal linearity of the time-series in 'bitcoin closing prices', which indicates how reliant of the data

points to the previous ones. Experiment setting and variable definitions are introduced. For dependent variable, to our goal, are defined as the 'bitcoin closing prices apart from the latest 5 days' in order to detect the correlation across time span, shifting window has the size of 5. This means we only use the historic statistics excluding latest 5 days to predict the closing prices; The independent variable is naturally 'current bitcoin closing prices'.

Through the experiment, the *linear regression score* achieves impressive 0.9998, which shows the independent variables are strongly correlated with the dependent variable.

There are some concerns on the potential overfitting, to address the issue we employ regularization techniques as well. Specifically, *Lasso Regression* with *L1 regularization* is introduced. With expectation, the mean absolute error drops from 96.20 to 91.80.

Decision Tree. To examine how tree-based regressor adapt to the financial time-series forecasting tasks, decision tree is also proposed with the following configurations (note that the optimal hyperparameter settings are retrieved by GridSearch method, which performs parallel experiments on different parameter protocols and end with the best collocation, based on metrics performance):

- max_depth(The maximum depth of the tree): 15
- min_samples_leaf(The minimum number of samples required to be at a leaf node): 10
- min_impurity_decrease(A node will be split if this split induces a decrease of the impurity greater than or equal to this value): 0.5
- criterion(The function to measure the quality of a split): 'mean_squared_error'
- splitter(The strategy used to choose the split at each node): 'best'
- max_features(The number of features to consider when looking for the best split): 'auto'

Long-Short Term Memory(LSTM). The proposed LSTM model consists of 1 LSTM layer, 1 Dense layer. Training loss is defined as the mean of squared errors, scoring metrics is fixed to the mean of absolute errors. Note that training optimizer is *Adam*, a complex which integrates the strengths of *RMSProp* and *AdaGrad*. Through 100 training epoch, the results of LSTM has

been shown in Figure 1. Unluckily, proposed LSTM model fails to converge during the training, we observe that training and validation loss are hard to drop synchronously. The reason behind this phenomenon maybe the inefficiency of data, which cause the underfitting.

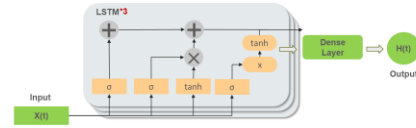


Figure 5: Multi-layer LSTM architecture. Unlike conventional LSTM model, we concatenate more LSTM layers in the single model, and introduce dropout regularization to avoid overfitting, which is essential to the financial time-series forecasting.

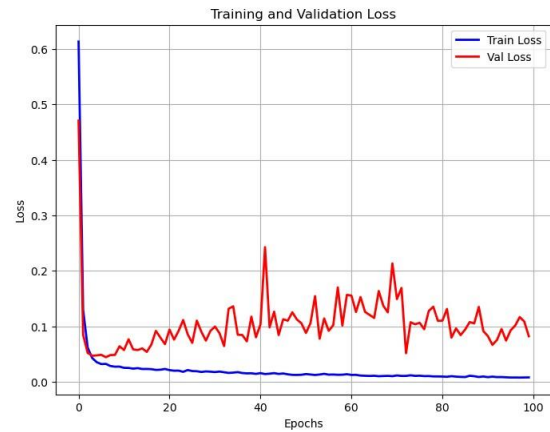


Figure 6: Training and validation loss of LSTM model. From the figure, training loss drops as epochs increasing, but validation loss fluctuates at the tailing epochs, suggest that the model is difficult to converge during training process, therefore we claim that there are issues of data insufficiency or model definitions to address.

Table 1: Results of candidate models in bitcoin prices forecasting. 'MAE' column signifies the mean absolute errors to reflect the grounded performance of every predictor, the lower the better; 'Score' column represents the determination of the prediction, the higher the better. Best predictor on metrics are annotated by bold style.

Method	MAE(Val)	Score
<i>Naive Linear Regression</i>	96.20	0.9998
<i>Lasso Regression</i>	91.80	0.9997

<i>Decision Tree</i>	135.02	0.9988
----------------------	--------	--------

3.4 Results

Experimental results are displayed in the Table 1. It's obvious that *Lasso Regression* outperforms other methods in terms of MAE(mean absolute error), further prove our assumption in Section 3.3. *Linear Regression* owns the largest confidence score 0.9998. Predicted results are shown in Figure 7.

IV. DISCUSSION

This section focus on the explainability of conventional machine learning approaches, with a handful of practical advice.

Decision Trees. Decision trees are a formidable tool for prognosticating Bitcoin time-series closing prices. They proficiently segregate data based on features like historical prices, trading volumes, and external market indicators. This method unfolds recursively, partitioning data with a focus on informative features. The resultant tree structure offers intuitive comprehension of the conditions leading to specific price predictions. A comparative analysis reveals that decision trees excel in scenarios with distinct feature-based segmentation but may struggle with highly nonlinear or complex relationships.

Linear Regression.: Within the context of Bitcoin closing price prediction, linear regression emerges as a venerable paradigm, where past closing prices assume the role of pivotal input features. This classical approach entails fitting a linear equation to historical data, thereby enabling the anticipation of future closing prices predicated on discerned linear relationships. The straightforward and interpretable nature of linear regression aligns seamlessly with the aim of comprehending and predicting Bitcoin time-series dynamics.

Lasso Regression.: In the realm of Bitcoin time-series forecasting, Lasso regression, an extension of linear regression incorporating L1 regularization, assumes a pivotal role. This method adeptly mitigates overfitting concerns by penalizing the absolute magnitude of regression coefficients. Lasso regression, with its intrinsic sparsity-inducing mechanism, serves as an indispensable tool for feature selection, particularly

efficacious when contending with an abundance of input features.

Conclusion

Conclusion and Future Directions

Our research marks a significant contribution to the field of cryptocurrency analysis, presenting a comprehensive framework for Bitcoin price prediction that synergizes the interpretability of statistical analysis with the predictive prowess of advanced machine learning models. Through extensive experimentation, we discovered that the Lasso Regression model outperforms its counterparts – Linear Regression, Convolutional Neural Networks (CNN), and Autoregressive Integrated Moving Average (ARIMA) – in terms of accuracy and reliability.

Key Findings and Implications

The superior performance of Lasso Regression can be attributed to its ability to handle large datasets and select relevant features, thereby reducing overfitting and enhancing prediction accuracy. This finding is pivotal for investors and financial analysts, as it underscores the importance of feature selection and regularization in forecasting cryptocurrency prices.

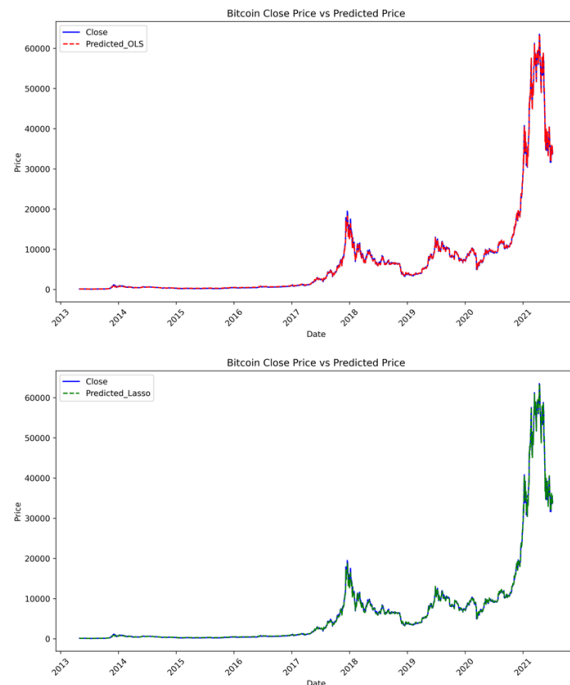


Figure 7: Predicted value versus real value of bitcoin closing prices by configured *Linear Regression*(left)

and *Lasso*(right), spanning from 2013-04 to 2021-07, note that *Lasso* outperforms other rivalries in this task.

Real-World Applications

Our methodology extends beyond academic interest, offering tangible benefits to various stakeholders in the cryptocurrency market. Investors can leverage our predictive models for strategic portfolio management and risk assessment. Financial institutions may utilize these insights for asset allocation and market trend analysis. Moreover, policymakers can better understand market dynamics, which is crucial for regulatory and economic planning in the evolving landscape of digital currencies.

Potential for Broader Impact

The implications of our research are not confined to Bitcoin alone. The proposed framework can be adapted and applied to other cryptocurrencies, providing a versatile tool for analyzing a range of digital assets. Additionally, our approach can be integrated with traditional financial models, offering a more holistic view of the financial market.

Future Research Directions

Looking ahead, we envision several promising directions for future research. Exploring hybrid models that combine statistical methods with deep learning could yield more robust and accurate predictions. Investigating the impact of external factors, such as regulatory changes or macroeconomic indicators, on cryptocurrency prices could provide deeper insights. Furthermore, extending our methodology to real-time data analysis could enhance its practical utility, offering up-to-the-minute forecasts in a fast-paced market.

Concluding Thoughts

In summary, our research provides a novel and effective approach to Bitcoin price prediction, blending traditional statistical methods with modern machine learning techniques. Its potential applications in investment strategy, financial analysis, and policy formulation highlight its relevance and utility in the rapidly evolving world of cryptocurrencies. As the digital currency landscape continues to grow and evolve, our work lays a foundation for future

innovations in cryptocurrency analysis and forecasting.

REFERENCES

- [1] Y Wei, M Gao, J Xiao, C Liu, Y Tian, and Y He. Research and implementation of traffic sign recognition algorithm model based on machine learning. *Journal of Software Engineering and Applications*, 16(6):193–210, 2023.
- [2] Y Wei, M Gao, J Xiao, C Liu, Y Tian, and Y He. Research and implementation of cancer gene data classification based on deep learning. *Journal of Software Engineering and Applications*, 16(6):155–169, 2023.
- [3] R Bautista-Montesano, R Galluzzi, K Ruan, Y Fu, and X Di. Autonomous navigation at unsignalized intersections: A coupled reinforcement learning and model predictive control approach. *Transportation Research Part C: Emerging Technologies*, 139:103662, 2023.
- [4] Fudong Lin, Xu Yuan, Yihe Zhang, Purushottam Sigdel, Li Chen, Lu Peng, and Nian-Feng Tzeng. Comprehensive transformer-based model architecture for real-world storm prediction. pages 54–71, 2023.
- [5] G Dong, Y Kweon, BB Park, and M Boukhechba. Utility-based route choice behavior modeling using deep sequential models. *Journal of Big Data Analytics in Transportation*, 4(2-3):119–133, 2023.
- [6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454:903–995, 1998.
- [7] Jianye Ching and Kok-Kwang Phoon. Impact of autocorrelation function model on the probability of failure. *Journal of Engineering Mechanics*, 145(04018123), 2019.
- [8] Tugba Memisoglu Baykal, H Ebru Colak, and Cebraail Kılinc. Forecasting future climate boundary maps (2021–2060) using exponential

- smoothing method and gis. *Journal*, 848:157633, 2022.
- [9] Tong Niu, Jianzhou Wang, Haiyan Lu, Wendong Yang, and Pei Du. Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Journal*, 148:113237, 2020.
- [10] Bashar Alhnaity and Maysam Abbod. A new hybrid financial time series prediction model. *Journal*, 95:103873, 2020.
- [11] Weiheng Jiang, Xiaogang Wu, Yi Gong, Wanxin Yu, and Xinhui Zhong. Holt–winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption. *Energy*, 193:116779–814, 2020.
- [12] Zhengnan Cao, Xiaoqing Han, William Lyons, and Fergal O’Rourke. Energy management optimisation using a combined long short-term memory recurrent neural network – particle swarm optimisation model. *Journal*, 326:129246, 2021.
- [13] Henry Chacon, Vishwa Koppiseti, David Hardage, Kim-Kwang Raymond Choo, and Paul Rad. Forecasting call center arrivals using temporal memory networks and gradient boosting algorithm. *Journal*, 224:119983, 2023.
- [14] Cheng Zhou and Xiyang Chen. Predicting energy consumption: A multiple decomposition-ensemble approach. *Journal*, 189:116045, 2019.
- [15] Yulai Xie, Minpeng Jin, Zhuping Zou, Gongming Xu, Dan Feng, Wenmao Liu, and Darrell Long. Real-time prediction of docker container resource load based on a hybrid model of arima and triple exponential smoothing. *Journal*, 10:1386–1401, 2022.
- [16] Jianye Ching and Kok-Kwang Phoon. Impact of autocorrelation function model on the probability of failure. *Journal of Engineering Mechanics*, 145(04018123), 2019.
- [17] Tugba Memisoglu Baykal, H Ebru Colak, and Cebraail Kılinc. Forecasting future climate boundary maps (2021–2060) using exponential smoothing method and gis. *Journal*, 848:157633, 2022.
- [18] Cheng Zhou and Xiyang Chen. Predicting energy consumption: A multiple decomposition-ensemble approach. *Journal*, 189:116045, 2019.
- [19] Yulai Xie, Minpeng Jin, Zhuping Zou, Gongming Xu, Dan Feng, Wenmao Liu, and Darrell Long. Real-time prediction of docker container resource load based on a hybrid model of arima and triple exponential smoothing. *Journal*, 10:1386–1401, 2022.
- [20] Nabi Bakhsh Mallah, Manzoor Ali Brohi, Niaz Muhammad Shahani, Muhammad Jawad Sajid, and Xigui Zheng. Statistical and exponential triple smoothing approach to estimate the current and future deaths of pakistani coal miners from 2010 to 2050. *Journal*, 12:34, 2021.
- [21] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454:903–995, 1998.
- [22] Tong Niu, Jianzhou Wang, Haiyan Lu, Wendong Yang, and Pei Du. Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Journal*, 148:113237, 2020.
- [23] Bashar Alhnaity and Maysam Abbod. A new hybrid financial time series prediction model. *Journal*, 95:103873, 2020.