Stock Price Index Prediction of Several ASEAN Countries Using Panel Data Regression and Machine Learning Algorithm

Raihan Adam Handoyo Winarso
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025241043@student.its.ac.id

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@its.ac.id

Agus Tri Haryono
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025231020@student.its.ac.id

Abdullah Faqih Septiyanto
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025221022@student.its.ac.id

Shoffi Izza Sabilla
Department of Medical Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
shoffi@its.ac.id

Yusril Falih Izzaddien
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025241026@student.its.ac.id

2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C) | 979-8-3315-0857-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/BTS-I2C63534.2024.10942143

***Abstract*—The analysis of major stock indices plays a crucial role in understanding market sentiment and predicting economic performance in the ASEAN region. Several studies have examined panel data regression and machine learning methods to analyze macroeconomic variables related to stock indices, including boosting models such as XGBoost, CatBoost, and LightGBM, which have proven effective in handling the complexity of multivariable financial data. This study aims to explore the use of the Fixed Effect Model (FEM) to measure the impact of variables such as gold prices, dollar exchange rates, bond yields, and commodity prices on major stock indices, as well as employing machine learning for short-term forecasting. Daily data from 2019-2024 for ASEAN countries was used, with models selected based on significance tests, accuracy, MAE, MSE, and R^2. The results show that LightGBM achieves the highest accuracy, with an MAE of 31.05 and 53.77, and an MSE of 3105.61 and 9786.65 on training and testing data, respectively, achieving 100% R^2 on both. CatBoost and XGBoost also perform well, with high accuracy and low error values. Results show LightGBM as a reliable tool for projecting ASEAN stock market trends.**

Keywords—adaboost, catboost, lightgbm machine learning, panel data regression, random forest, stock index asean, xgboost.

I. INTRODUCTION

The Association of Southeast Asian Nations (ASEAN) is a regional organization established by Southeast Asian countries to enhance cooperation in various fields such as politics, economy, and socio-culture. After the global crisis in 2008, ASEAN countries began to improve their economies. Overall, the economic conditions of ASEAN plays a significant role in the global economy, reflected in the capital market as an economic health indicator [1]. One of the indicators reflecting stock market performance is the main stock index. This index serves as an indicator of stock market performance in each country. The main stock index is influenced by investor sentiment, economic conditions, and commodity prices [2].

Understanding the extent of the influence of a country's variables on main stock price index in ASEAN countries requires an analysis using Panel Data Regression. This method measures the impact of various factors, on stock indices while accounting for differences across countries and time dynamics. Additionally, machine learning methods such as CatBoost, AdaBoost, XGBoost, LightGBM, and Random Forest are employed to predict future movements of main stock indices.

Previous research has been conducted on the use of panel data regression to analyze the factors influencing composite stock indices in Indonesia and Malaysia [18]. Also, the prediction of stock indices using machine learning algorithms, such as XGBoost, has shown promising results by utilizing investor sentiment to predict the direction of stock indices in the Chinese market [10]. Similarly, in stock price prediction using Random Forest and SVM, it was found that Random Forest effectively classifies price patterns [19]. Furthermore, in the Hybrid AdaBoost-LSTM model, which combines feature selection with LSTM to improve the accuracy of futures index prediction, it outperforms other models like Random Forest and Deep Belief Network [20]. These studies often face the challenge of data scarcity, which can limit model performance. Despite these challenges, the models in this study focus on optimizing available data without the use of data augmentation or transfer learning techniques, relying instead on the inherent structure and features of the dataset to enhance prediction accuracy.

The purpose of this panel data analysis is to examine how macroeconomic variables, commodity prices, and consumer sentiment influence the main stock price indices across several ASEAN countries. The significant variables identified will be used in predictive models with machine learning methods to determine the best-performing algorithm based on accuracy and effectiveness. This approach aims to provide more precise predictions and support better-informed investment decisions within ASEAN markets [3]. The selected methods XGBoost, LightGBM, Random Forest, AdaBoost, and CatBoost are chosen for their ability to handle complex data, improve prediction accuracy, and optimize model performance [10].

II. LITERATURE REVIEW

A. Panel Data Regression

Panel data regression is a statistical method used to analyze relationships between variables by combining time series and cross-sectional data, such as individuals, companies, or countries. It allows researchers to identify patterns over time while observing differences between entities. The general formula for Panel Data Regression is shown in Equation 1.

1. Panel Data Regression Model

$$y_{\mathrm{it}} = \mathbf{x}'_{\mathrm{it}}\,\boldsymbol{\beta} + \mathbf{Z}'_i\boldsymbol{\alpha} + \varepsilon_{\mathrm{it}}$$
$$i = 1, \ldots, K; t = 1, \ldots, T \quad (1)$$

Equation (1) represents the general form of panel data regression, where *i* represents the cross-sectional units numbered *K*, while *t* represents the time periods numbered *T*. There are p independent variables in $\mathbf{x_{it}}$, excluding the constant term. The individual-specific effect is $\mathbf{Z_i'}\,\boldsymbol{\alpha}$, where $\mathbf{Z_i}$ consists of the constant and individual-specific effects, which may be either observable or unobservable. $\boldsymbol{\beta}$ is a slope matrix of size *p* x *1* [6].

2. General Model Structure

Estimating a panel regression model relies heavily on the assumptions made regarding the intercept, slope coefficients, and error term. Based on these assumptions and their influencing factors, the model structure is categorized into three types: Pooled Regression, Fixed Effect, and Random Effect. [7].

a. Pooled Regression

This model assumes that $\mathbf{Z_i}$ consists only of a constant term, meaning that there are no individual-specific effects [8]. This model structure is often referred to as the Common Effect Model (CEM) and is expressed in Equation (2).

$$y_{it} = \alpha + \mathbf{x}_{\mathrm{it}}'\,\boldsymbol{\beta} + \varepsilon_{\mathrm{it}}$$
$$i = 1, \dots, K; t = 1, \dots, T \qquad (2)$$

Equation (2) describes the components of the panel data regression model, where α is the intercept coefficient (constant), which is a scalar value, $\boldsymbol{\beta}$ is a slope matrix of size *p* x *1* and $\mathbf{x}_{\mathrm{it}}$ represents the observation at the *i*-th unit and *t*-th time for the p-th explanatory variable.

b. Fixed Effect

The fixed effect model structure is a model that accounts for the variability of independent variables across individuals [8]. The Fixed Effect Model is expressed in Equation (3).

$$y_{it} = \mathbf{x}_{\mathrm{it}}'\,\boldsymbol{\beta} + \alpha_{\mathrm{i}}\,\varepsilon_{\mathrm{it}} \qquad (3)$$

Equation (3) further specifies the panel data regression model, where $\alpha_{\mathrm{i}} = \mathbf{Z_i'}\,\boldsymbol{\alpha}$ represents all observed effects and specifies an estimable conditional mean. It is treated as an unknown fixed parameter to be estimated. $\mathbf{Z_i}$ is assumed to be unobserved and correlated with the independent variables. $\varepsilon_{\mathrm{it}}$ is a stochastic error term, independently and identically distributed with a mean of 0 and variance σ_ε^2. The independent variable x_{it} assumed to be independent of the error $\varepsilon_{\mathrm{it}}$ for all *i* and *t*.

c. Random Effect

If the individual effect $\mathbf{Z_i}$ has no correlation with the independent variables, this model structure is known as the Random Effect Model, which is expressed in Equation (4) [8].

$$y_{it} = \mathbf{x}_{\mathrm{it}}'\,\boldsymbol{\beta} + \alpha + \eta_{\mathrm{it}} \qquad (4)$$

where

$$\eta_{\mathrm{it}} = \mathrm{u_i} + \varepsilon_{\mathrm{it}}$$
$$\alpha = \mathrm{E}[\mathbf{Z_i'\alpha}]$$
$$\mathrm{u_i} = \left[\mathbf{Z_i'\alpha} - \mathbf{E}[\mathbf{Z_i'}]\right]$$

Equation (4) introduces additional elements of the panel data regression model, where there are P independent variables, including the constant. α represents the mean of the unobserved individual effects, and u_{i} is the random effect specific to the iii-th observation. In this model, u_{i} is assumed to be independent of $\varepsilon_{\mathrm{it}}$ additionally, the independent variable x_{it} is assumed to be independent of both u_{i} and $\varepsilon_{\mathrm{it}}$.

3. Model Selection

a. Chow Test

The Chow Test is used to determine whether the Fixed Effect Model (FEM) is better than the Common Effect Model (CEM) [8]. The following is the test statistic used as shown in Equation (5).

$$F\ hitung\ \frac{(RSS_1 - RSS_2)/(K-1)}{RSS_2/(KT-K-P)} \sim F_{(\alpha,(K-1),(KT-K-P))} \qquad (5)$$

Equation (5) defines the components of the Chow Test statistic, where *K* is the number of sectors, T is the observation time period, P is the number of parameters in the fixed effects model, $\mathrm{RSS_1}$ is the residual sum of squares for the common effects model, while $\mathrm{RSS_2}$ is the residual sum of squares for the fixed effects model. If the calculated F statistic is greater than the F table value F table $F_{(\alpha,(K-1),(KT-K-P))}$ at a certain α level, then the selected model is the Fixed Effect Model (FEM).

b. Hausman Test

The Hausman Test is used to choose between the Fixed Effect Model (FEM) and the Random Effect Model (REM) [8]. Following the Wald criterion, the Hausman test statistic can be calculated using the following formula as shown in Equation (6).

$$\mathrm{W} = X_{(P)}^2 = [\mathrm{b} - \beta]'\psi^{-1}[\mathrm{b} - \beta] \qquad (6)$$

Where

$$\psi = \mathrm{Var}[\mathrm{b}] - \mathrm{Var}[\beta]$$

Equation (6) provides the parameters used in the Hausman Test, b is the parameter (without intercept) of the random effect, and β is the parameter of the fixed effect using LSDV. Var[b] represents the covariance matrix of the random effect parameters (without intercept), and Var[β] is the covariance matrix of the fixed effect parameters. if $\mathrm{W} > x_{(\alpha.P)}^2$ then the selected model is the Fixed Effect Model (FEM), where P is the number of independent variables.

4. Significance Testing

a. Simultaneous Test (F-Test)

The F-Test is conducted to examine whether the estimated regression model shows that the independent variables collectively have an influence on the dependent variable [8]. The F-test statistic is formulated as shown in Equation (7).

$$\mathrm{F\ Test} = \frac{Mean\ Square\ Regresi}{Mean\ Square\ Residual} \qquad (7)$$

Equation (7) specifies the criteria for the F-Test, indicating that the independent variables collectively have a significant influence on the dependent variable if the calculated F-value $\mathrm{F_{test}} > F_{(\alpha,(K-1),(KT-K-P))}$. here *K* is the number of cross-sectional units, T is the number of time periods, and P is the number of independent variables.

B. Machine Learning

Machine Learning is a branch of Artificial Intelligence (AI) that enables computers or systems to learn from data and make predictions or decisions without being explicitly programmed [9].

1. XGBoost

XGBoost is a boosting method based on the gradient boosting algorithm, designed to improve model accuracy by building models iteratively and reducing residual errors at each step. XGBoost is known for its speed and efficiency, as well as its ability to handle missing data. This algorithm is frequently used for prediction and classification tasks on large datasets [9].

XGBoost constructs models as a combination of multiple decision trees optimized through gradient boosting. The resulting model is a combination of trees, as shown in Equation (8).

$$f(x) = \sum_{k=1}^{K} T_k(x) \quad (8)$$

Equation (8) defines the structure of the XGBoost model, with $T_k(x)$ representing the k-th decision tree. In each iteration, XGBoost minimizes the cost function, as shown in Equation (9) for the loss function.

$$l = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(T_k) \quad (9)$$

Equation (9) specifies the components of the XGBoost loss function, where l is the loss function, and $\Omega(T_k)$ is the regularization term for complexity to prevent overfitting.

2. CatBoost

CatBoost is a boosting algorithm specifically designed to handle categorical variables more effectively, reducing the need for preprocessing categorical data. This method uses ordered boosting, which avoids target leakage during the training process, making it more accurate for categorical data [10].

CatBoost optimizes an objective function similar to other boosting algorithms. The model is defined as shown in Equation (10).

$$F(X) = \sum_{m=1}^{M} f_m(x; \theta_m) \quad (10)$$

Equation (10) defines the structure of the CatBoost model, with each tree f_m added based on error improvement. The loss function typically used for regression is Mean Squared Error (MSE), as shown in Equation (11).

$$l = \frac{1}{n} \sum_{i=1}^{n} (y_i, \hat{y}_i)^2 \quad (11)$$

Equation (11) describes how CatBoost minimizes this loss function using ordered boosting and updates the model based on the gradient.

3. AdaBoost

AdaBoost is an adaptive boosting method that combines multiple weak models (e.g., shallow decision trees) to form a strong model. Each weak model is trained sequentially, where the error weights from previous predictions are increased so that the next model focuses on difficult observations [11].

In each iteration *t*, AdaBoost produces a weak model $h_t(x)$ and calculates the error ϵ_t, as shown in Equation (12).

$$\epsilon_t \frac{\sum_{i=1}^{n} w_i^{(t)} . I(y_i \neq h_t(x_i))}{\sum_{i=1}^{n} w_i^{(t)}} \quad (12)$$

The weight of the weak model α_t is defined as shown in Equation (13).

$$\alpha_t = ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \quad (13)$$

The final model is a combination of the weak models, as shown in Equation (14).

$$H(X) = sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \quad (14)$$

4. LightGBM

LightGBM is a tree-based boosting algorithm optimized for speed and efficiency. LightGBM uses Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) techniques to reduce data complexity, making it faster on large datasets [12].

The resulting model is an ensemble of trees, similar to other boosting algorithms, as shown in Equation (15).

$$f(x) = \sum_{k=1}^{K} T_k(x) \quad (15)$$

LightGBM minimizes an objective function in the form of a loss function with added regularization, as shown in Equation (16).

$$l = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \lambda \sum_{j=1}^{P} \|w_j\|^2 \quad (16)$$

Equation (16) defines the objective function for LightGBM, where l is the loss function, λ is the regularization parameter, and w_j represents the weights.

5. Random Forest

Random Forest is an ensemble learning algorithm that combines a large number of independent decision trees trained on different subsets of data. The final result is the average (for regression) or majority vote (for classification) of each tree's predictions, thereby improving accuracy and reducing overfitting [13].

The regression model of Random Forest is the average of B decision trees, as shown in Equation (17).

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \quad (17)$$

Equation (17) defines the Random Forest regression model, where $T_b(x)$ is the prediction of the *b*-th tree for input *x*. Each tree is trained on a randomly selected data subset using the bagging technique.

C. Macroeconomic Variables

1. Bond Yield

Bond yield is the rate of return or profit earned by investors from bond investments, typically expressed as a percentage.. Bond yield is crucial for investors as it reflects potential returns and risks and can be influenced by factors such as interest rates and inflation [14].

2. Dollar Selling Rate

The dollar selling rate is the exchange rate or price set by banks or financial institutions when they sell US dollars to buyers or consumers in need of dollars. The dollar selling rate is influenced by various factors, including market demand and supply, interest rates, global economic conditions, and monetary policy [15].

D. Key Commodity Price Variables

1. Copper

Copper is a reddish metal known for its excellent electrical and thermal conductivity. Copper prices in the market are often considered an indicator of global economic health due to its high demand in the construction and manufacturing sectors [16].

2. Gold

Gold is a rare and highly valuable precious metal with a shiny yellow color. Gold price fluctuations can be influenced by factors such as inflation rates, interest rates, and global economic conditions [17].

3. Crude Oil

Crude oil is a liquid raw material extracted from natural resources and is essential for producing various energy products such as gasoline, diesel, and jet fuel. Crude oil prices are highly volatile and influenced by factors such as supply, global demand, political stability in oil-producing countries, and policy decisions from OPEC (Organization of the Petroleum Exporting Countries) [16].

4. Natural Gas

Natural gas, or earth gas, is a fossil energy source in gas form found beneath the earth's surface, often alongside crude oil. Natural gas prices are influenced by factors such as seasonal demand (e.g., higher in winter), supply conditions, and developments in energy infrastructure [16].

E. Bitcoin

Bitcoin is a form of digital currency or cryptocurrency created in 2009 by an individual or group under the pseudonym Satoshi Nakamoto. Bitcoin's value tends to be highly volatile, influenced by market demand, user adoption, and global sentiment toward digital assets [17].

III. PROPOSED METHODS

A. Dataset

This research uses secondary data obtained from the website investing.com, covering the period from September 27, 2019, to September 27, 2024, with a total of 1,216 observations for each variable. Data was collected on Friday, September 27, 2024. However, the scope of this study is limited to data from Indonesia, Malaysia, Singapore, Thailand, the Philippines, and Vietnam.

The proposed method is shown in Figure 1. It starts with data pre-processing, followed by selecting the best panel data regression model: Fixed Effects (FEM), Random Effects (REM), or Common Effects (CEM). The chosen model is used to make predictions, which are then compared with machine learning predictions to find the best approach.

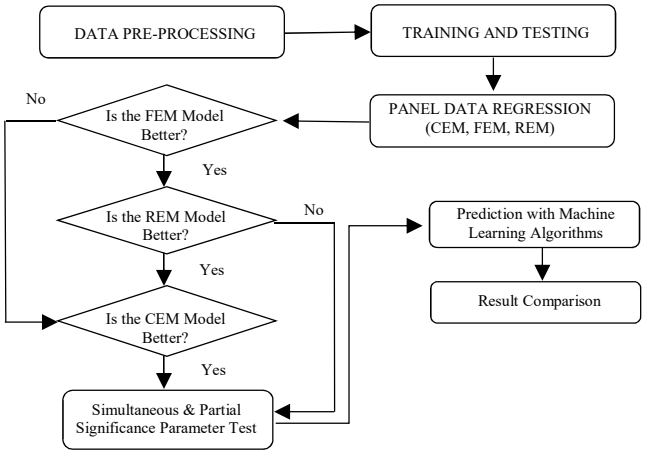

Fig 1. Proposed Methods

B. Pre-Processing

Data collected from the field often requires cleaning to address missing values and outliers before analysis. The preprocessing steps include checking for missing values, which, if found, are handled using median imputation, as it is more resilient to outliers than mean imputation. Next, outlier detection is conducted to identify extreme values that could skew the analysis or reduce model accuracy. These outliers are then managed through winsorization, a technique that adjusts extreme values while preserving as much of the dataset as possible without discarding potentially valuable data points. This structured approach ensures a cleaner, more reliable dataset for research analysis.

C. K-Fold Cross Validation

K-Fold Cross Validation is a method to validate model accuracy by splitting a dataset into multiple parts. The data is divided into k equal folds, and in each iteration, one fold serves as testing data while the remaining k-1 folds are used for training. This reduces bias and ensures that every part of the data is used for both training and testing. This study uses 10-Fold Cross Validation, where more folds improve the reliability of the results and help in detecting any overfitting. It provides an accurate measure of model performance, helping to develop robust, generalizable models that perform well on unseen data.

D. Panel Data Regression

The panel data regression analysis involves estimating with CEM, FEM, and REM models, applying Chow, Hausman, and LM tests to identify the best model, and then conducting simultaneous and partial significance tests on the research variables. These tests are essential in determining the most suitable model for analyzing panel data, ensuring it accurately reflects both individual and temporal variations. This approach leads to more reliable insights into the relationships between macroeconomic factors and stock price indices in ASEAN countries, with the significance tests further validating each variable's contribution and reinforcing the model's robustness.

E. Prediction with Machine Learning

The process of making predictions using machine learning involves several key steps. First, predictions are performed using significant variables identified in panel data analysis, utilizing algorithms like XGBoost, CatBoost, AdaBoost, LightGBM, and Random Forest. After generating predictions, the model's performance is evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) to assess accuracy and reliability. The best-performing model is then selected based on these evaluations. Finally, the prediction results are visualized by creating plots using the chosen model, allowing for a clear presentation and interpretation of the outcomes.

IV. RESULT AND DISCUSSION

A. Data Panel Regression

The panel regression estimates the relationship.

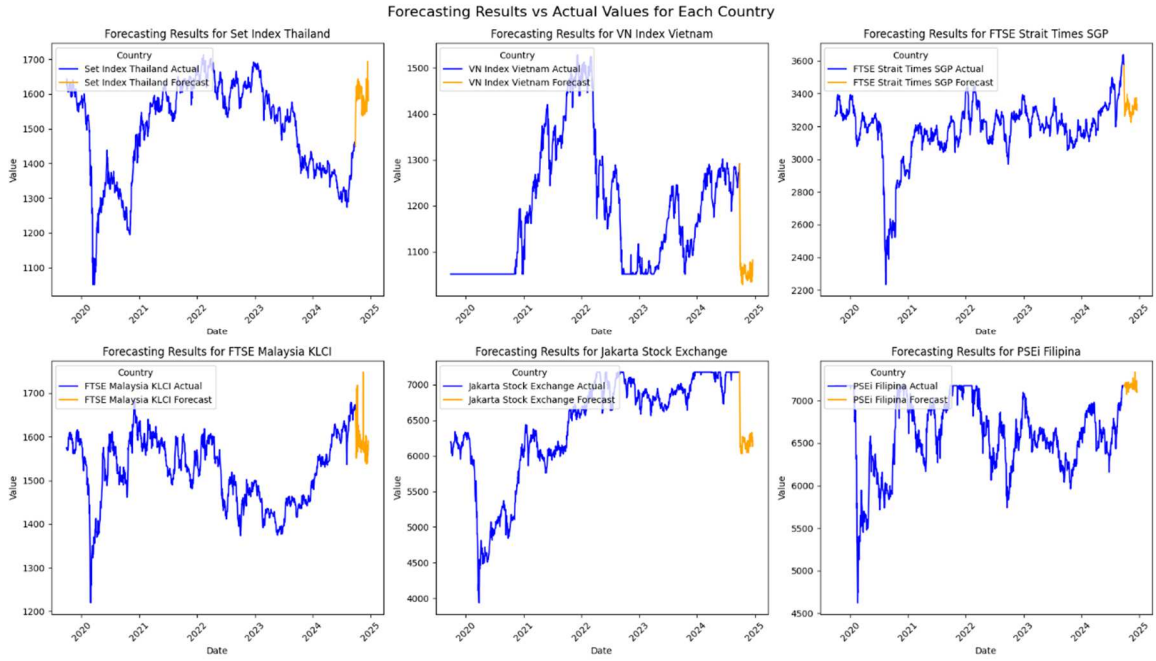

Fig 2. Forecast Plot of Major ASEAN Stock Price Index Using LightGBM

TABLE I. Chow Test

F_{Test}	$F_{0.05(5;7284)}$	p-value
23.64	2.215	0.000

TABLE II. Hausman Test

χ^2	$\chi^2_{(0.05,7)}$	p-value
11752.247	14.067	0.000

TABLE III. Simultaneous Test

F_{Test}	$F_{0.05(5;7284)}$	p-value
8.945	2.215	0.000

TABLE IV. Partial Parameter Significane Test

Variable	$\lvert t_{test} \rvert$	$t_{(0.05;6078)}$	Decision
Gold Price	61.536	1.645	Reject H_0
Crude Oil Price	11.413		Reject H_0
Bond Yield	23.834		Reject H_0
Dollar Exchange Rate	41.583		Reject H_0
Copper Rice	37.393		Reject H_0
Natural Gas	45.586		Reject H_0
Bitcoin	81.665		Reject H_0

The Chow test in Table I shows an F-Test value of 23.64, which exceeds the critical value of 2.215, indicating that the Fixed Effects Model (FEM) is preferable to the Common Effects Model (CEM).

Similarly, the Hausman test in Table II presents a χ^2 value of 23.64, surpassing the critical value of 14.067, thereby confirming FEM over the Random Effects Model (REM). Both tests validate FEM as the most suitable model for further parameter significance testing.

TABLE V. Evaluation of Model Goodness Criteria

Data	Algorithm	MAE	MSE	R^2
Training	AdaBoost	673.52	917025.14	83%
	CatBoost	82.44	15379.07	100%
	XGBoost	33.77	3160.31	100%
	LightGBM	**31.05**	**3105.61**	**100%**
	RF	27.28	7145.91	100%
Testing	AdaBoost	690.10	944755.84	82%
	CatBoost	95.55	25909.47	100%
	XGBoost	68.26	32818.04	99%
	LightGBM	**53.77**	**9786.65**	**100%**
	RF	61.85	36282.60	99%

B. Simultaneous Parameters Significance Test

The simultaneous parameter significance test was conducted to evaluate whether the predictor variables used in the study significantly impact the stock price index. This test examines the collective influence of all variables on the main stock price index for various ASEAN countries.

As demonstrated in Table III, the results from the simultaneous test show that the F-Test value exceeds the critical threshold. As a result, the null hypothesis (H0) is rejected, implying that at least one of the predictor variables has a statistically significant impact on the stock price index. This outcome highlights the relevance of the chosen predictor variables in explaining fluctuations in stock prices across ASEAN countries.

C. Partial Parameter Significance Test

The partial significance test statistics for each variable used in the study are presented in Table IV below, showing that, based on the partial parameter significance test, all variables used in the study have a significant effect on the main stock price index of various ASEAN countries.

D. Forecasting the Main Stock Price Index of Various ASEAN Countries Using Machine Learning.

The dataset is evaluated using 10-fold cross-validation, where each iteration applies an 80:20 train-test split, ensuring model reliability by testing on varied data subsets. Machine

learning algorithms are trained, and performance metrics Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) are averaged across iterations for consistency.

Table V presents a comparison of five machine learning models, with LightGBM emerging as the top performer across all metrics for both training and testing data. The selection of parameters for each model was based on prior research and empirical tuning, focusing on hyperparameters like learning rate, number of estimators, and maximum depth, which significantly influenced model performance. LightGBM's optimized parameters ensured high accuracy and efficiency in forecasting. The actual and predicted stock index values across ASEAN countries show minimal deviations, demonstrating the model's robustness and its suitability for forecasting complex financial data.

V. CONCLUSION

ASEAN stock indices mirror the region's economic health, yet complex factors like macroeconomic variables, commodity prices, and consumer sentiment shape these indices, warranting further exploration. By employing panel data regression and machine learning, this study seeks to uncover these relationships, providing a valuable framework for investors and policymakers to enhance decision-making based on a structured approach.

The analysis identifies the Fixed Effect Model (FEM) as the optimal choice for assessing the influence of key macroeconomic variables on ASEAN stock indices. Statistical tests, including the Hausman and Chow tests, confirmed FEM's suitability, indicating that each variable significantly affects stock indices across the region. Additionally, among machine learning models, LightGBM emerged as the most accurate in predicting stock index trends, highlighting its utility in financial forecasting.

Future research could deepen this analysis by including more variables, such as inflation and broader economic trends, alongside longer time series data. Such expansions would enhance insights into ASEAN market dynamics, capturing how external economic shifts impact financial performance. This refinement of predictive models would support more nuanced decision-making for both investors and policymakers.

ACKNOWLEDGMENT

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Kolaborasi Pusat and under the Project Scheme of the Publication Writing and Intellectual Property Rights (IPR) Incentive (Penulisan Publikasi dan Hak Kekayaan Intelektual/PPHKI) Program 2024.

REFERENCES

[1] A. T. Haryono, R. Sarno, and K. R. Sungkono, "Stock Price Forecasting in Indonesia Stock Exchange Using Deep Learning: A Comparative Study," International Journal of Electrical and Computer Engineering (IJECE), vol. 14, no. 1, pp. 861–869, Feb. 2024.

[2] A. T. Haryono, R. Sarno, and K. R. Sungkono, "Stock price forecasting in Indonesia stock exchange using deep learning: a comparative study," International Journal of Electrical and Computer Engineering (IJECE), vol. 14, no. 1, pp. 861-869, Feb. 2024.

[3] Hariani, S., Halim, A., Hendryadi, & Budiharjo, R. (2024). "Dynamics of ASEAN, US, and China Capital Market Relations: Before, During and Post COVID-19." Jurnal Reviu Akuntansi dan Keuangan, 14(3).

[4] Narajewski, M., Kley-Holsteg, J., & Ziel, F. (2021). "tsrobprep — an R package for robust preprocessing of time series data." *SoftwareX*, 16, 100809.

[5] Mohr, F., & van Rijn, J. N. (2023). "Fast and Informative Model Selection Using Learning Curve Cross-Validation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8), 9669-9680.

[6] N. Dziubanovska, V. Maslii, J. Novakk, and V. Rusin, "Governance Quality's Impact on FDI in EU: Comparative Analysis via Regression and Panel Models," 2024 14th Int. Conf. Adv. Comput. Inf. Technol. (ACIT), Ceske Budejovice, Czech Republic, 2024, pp. 322-325.

[7] L. Li and C. Bing, "Measure the Evolution of Interbank Network Structure: Panel Data Analysis to Shibor Network," 2014 International Conference on Management of e-Commerce and e-Government, Shanghai, China, 2014, pp. 49-53.

[8] Y. Wang, J. Tsai, and X. Lu, "Factors Driving Cross-border RMB Settlement: Panel Data Analysis of 65 Countries," 2021 5th Int. Conf. Data Sci. Bus. Anal. (ICDSBA), Changsha, China, 2021, pp. 421-427.

[9] M. A. Hasan, R. Sarno, and S. I. Sabilla, "Optimizing Machine Learning Parameters for Classifying the Sweetness of Pineapple Aroma Using Electronic Nose," 2024 International Conference on Advanced Informatics and Emerging Technologies (ICAIE), Pekanbaru, Indonesia, 2024, pp. 1-10.

[10] S. Deng, X. Huang, Y. Zhu, Z. Su, Z. Fu, and T. Shimada, "Stock index direction forecasting using an explainable eXtreme Gradient Boosting and investor sentiments," North American Journal of Economics and Finance, vol. 64, p. 101848, 2023.

[11] B. S. Rintyarna, R. Sarno, and A. L. Yuananda, "Automatic Ranking System of University Based on Technology Readiness Level Using LDA-Adaboost.MH," Universitas Muhammadiyah Jember, 2023, pp. 1–8.

[12] Y. Tan et al., "Long-Term Load Forecasting Using Feature Fusion and LightGBM," 2021 IEEE 4th Int. Conf. Power Energy Appl. (ICPEA), Busan, Korea, 2021, pp. 104-109.

[13] I. C. Suherman, R. Sarno, and Sholiq, "Implementation of Random Forest Regression for COCOMO II Effort Estimation," Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2023, pp. 1–12

[14] D. Safitri et al., "Yield Fluctuation Prediction on Indonesian Gov. Bonds Using Rough Sets and Fuzzy Rough Sets," 2023 1st Int. Conf. Adv. Eng. Technol. (ICONNIC), Kediri, Indonesia, 2023, pp. 334-338.

[15] A. B. Muharam et al., "WNFS for Predicting Rupiah-USD Exchange Rate," 2021 Int. Conf. Artif. Intell. Big Data Analyt. (ICAIBDA), Bandung, Indonesia, 2021, pp. 1-5.

[16] H. Sakamoto et al., "Polymer Hybrid Bonding for Copper-Copper at 200–250°C," 2024 Int. Conf. Electron. Packag. (ICEP), Toyama, Japan, 2024, pp. 73-74.

[17] A. Gurbaxani, "Digital Gold in Emerging Markets: An Investor's Perspective," 2023 International Conference on Sustainable Islamic Business and Finance (SIBF), Bahrain, 2023, pp. 81-84.

[18] S. Idhia and F. Herlin, "Analisis Model Indeks Harga Saham dengan Metode Regresi Data Panel," *Semirata 2017 in the Field of MIPA BKS-PTN West Region*, Ratu Convention Center, Jambi, Indonesia, 12-14 May 2017.

[19] P. K. Illa, B. P. Parvathala, and A. K. Sharma, "Stock price prediction methodology using random forest algorithm and support vector machine," *Materials Today: Proceedings*, vol. 56, part 4, pp. 1776–1782, 2022.

[20] W. L. Yan, "Stock index futures price prediction using feature selection and deep learning," *The North American Journal of Economics and Finance*, vol. 64, p. 101867, 2023.