INTI JOURNAL | Vol.2025, Issue 3, No.4
eISSN:2600-7320

Stock Price Prediction of Bank Rakyat Indonesia Using an Ensemble Stacking Model of K-Nearest Neighbors (KNN) and Support Vector Machine (SVM)

Muhammad Tamami[1*], Sefto Pratama[2], Zaenuddin[3], Haldi Budiman[4], Erfan Karyadiputra[5], Desy Ika Puspitasari[6]

[1,2,3,4,5.6] Faculty of Information Technology Islamic University of Kalimantan Muhammad Arsyad Al-Banjari, Indonesia

Email: muhammadtamamitkj@gmail.com[1*], seftoprtama.bjm@gmail.com[2], zaenuddin.uniska@gmail.com[3], haldibudiman@gmail.com[4], erfantsy@gmail.com[5], smile4desyka@gmail.com[6]

Abstract

Need for future price forecasts by investors faces difficulties in achieving accurate predictions because market changes exist. Standard single models do not accurately model stock market behaviors because of their complex nature. The problem solution implemented by the study involves combining K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) to create ensemble stacking. Research personnel collected Bank Rakyat Indonesia's (BRI) historical stock price data using KNN and SVM models. Studio performance delivers superior predictive results with lower error rates than KNN and SVM models that operate individually. Study results demonstrate stacking technology produces the most desirable results for stock market price prediction.

Keywords

Stock Prediction, Machine Learning, Ensemble Stacking, KNN, SVM

Introduction

The constant fluctuations of stock market prices make the development of trustworthy market forecasts extremely difficult. Numerous forecasting approaches struggle to identify advanced price patterns that emerge in market conditions. Pattern detection receives effective solutions through machine learning techniques as Fama (1970) and Zhang et al. (2020) demonstrate.

The research examines Bank Rakyat Indonesia's (BRI) stock price history available on the Kaggle database for analysis. A set of normalization and feature selection techniques prepares the data for modeling applications by Han, Kamber, and Pei (2011). KNN algorithms track historical trends,

Submission: 26 May 2025; **Acceptance:** 2 September 2025; **Available online:** September 2025

so SVM provides generalization by determining optimal decision boundaries as described in Cortes and Vapnik (1995).

The study integrates KNN and SVM through stacking ensemble methods because the models provide different advantages in prediction. The algorithm generates predictive results with higher accuracy levels according to research findings (Wolpert, 1992).

The implementation of several predictive models leads to better stock price forecasting results based on Dietterich (2000) and Zhou (2012). The stacking ensemble technique should be employed to integrate KNN and SVM models because these models possess different advantages that enhance predictive accuracy.

Methodology

KNN (K-Nearest Neighbors)

The K-Nearest Neighbors (KNN) algorithm implements non-parametric logic whereby it assigns unclassified data points to the predominant class found among their nearest adjacent points. The measurement techniques for the distance between KNN data points include Euclidean distance and other metrics that enhance pattern recognition in financial markets. The KNN algorithm reacts to noisy data while its performance quality decreases when dealing with extensive datasets as noted by Altman (1992).

SVM (Support Vector Machine)

Support Vector Machine (SVM) functions as a classification tool that detects the most optimal hyperplane to divide different categories of data. Assessable data through kernel functions enables SVM to successfully process complex datasets consisting of high dimensions. The algorithm of SVM reaches its best predictive accuracy while minimizing overfitting by finding optimal class margins (Cortes & Vapnik, 1995).

Stacking Ensemble Model

The methodology of stacking uses two or more basic models to generate combined predictions which boost the overall predictive precision. The researchers unite KNN and SVM while adding a meta-model to refine their combined output. The approach employs both algorithms effectively to overcome individual weaknesses which produces upgraded stock price forecasting outcomes (Wolpert, 1992).

Data Preprocessing

The research analyzes stock pricing data of Bank Rakyat Indonesia (BRI) available in Kaggle's database. Data preprocessing requires proper execution because it leads to both better model results and improved data quality. The steps involved include:

- A data cleaning process removes missing values because it upholds data integrity as mentioned in Rahm and Do (2000).
- Min-max normalization transforms all features into values between 0 to 1 because it maintains equal variable participation throughout the training process (Han, Kamber, & Pei, 2011).

- Feature Selection keeps essential features including Open, High, Low, Close, and Volume by following a domain-based assessment (Guyon & Elisseeff, 2003).

1	Date	Open	High	Low	Close	Adj Close	Volume	label_harga
2	1/1/2019	3327.21533	3327.22	3327.22	3327.22	2510.63	0	1
3	1/2/2019	3281.76148	3309.03	3263.58	3281.76	2476.33	82441033	0
4	1/3/2019	3254.48926	3309.03	3254.49	3290.85	2483.19	102805905	0
5	1/4/2019	3290.8523	3327.22	3281.76	3327.22	2510.63	125907734	0
6	1/7/2019	3363.57837	3372.67	3327.22	3327.22	2510.63	82106628	0

Figure 1. Original Dataset

Target Variable: The Close price of the next trading day is selected as the prediction target to model future prices. Normalization is applied because it helps to eliminate scale bias among features, accelerates convergence during training, and improves the model's ability to learn patterns from the data efficiently. Below is the normalized dataset.

Table 1. Normalized Dataset

Date	Open	High	Low	Close	Adj Close	Volume	label_harga
1/1/2019	0.332722	0.332722	0.332722	0.332722	0.251063	0	1
1/2/2019	0.328176	0.330903	0.326358	0.328176	0.247633	8244.103	0
1/3/2019	0.325449	0.330903	0.325449	0.329085	0.248319	10280.59	0
1/4/2019	0.329085	0.332722	0.328176	0.332722	0.251063	12590.77	0
1/7/2019	0.336358	0.337267	0.332722	0.332722	0.251063	8210.663	0

Result and Discussion

Table 3. Comparison of results

Algorithm	Splitting Dataset	Accuracy
KNN	80% Train, 20% Test	0.5883
SVM	80% Train, 20% Test	0.9120
Stacking KNN + SVM	**80% Train, 20% Test**	**0.9630**

Applied to our ensemble stacking model these two methods provide KNN-based local trend sensitivity and SVM-based generalization ability which leads to higher predictive accuracy.

The research analyzes stock prediction through machine learning algorithms and historical stock market records. The analysis includes the implementation and comparison of K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and an ensemble stacking model as predictive models. The performance levels of these models differ from one another thus enabling users to generate meaningful conclusions about financial prediction methods.

The ensemble stacking model proves to be the most effective method with 96.30% accuracy based on the research outcomes. The combination of SVM and KNN elements within the model produces enhanced predictive accuracy while minimizing errors. The process of market fluctuation adaptation makes this methodology a dependable method for predicting stock prices.

The accuracy reached by SVM models reached 91.20% which placed the technique in second position. The identification of patterns in historical stock price information by this system functions successfully yet its performance is negatively affected by both noise and the selection of parameters. SVM brings value to price predictions that show consistent patterns that link past data points.

Between the examined models KNN presents the most inaccurate results by producing 58.83% accuracy. The stoc k price prediction abilities of KNN suffer due to its inability to manage complex financial data together with highly volatile market conditions. The data set requirements for KNN utilization include situations when data points can be divided linearly.

The application of ensemble stacking models with machine learning processes leads to higher stock price prediction accuracy. KNN and SVM integration into stacking models enhances both the prediction accuracy along the model's reliability performance. The ensemble stacking model shows superior financial forecasting capacity through its 96.30% accuracy level that yields improved investment decisions.

KNN and SVM show subpar performance compared to stacking yet they hold worth in financial market scenarios that align with their capabilities. KNN detects market patterns effectively but SVM works best when dealing with datasets whose data points form distinct linear relationships. The present research uses stock price historical data as its only limitation although future analyses should incorporate market sentiment data and macroeconomic data for improved forecast results. The accurate prediction of stock prices becomes possible with deep learning models as they represent sophisticated machine learning approaches.

Conclusion

The ensemble stacking model outperformed individual algorithms, achieving 96.30% accuracy compared to 58.83% for KNN and 91.20% for SVM. This confirms its effectiveness in handling complex stock market data and providing reliable forecasts. The model can serve as a valuable tool for investors and analysts in improving decision-making and risk management. Future work may include larger datasets, macroeconomic variables, and deep learning approaches to further enhance performance.

Recommendation

The research should expand its dataset by adding financial indicators like macroeconomic factors market sentiment and worldwide economic trends to increase the predictive power of the models. The examination of ensemble learning techniques should be expanded through research efforts that evaluate multiple diverse base learners and the introduction of deep learning solutions into ensemble strategies. Model interpretation for financial analysts along with time-sensitive data integration would allow the development of dynamic decision support platforms.

Acknowledgement

The authors would like to express their gratitude to the Faculty of Computer Science and Information Technology (FTI), Universitas Islam Kalimantan Muhammad Arsyad Al-Banjari (UNISKA), for the support provided in this research.

References

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175–185. https://doi.org/10.1080/00031305.1992.10475879
Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
Dietterich, T. G. (2000). Ensemble methods in machine learning. In Lecture Notes in Computer Science: Multiple Classifier Systems (Vol. 1857, pp. 1–15). Springer. https://doi.org/10.1007/3-540-45014-9_1
Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417. https://doi.org/10.2307/2325486
Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157–1182. Retrieved from https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf
Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann. https://doi.org/10.1016/C2009-0-61819-5
Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Engineering Bulletin, 23(4), 3–13. Retrieved from https://dbs.uni-leipzig.de/files/research/publications/2000-1/pdf/TBDE2000.pdf
Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1
Zhang, X., Hu, Y., & Zhang, B. (2020). Stock market prediction based on deep learning: A review. Financial Innovation, 6(1), Article 1. https://doi.org/10.1186/s40854-020-0174-6
Zhou, Z.-H. (2012). Ensemble Methods: Foundations and Algorithms. Chapman & Hall/CRC. https://doi.org/10.1201/b12207