Journal of Intelligent Decision Support System (IDSS) 7 (1) (2024), pp. 1-9

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Machine learning algorithm-based decision support system for prime bank stock trend prediction

Gunawan Gunawan[1], Wahyu Budiono[2], Wresti Andriani[3], Naella Nabila Putri Wahyuning Naja[4]

[1,2,3]STMIK YMI TEGAL, Tegal City, Indonesia
[4]Universitas Negeri Semarang, Semarang City, Indonesia

Article Info

Article history:

Received Mar 16, 2024
Revised Mar 16, 2024
Accepted Mar 21, 2024

Keywords:

Artificial Neural Network (ANN)
Machine Learning Algorithms
Predictive Analytics
Stock Trend Prediction
Support Vector Machine (SVM)

ABSTRACT

In the complex landscape of financial markets, predicting bank stock trends is a critical aspect that supports more accurate investment decision-making. This study aims to develop and evaluate machine learning algorithms—Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN)—for predicting the trends of major bank stocks in Indonesia using the IDX-PEFINDO dataset from January 1, 2020, to December 31, 2023. The adopted methodology includes collecting historical data, initial processing, feature selection, and training and validating models using evaluation metrics such as Accuracy, Precision, Recall, F1-Score, MAE, and RMSE. Results indicate that although no single algorithm is dominant, SVM and ANN perform better within the given data context. This research underscores the importance of a tailored approach to maximize the potential of machine learning algorithms in stock prediction, providing new insights into developing decision support systems for bank stock investments. This study implies that it recommends the integration of broader economic indicators and the exploration of advanced machine-learning techniques to enhance stock prediction accuracy in the future.

Corresponding Author:

Gunawan Gunawan,
Informatics Engineering,
STMIK YMI TEGAL,
#1 Pendidikan Street Tegal City, Central Java, 52142, Indonesia
Email: gunawan@stmik-tegal.ac.id

Introduction

In the intricate financial markets, the prediction of stock trends has always been a focal point of analysis and research, offering pivotal insights for investors, policymakers, and the academic community (Sheng et al., 2021). Remarkably, the banking sector, as a barometer of a nation's economic health, commands considerable attention, with its stock performance being a critical indicator of broader economic dynamics (Razzaq et al., 2022). In this context, the advent of machine learning (ML) technologies has ushered in unprecedented opportunities for enhancing the accuracy and reliability of financial predictions, thereby supporting more informed decision-making processes (Allioui & Mourdi, 2023).

The challenge at hand revolves around the volatile and unpredictable nature of bank stock prices, influenced by various factors ranging from macroeconomic indicators to market sentiment (Ellis et al., 2022). Traditional statistical methods, while valuable, often fall short of capturing the complexity and non-linearity inherent in stock market data (Deng et al., 2022). Consequently, there is a burgeoning interest in leveraging advanced machine learning algorithms adept at unraveling intricate patterns and dependencies within large datasets, thus offering more nuanced and predictive insights (Zhao et al., 2023), (Yao et al., 2023).

Despite the acknowledged potential of machine learning (ML) algorithms to revolutionize financial predictions, especially for prime bank stock trends, significant gaps persist in comparative analyses of these algorithms' effectiveness. Previous studies have highlighted the transformative impact of ML on improving prediction accuracy and decision-making within the banking sector. However, they often fall short in conducting exhaustive comparative studies across different ML strategies, such as Random Forest, Support Vector Machine (SVM), and Artificial Neural Networks (ANN), and utilizing a comprehensive array of evaluation metrics to assess their performance. This research seeks to bridge these gaps by thoroughly comparing various ML algorithms, using a broad spectrum of evaluation metrics, to identify the most efficacious models for bank stock trend prediction. The aim is to advance academic knowledge and provide practical contributions by enhancing the reliability and utility of predictive models, thereby supporting the development of robust decision-support systems crucial for financial market stability and investor confidence.

This research aims to address this gap by meticulously developing and evaluating various machine learning algorithms, specifically Random Forest, Support Vector Machine (SVM), and Artificial Neural Networks (ANN), in the context of the IDX-PEFINDO Prime Bank Stock dataset spanning from January 1, 2020, to December 31, 2023. The evaluation will focus on a comprehensive suite of metrics, including Accuracy, Precision, Recall, F1-Score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), to ascertain each algorithm's predictive prowess.

By comparing these algorithms' effectiveness, the study intends to unveil nuanced insights into their respective strengths and limitations within the ambit of bank stock trend prediction. The proposed research is not only timely but imperative, as it seeks to fortify the decision-making arsenal of investors and financial analysts with advanced, data-driven tools. The anticipated outcome is a significant stride toward refining predictive models in finance, culminating in a decision support system that marries empirical rigor with practical utility, thereby making a meaningful contribution to the field and aiding stakeholders in navigating the complexities of the stock market.

Method

This study's selection of research subjects aims to predict the trend of significant bank stocks in Indonesia using machine learning algorithms. This design allows researchers to compare the results of predictions generated by each algorithm without controlling for all variables that might affect the stock market, given the complexity and dynamic nature of stock market data.

This research design tests the effectiveness of machine learning algorithms, Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) in predicting Prime Bank share price trends. This design allows researchers to compare the results of predictions generated by each algorithm without controlling for all variables that might affect the stock market, given the complexity and dynamic nature of stock market data.

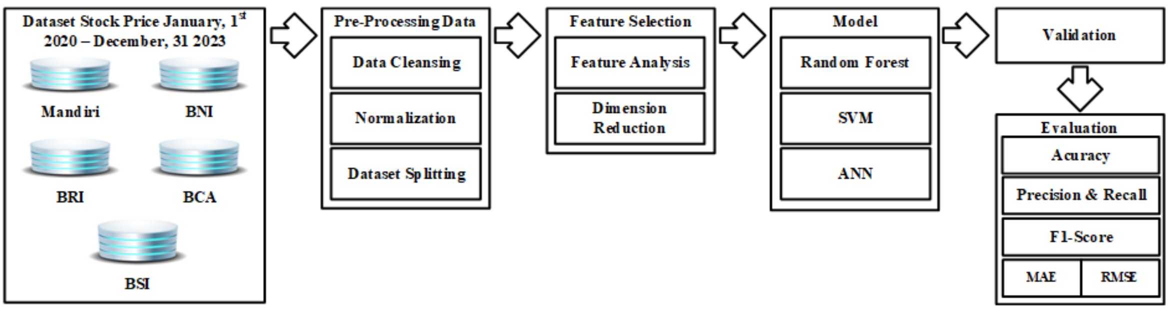

Figure 1. Research Flow

Figure 1 shows the research procedure for predicting stock price trends using machine learning algorithms, and the study began with collecting stock price datasets. Data cleansing is the first step in pre-processing to eliminate inconsistencies or missing values. Next, the data is normalized to ensure that the different features of the value scale can be analyzed relatively. The data is then divided into two sets: one for training machine learning algorithms and another for testing trained models. Feature Analysis to determine which features are most relevant to stock price trends. Dimension Reduction will be used in the model to eliminate redundancy and improve computational efficiency. Three machine learning algorithms were used: Random Forest, Support Vector Machine (SVM), and Artificial Neural Networks (ANN). Each of these models will be trained with data that has been processed. Once the model is trained, a validation process is performed to evaluate performance. The model is assessed based on evaluation metrics: Accuracy, Precision and Recall, F1-Score, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Data gathering

Researchers collected historical datasets of stocks from five central banks in Indonesia, Bank Mandiri (BMRI. JK), Bank BNI (BBNI. JK), Bank BRI (BBRI. JK), Bank BCA (BBCA. JK), and Bank Syariah Indonesia (BRIS. JK), using Yahoo Finance as a data source. The data collected includes essential information such as daily opening and closing prices, highs and lows, and daily trading volume. The period set for data collection is from January 1, 2020, to December 31, 2023, which provides sufficient data to capture various market conditions and allows the model to learn from historical trends. The collection process is done by downloading data provided by Yahoo Finance. This data is then stored in comma delivated (CSV) format.

Data pre-processing

In this study, data pre-processing begins with a data cleansing step, which involves identifying and filling in missing values and correcting or removing outliers that may affect the analysis (Oleghe, 2020). The missing values are filled in using statistical techniques, using the mean or median (Emmanuel et al., 2021). Outlier correction is done by checking the data against historical parameters and known market trends to ensure the consistency and reliability of the data used (Ramos et al., 2020).

Next, data normalization is implemented to ensure all variables have a similar scale (Yang et al., 2020). After normalization, the dataset is divided into a training set to develop a model and a test set to evaluate model performance (Majkowska et al., 2020). Data sharing is done to guarantee that the model will be tested, avoid overfitting and ensure the validity of the resulting predictions (Salazar et al., 2022).

Feature selection

Feature selection is done to identify the most influential features of stock price trends (Shen & Shafiq, 2020). First, a thorough feature analysis is carried out, where each feature available in the dataset, namely the opening, closing price, high, low, and trading volume, is evaluated to determine its effect on stock price movements. After feature identification, the next step is dimensionality reduction so that the analysis focuses only on relevant features and eliminates features that do not provide additional information.

Proposed models

In this study, the proposed model or algorithm for stock price trend prediction involves using three machine learning algorithms: Random Forest, Support Vector Machine (SVM), and Artificial Neural Networks (ANN).

Random Forest

Random Forest is an ensemble algorithm consisting of many decision trees trained with different pieces of data (Dhivyaa et al., 2020). The final prediction is based on the majority of votes from all trees (Kim & Upneja, 2021). The illustration for Random Forest can be described as a collection of many decision trees that each provide predictions, and the final result is chosen by the trees the most (Zhou et al., 2020). The basic equation for Random Forest is not as simple as other algorithms because it involves many decision trees (Jun, 2021). However, the equation for prediction from a single decision tree can be represented as:

$$Y = f(X) + \epsilon \quad \text{...(1)}$$

Where $\boldsymbol{Y}$ is the target variable, $\boldsymbol{X}$ is an input feature, $\boldsymbol{f(X)}$ Does the decision tree provide the function, and ϵ is an error or residue.

Support Vector Machine (SVM)

SVM is an algorithm that looks for a maximum-margined hyperplane separating data classes in a feature space (Taneja et al., 2023). SVM illustration is a line (in two dimensions) or plane (in three or more dimensions) that divides the data space so that the classes of each data point are separated as far as possible (Thudumu et al., 2020). The basic equation for SVM (in an elementary form) is:

$$w \cdot x + b = 0 \qquad \text{.....(2)}$$

Where $\boldsymbol{w}$ is a weight vector, $\boldsymbol{x}$ is the input vector, and b is the bias. $\boldsymbol{w} \cdot \boldsymbol{x}$ is the product of the point between the weight vector and the input, and $\boldsymbol{b}$ is the offset of origin.

Artificial Neural Network (ANN)

ANN is a network of interconnected nodes or neurons, where each connection has an adjustable weight during training (Kumar et al., 2020). In this study, the ANN layer consists of an input, hidden, and output layer. The equation for a single neuron in an ANN is:

$$y = \phi\left(\sum_{i=1}^{n} w_i x_i + b\right) \qquad \text{.....(3)}$$

Where $\boldsymbol{y}$ is the output of neurons, $\boldsymbol{\phi}$ is an activation function (such as sigmoid or ReLU), $\boldsymbol{w_i}$ Is the weight of the ith input, $\boldsymbol{x_i}$ It is the value of the ith input, and $\boldsymbol{b}$ is the neuron's bias.

Validasi dan Evaluasi Model

Model validation involves using datasets not used during model training to test the model, performed with cross-validation techniques (Yates et al., 2023). After validation, model performance is evaluated using these evaluation metrics involving the calculation of Accuracy, Precision, Recall, F1-Score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Accuracy is the percentage of overall correct predictions:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Population} \qquad \text{.....(4)}$$

Precision adalah proporsi prediksi positif yang sebenarnya benar:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad \text{.....(5)}$$

Recall yaitu proporsi positif aktual yang diidentifikasi dengan benar:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad \text{.....(6)}$$

F1-Score yaitu harmonic mean dari Precision dan Recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad \text{.....(7)}$$

Mean Absolute Error (MAE) is the average of the absolute error between the prediction and the actual value:

$$MAE = \frac{\sum|y_i - \hat{y}_i|}{n} \quad \text{..........(8)}$$

Where y_i is the real value and $\hat{y}_i$ Is the predicted value. Root Mean Square Error (RMSE) is the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} \quad \text{..........(9)}$$

Results and Discussions

Based on the Bank's primary stock data provided, this study analyzes the stock performance of PT Bank Mandiri (Persero) Tbk (BMRI. JK), PT Bank BNI (Persero) Tbk (BBNI. JK), PT Bank BRI (Persero) Tbk (BBRI. JK), PT Bank Central Asia Tbk (BBCA. JK), and PT Bank Syariah Indonesia Tbk (BRIS. JK) uses Random Forest machine learning algorithms, Support Vector Machine, and Artificial Neural Network. The analysis focuses on each stock's closing price over a specified period.

Table 1. IDX-PEFINDO Prime Bank Stock

Date	BMRI.JK	BBNI.JK	BBRI.JK	BBCA.JK	BRIS.JK
2/23/2024	7050	5900	6125	9825	2450
2/22/2024	7100	5950	6250	9875	2460
2/21/2024	7250	6000	6300	9975	2440
2/20/2024	7150	6025	6300	10025	2480
2/19/2024	7150	5875	6100	9875	2380
2/16/2024	7200	6000	6150	9950	2330
⋮	⋮	⋮	⋮	⋮	⋮
2/28/2023	5000	4387.5	4670	8750	1520
2/27/2023	5075	4475	4810	8775	1520
2/24/2023	5062.5	4450	4760	8675	1630

The data was divided into two sets to perform this analysis: 80% for training and 20% for testing. Those machine learning models are trained using training sets to learn patterns in a stock's historical data. Then, the model's effectiveness in predicting stock prices is tested using a test set.

The model evaluation uses predefined metrics: accuracy, precision, recall, F1-Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The performance of each model was compared to determine the most effective algorithm in predicting IDX-PEFINDO Prime Bank's share price trend.

Early results show that each algorithm has its strengths and weaknesses in the context of this stock data. For instance, the Artificial Neural Network might excel in capturing the complexity of nonlinear stock price patterns, whereas Random Forest and Support Vector Machine might be more effective in specific scenarios, depending on the data characteristics. This finding aligns with previous research, such as (Ali et al., 2021), which highlighted the accuracy of Support Vector Machine (SVM) in stock price prediction using historical financial data, and (Nabipour et al., 2020) who suggested that a combination of algorithms, including Random Forest and Neural Networks, could enhance stock prediction accuracy. Our study contributes to the scientific community by directly comparing the performance of Random Forest, SVM, and ANN in predicting prime bank stock trends, employing a comprehensive suite of evaluation metrics. The results confirm the significance of selecting the appropriate algorithm based on specific data characteristics and demonstrate variability in model performance, echoing the findings of (Nabipour et al., 2020) and (Ali et al., 2021). Furthermore, our research underscores the necessity of integrating external data, such as economic indicators and market news, to improve the accuracy of ML models in stock prediction, offering new insights for developing more effective ML-based decision support systems.

Table 2. Application of Random Forest, SVM, and ANN algorithms to IDX-PEFINDO Prime Bank stock dataset

Evaluation Metrix	Random Forest	SVM	ANN

Accuracy	35.42%	47.92%	43.75%
Precision	29.63%	40.00%	33.33%
Recall	40.00%	50.00%	35.00%
F1-Score	34.04%	44.44%	34.15%
MAE	0.6458	0.5208	0.5625
RMSE	0.8036	0.7217	0.7500

Table 2 shows that among the three algorithms tested, SVM showed better performance in terms of accuracy and precision, while SVM and ANN achieved the best recall and showed moderate performance among the three models. However, these results may vary depending on feature selection, model parameters, and validation strategies.

Additional data such as economic indicators, market news, and others are needed to determine which bank stocks are the best. This analysis will be limited by the information available. We will use the daily price change as each stock's predicted target (up/down).

Here is the Pseudocode for analysis using the SVM model for each share: (1) Prepare dataset with targets as 'up' or 'down' from day to day, (2) Split dataset into training and test sets, (3) Apply and train SVM model on the training set, (4) Prediction on the test set, (5) Evaluate model with accuracy, precision, recall, F1-Score, MAE and RMSE metrics.

Here is a model evaluation using Random Forest, SVM and ANN algorithms for stock movement prediction using Accuracy, Precision, Recall, F1-Score, MAE and RMSE evaluation metrics:

Table 3. Comparison of Evaluation Results of the Application of the Random Forest Algorithm

Bank	Accuracy	Precision	Recall	F1-Score	MAE	RMSE
BMRI.JK	43.75%	33.33%	35.00%	34.15%	0.56250	0.75000
BBNI.JK	54.17%	36.36%	50.00%	42.11%	0.45833	0.67700
BBRI.JK	54.17%	63.16%	44.44%	52.17%	0.45833	0.67700
BBCA.JK	35.42%	25.00%	23.81%	24.39%	0.64583	0.80364
BRIS.JK	37.50%	45.45%	35.71%	40.00%	0.62500	0.79057

Based on Table 3, an analysis of the evaluation results shows that Bank BNI and Bank BRI performed relatively better than other banks, with higher accuracy and an F1 score. Bank BRI stands out with the highest precision (63.16%), indicating that the prediction is likely correct when the model predicts an increase in stock prices. Bank BCA and Bank Syariah Indonesia have lower accuracy and F1-Score than other banks, with Bank BCA showing the lowest performance in all metrics. This suggests that the Random Forest model is less effective at predicting stock movements for these two banks. Precision and recall vary among banks, with some banks showing a higher tendency for precision or recall. This suggests a difference in how the model can identify actual stock price increases instead of avoiding false positives. The MAE and RMSE show the prediction error rate, with Bank BCA and Bank Syariah Indonesia having higher RMSEs, signaling a more significant prediction error in general.

Table 4. Comparison of Evaluation Results of the Application of Support Vector Machine (SVM) Algorithm

Bank	Accuracy	Precision	Recall	F1-Score	MAE	RMSE
BMRI.JK	47.92%	40.00%	50.00%	44.44%	0.52083	0.72169
BBNI.JK	50.00%	31.82%	43.75%	36.84%	0.50000	0.70711
BBRI.JK	54.17%	64.71%	40.74%	50.00%	0.45833	0.67700
BBCA.JK	56.25%	0.00%	0.00%	0.00%	0.43750	0.66144
BRIS.JK	41.67%	0.00%	0.00%	0.00%	0.58333	0.76376

Based on Table 4, Analysis of the evaluation results that, Bank BRI shows the highest precision (64.71%), which indicates that when the model predicts an increase in stock price, the prediction is very likely to be correct, although the recall is lower indicating the model is less able to identify all actual stock price increases. Bank BCA and Bank Syariah Indonesia showed 0% precision and recall, indicating that the model failed to predict the actual stock increase for these two banks. Bank Mandiri and Bank BNI showed moderate performance, with Bank BNI having slightly higher accuracy but lower precision than Bank Mandiri. F1's low scores for BBCA and BRIS indicate a poor balance between precision and recall, indicating that the model is ineffective for both stocks. The MAE and RMSE show the prediction error rate, with Bank Syariah Indonesia having the highest RMSE, signaling a more significant prediction error in general.

Table 5. Comparison of Evaluation Results of the Application of the ANN Algorithm

Bank	Accuracy	Precision	Recall	F1-Score	MAE	RMSE
BMRI.JK	58.33%	50.00%	65.00%	56.52%	0.41667	0.64550
BBNI.JK	54.17%	36.36%	50.00%	42.11%	0.45833	0.67700
BBRI.JK	62.50%	66.67%	66.67%	66.67%	0.37500	0.61237
BBCA.JK	39.58%	30.00%	28.57%	29.27%	0.60417	0.77728
BRIS.JK	54.17%	61.54%	57.14%	59.26%	0.45833	0.67700

Based on Table 5, an analysis of the evaluation results shows that Bank BRI showed the best performance among the five banks, with the highest score in almost all evaluation metrics, indicating the effectiveness of the ANN model in predicting the movement of its shares. Bank BCA has the lowest performance with the lowest accuracy and metric values, indicating that the ANN model is less suitable for predicting Bank BCA's stock movements than other banks. Bank Mandiri and Bank Syariah Indonesia showed competitive results, with Bank Syariah Indonesia standing out in precision and F1-Score. The overall performance of ANN models varies between banks, which shows the importance of selecting the right features and adjusting the model for each use case.

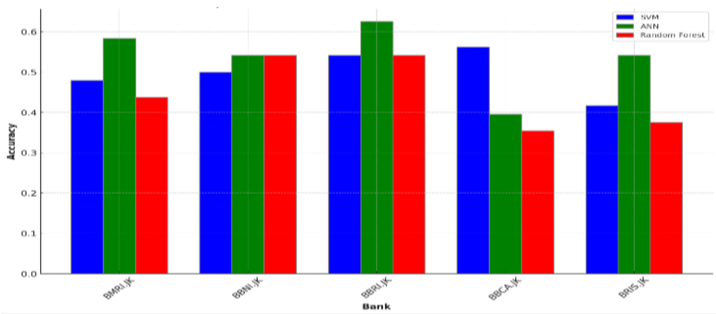

Figure 2. Comparison of Model Accuracy Across Different Banks

Figure 2 shows a comparison graph of the accuracy of three different algorithm models (Support Vector Machine - SVM, Artificial Neural Networks - ANN, and Random Forest - RF) in predicting stock movements for five different banks: Bank Mandiri (BMRI. JK), Bank BNI (BBNI. JK), Bank BRI (BBRI. JK), Bank BCA (BBCA. JK), and Bank Syariah Indonesia (BRIS. JK).

From Figure 2, several key conclusions can be drawn, and there is significant variability in accuracy performance among the three models for each Bank, indicating that no single model is consistently superior in all cases. This confirms the importance of conducting model exploration of various algorithms to find the best fit for a specific dataset. In general, Bank BRI appears to have higher prediction accuracy than other banks when using ANN and RF models, indicating that data for Bank BRI may have patterns that are easier to learn by these models. Bank BCA showed suboptimal performance in SVM and RF models but did not see significant improvement with ANN models. This indicates that Bank BCA's stock movement prediction may be more challenging than other banks. The effectiveness of the ANN Model shows significant accuracy improvements for some banks (such as BMRI. JK and BBRI. JK) compared to SVM and RF models, demonstrating the power of ANN in capturing the complexity of nonlinear data. Prediction difficulties for BCA and BRIS banks are present as these banks show relatively lower accuracy across all models, indicating that external factors or data characteristics may make prediction more difficult.

Conclusions

The conclusion of this research indicates that while the Support Vector Machine (SVM) algorithm exhibits superior accuracy and precision compared to Random Forest and Artificial Neural Network (ANN) in predicting the trend of prime bank stocks, each algorithm has its strengths and weaknesses influenced by the data characteristics. The primary limitation of this study lies in its reliance on a limited dataset without incorporating external economic indicators or market news, which could impact stock price dynamics. Future research should integrate these additional data and explore other ML algorithms that may offer new perspectives in stock prediction. This study is expected to provide theoretical implications by expanding the understanding of the effectiveness of various ML algorithms in the financial context and practically, as well as assist investors and financial analysts in making more informed decisions by developing more sophisticated and market-responsive decision support systems.

References

Ali, M., Khan, D. M., Aamir, M., Ali, A., & Ahmad, Z. (2021). Predicting the direction movement of financial time series using artificial neural network and support vector machine. *Complexity*, *2021*, 1–13.

Allioui, H., & Mourdi, Y. (2023). Exploring the full potentials of IoT for better financial growth and stability: A comprehensive survey. *Sensors*, *23*(19), 8015.

Deng, C., Huang, Y., Hasan, N., & Bao, Y. (2022). Multi-step-ahead stock price index forecasting using long short-term memory model with multivariate empirical mode decomposition. *Information Sciences*, *607*, 297–321.

Dhivyaa, C. R., Sangeetha, K., Balamurugan, M., Amaran, S., Vetriselvi, T., & Johnpaul, P. (2020). Skin lesion classification using decision trees and random forest algorithms. *Journal of Ambient Intelligence and Humanized Computing*, 1–13.

Ellis, S., Sharma, S., & Brzeszczyński, J. (2022). Systemic risk measures and regulatory challenges. *Journal of Financial Stability*, *61*, 100960.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*, 1–37.

Jun, M.-J. (2021). A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: The case of the Seoul metropolitan area. *International Journal of Geographical Information Science*, *35*(11), 2149–2167.

Kim, S. Y., & Upneja, A. (2021). Majority voting ensemble with a decision trees for business failure prediction during economic downturns. *Journal of Innovation & Knowledge*, *6*(2), 112–123.

Kumar, P. S., Behera, H. S., Kumari, A., Nayak, J., & Naik, B. (2020). Advancement from neural networks to deep learning in software effort estimation: Perspective of two decades. *Computer Science Review*, *38*, 100288.

Majkowska, A., Mittal, S., Steiner, D. F., Reicher, J. J., McKinney, S. M., Duggan, G. E., Eswaran, K., Cameron Chen, P.-H., Liu, Y., & Kalidindi, S. R. (2020). Chest radiograph interpretation with deep learning models: assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology*, *294*(2), 421–431.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data, a comparative analysis. *Ieee Access*, *8*, 150199–150212.

Oleghe, O. (2020). A predictive noise correction methodology for manufacturing process datasets. *Journal of Big Data*, *7*(1), 89.

Ramos, D., Faria, P., Vale, Z., Mourinho, J., & Correia, R. (2020). Industrial Facility Electricity Consumption Forecast Using Artificial Neural Networks and Incremental Learning. In *Energies* (Vol. 13, Issue 18). https://doi.org/10.3390/en13184774

Razzaq, A., Sharif, A., An, H., & Aloui, C. (2022). Testing the directional predictability between carbon trading and sectoral stocks in China: New insights using cross-quantilogram and rolling window causality approaches. *Technological Forecasting and Social Change*, *182*, 121846.

Salazar, J. J., Garland, L., Ochoa, J., & Pyrcz, M. J. (2022). Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. *Journal of Petroleum Science and Engineering*, *209*, 109885.

Shen, J., & Shafiq, M. O. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*, *7*, 1–33.

Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2021). COVID-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions. *British Journal of Management*, *32*(4), 1164–1183.

Taneja, K., Arora, V., & Verma, K. (2023). Classifying the heart sound signals using textural-based features for an efficient decision support system. *Expert Systems*, *40*(6), e13246.

Thudumu, S., Branch, P., Jin, J., & Singh, J. (Jack). (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, *7*(1), 42. https://doi.org/10.1186/s40537-020-00320-x

Yang, D., Alessandrini, S., Antonanzas, J., Antonanzas-Torres, F., Badescu, V., Beyer, H. G., Blaga, R., Boland, J., Bright,

J. M., & Coimbra, C. F. M. (2020). Verification of deterministic solar forecasts. *Solar Energy*, *210*, 20–37.

Yao, Z., Wang, H., Yan, W., Wang, Z., Zhang, W., Wang, Z., & Zhang, G. (2023). Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images. *European Journal of Radiology*, 110934.

Yates, L. A., Aandahl, Z., Richards, S. A., & Brook, B. W. (2023). Cross validation for model selection: a review with examples from ecology. *Ecological Monographs*, *93*(1), e1557.

Zhao, J., Han, X., Ouyang, M., & Burke, A. F. (2023). Specialized deep neural networks for battery health prognostics: Opportunities and challenges. *Journal of Energy Chemistry*.

Zhou, X., Lu, P., Zheng, Z., Tolliver, D., & Keramati, A. (2020). Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety*, *200*, 106931.