Jurnal Informasi dan Teknologi

https://jidt.org/jidt

2024 Vol. 6 No. 2 Hal: 23-28 e-ISSN: 2714-9730

The Application of Neural Network Models to Explain the Relationship Between Stock Value, Returns Value, and Information on Prices

Ni Kadek Dessy Hariyanti[1]✉, Iwan Henri Kusnadi[2], Rezky Nurbakti[3], Enny Diah Astuti[4], Rahmat Taufik R.L Bau[5]

[1]Politeknik Negeri Bali
[2]Universitas Subang
[3]Institut Ilmu Sosial Dan Bisnis Andi Sapada
[4]Politeknik LP3I Jakarta
[5]Universitas Negeri Gorontalo

dessyhariyanti@pnb.ac.id

Abstract

This study aims to identify features for stock market value predictions based on time series grouping. This research uses various data sets for different activities. We have the balance sheet data set, which includes a series of quarterly balance sheets. Next, there is the ratio data set. The clustering data set consists of a series of daily prices. There are two data sets for testing activities: pilot forecasts and investments with daily data, projections, and investments. To speed up the development and exploration of different methods, testing begins with a subset of the data, with additional shares of witnesses. This research uses ARIMA and artificial neural networks to predict stock prices. Predicted results provide an investment strategy that compares the results obtained with what happens, resulting in generally positive profits. The capacity of artificial neural networks to manage non-linearities in financial data explains this. Second, many experiments demonstrate that not all data can be utilized using predicting techniques. These results indicate that improved estimations do not always result from more significant data. Forecast accuracy and performance on the test set will likely be negatively impacted by noise and competing signals introduced by data from unrelated stocks. Put another way, by including unrelated data, the information about the action supplied by the group of which it is a member is obscured. As such, the forecasts more closely align with the market's average behavior than with the performance of a specific stock.

Keywords: Stock, Financial Data, Neural Networks, Profits.

1. Introduction

Putting data from different stock exchanges and the companies listed on them together to get more information for price prediction models or market movements has emerged as an intriguing study area. Some additional sources of information used by researchers include company fundamental data, quarterly financial reports, forum opinions, and company news [1]. The first grouping is often based on indicators produced by risk rating agencies. In recent years, many studies have used static data clustering approaches to search for correlations or additional information, using static data from balance sheets to group companies and examining stock price growth in both groups. Some researchers use a different approach, modifying the balancing characteristic vectors and grouping them to find brief variations in stock prices when balance sheets are released. Time series data clustering is still in its infancy, especially for multivariate data series, but static data clustering has been thoroughly studied [2].

Nonetheless, several studies have demonstrated that time series clustering plays a significant role in trend and price prediction. Additionally, researchers have presented techniques such as the extended Frobenius norm to cluster multivariate time series. These techniques produce better results than conventional methods by reducing dimensionality using principal component analysis and comparing time series using the Frobenius norm [3].

Whether using time series or static data, similar scenarios in trend or stock price prediction models frequently refer to current balance sheet data contrasted with observed outcomes in other similar organizations [3]. Different models are applied, ranging from machine learning-based techniques like decision trees, neural networks, and evolutionary algorithms to univariate and multivariate statistical methods like VARIMA. Because of the volatility of the stock market, many studies have been done on applying neural networks for stock price and trend prediction [4]. Novel methods, such as generative adversarial networks, neural networks with convolutional and recurrent layers, or using genetic algorithms to enhance predictions, have been suggested by

Receipt: 10-11-2023 | Revision: 10-03-2024 | Publish: 06-04-2024 | doi: 10.60083/jidt.v6i2.525

certain researchers [5]. The question of whether statistical techniques yield more accurate forecasts than neural networks is up for dispute. Neural networks are deemed superior by specific experts, particularly regarding applications such as predicting bitcoin prices. However, this opinion is not absolute, and research still shows the superiority of statistical methods in some contexts [6].

Combining clustering with various prediction methods has become an exciting area of research. Some studies suggest performing clustering first to prepare the data before feeding it into a prediction model. Prior work proposed employing self-organizing maps to cluster action time series data, which were then put into a genetic network to produce good prediction accuracy [7]. However, other research suggested clustering price time series using k-means and morphological similarity distance to identify several series types and train distinct neural networks for each type. Some writers employ clustering with distance-based techniques like dynamic time warping or time series properties like stability, entropy, autocorrelation coefficients from the fast Fourier transform, or coefficients from the continuous wavelet transform to generate features [8]. A neural network is then given this data along with the time series to enhance predictions.

Furthermore, studies have been conducted to detect and classify investment structures using public stock information and market transaction records and to predict market trends using a combination of sentiment and market data [9]. The findings demonstrate how grouping factors might enhance artificial neural network predictions. Although price-based time series grouping has been applied extensively, time series based on quarterly financial reports have not been as frequently grouped. Consequently, we can expand on this method by using more representative distance measures for multivariate time series [10].

The stock market is a vital center of economic activity, dominated by two large stock exchanges: the NYSE and Nasdaq. The NYSE, with a history spanning over a hundred years, lists more than 2,000 stocks with a market value of more than 30 billion dollars in 2023. Meanwhile, the Nasdaq lists more than 3,000 stocks with a market value of more than 20 trillion dollars in the same year [11]. These two stock exchanges list companies from all over the world, representing a sizeable global action market. The stock market is dynamic and attractive to investors because of its diversity and liquidity. Investors can transact through various instruments such as shares, options, indices, and staked securities [12]. To list its shares, a company must formally communicate relevant facts, including publishing a quarterly balance sheet containing accounting and financial information about its condition. The stock market usually operates on weekdays, but little activity is outside the main market hours. Share prices vary over time in response to supply and demand and intrinsic and external factors that influence listed companies [13].

The stock market offers several basic activities for investors to do to earn profits. One is long-buy, where investors buy an instrument hoping the price will rise in the future [14]. Meanwhile, short selling allows investors to rent an instrument at a specific daily price for resale by betting on a decline in the instrument's price. To carry out both types of transactions, investors can choose from various operations [15]. For example, transactions can be carried out at the market price, which is the market value when the transaction is completed without a price limit, or use a limit to limit the transaction price. In addition, investors can use Stop to execute transactions when the instrument reaches a specific price or trailing Stop to set price limits based on changes in the instrument's price [16]. Using various types of transactions and methods is an integral part of investment strategies for professional and amateur investors to minimize losses or maximize profits.

2. Research Methods

In general, historical price data can be obtained from a variety of widely available sources. In this research, the finance service is used via the Python finance library because it provides a simple way to download data. It must be noted that the downloaded price data has been adjusted for stock division or consolidation, making it easier for users to adjust it manually. Meanwhile, access to quarterly reports is generally not easy. The latest quarterly reports are freely accessible on some sites, but historical data often requires a paid service. In this study, Alpha Vantage services were used through their public API because they offer the possibility of obtaining data at a low cost. In this research, various data sets are used for different activities. First, we have the Balance Sheet data set, which includes a series of quarterly balance sheets. Next, there is the Ratio data set. The clustering data set consists of a series of daily prices. Next are two data sets for testing activities: pilot forecasts and investments with daily data and projections and investments. To speed up the development and exploration of different methods, testing begins with a subset of the data, with additional shares of witnesses. In this research, we use artificial neural networks to predict stock prices. As a result, we generated subsets of all data for additional testing. After the testing phase was completed and with algorithm improvements, testing was carried out on the total data obtained from all data. This is a systematic and comprehensive approach to combining various data sets in the study.

3. Results and Discussion

This study evaluated prediction quality using two general metrics. We use precision to measure the accuracy of predictions in transaction classification, specifically the correlation between the number of correctly classified transactions and the total number of transactions. We use average performance to gauge an investment fund's overall return, indicating the effectiveness of the predictions in generating investment returns. We also measure prediction accuracy by observing the actual execution of predicted transactions, such as purchases or sales. We employ mean absolute scaled error data and symmetric mean absolute percentage error to quantify the degree of prediction error. The evaluation results indicate that the precision is often substantially less than random prediction, except for the multiple intraday models for entry prices, and is not significantly more significant. The naive approach, which forecasts the following day based on the price or performance of the previous day, is substantially different from what the sMAPE and MASE numbers show. This divergence is only corrected for prices using several intraday models.

This research evaluates prediction quality by back-testing investment recommendations, such as buy and sell orders and comparing them to profitability. The results demonstrate that while the accuracy of the prediction results is not perfect, we cannot accurately predict market movements as a whole, nor are they completely random. This shows that although there are factors that can help in making predictions, such as using mathematical models or complex algorithms, there are also other factors outside our control that can influence market movements. For example, political news, economics, or global events can significantly impact the market, which is difficult to predict precisely using traditional mathematical models. However, although market predictions are not always accurate, this research provides valuable insights for investors and other market participants to understand the complexity and uncertainty of making investment decisions. Enhanced comprehension of the market's influencing aspects enables investors to devise more effective risk management and profit-maximizing tactics.

According to empirical simulations, all forecasting techniques outperform all reference techniques except ARIMA intraday price multiples. This means that predicting pricing with additional data will probably work better than only using the data provided for each activity. Except for ARIMA estimation with prices, the results are inferior to other approaches when stock data is used in the sample. These results highlight the importance of additional information in improving the quality of stock price predictions. Further information, such as financial ratios or company fundamentals, can provide better insight into future stock price behavior. However, it should be noted that using additional information can also increase the complexity of the model and require the processing of more extensive data, which can affect overall performance.

Furthermore, the results show that using stock data in the sample can lead to less accurate predictions. The model's inability to generalize to new data may be caused by bias in the past data used to train it. Therefore, it's crucial to consider the data sources carefully while creating stock price forecasting models. These empirical simulations provide valuable insights for researchers and practitioners in developing better stock price forecasting models. Considering the use of relevant additional information and carefully selecting the data sources used can improve the quality of stock price predictions and help investors make better investment decisions.

Furthermore, artificial neural networks' prediction capacity generally exceeds ARIMA's. This may be due to the neural network's ability to handle non-linearity and extract additional information. Furthermore, the results of price prediction generally outperform those of yield prediction. We obtain a tiny average performance with a low standard deviation around the average performance by aggregating the outcomes of all conducted transactions. Out of all the techniques, the random strategy has the lowest standard deviation and an average that approaches 0%. These results provide an opportunity to optimize algorithms and strategies that can process company data more efficiently and test these strategies in most stock markets. These results confirm the findings at the exploratory stage but with a volume that makes it possible to verify many. This demonstrates that the proposed approach has the potential to be a highly accurate method for stock price prediction, making it a valuable tool for investors and stock market analysts.

In the series of financial ratios, we find suitable values for dynamic time warping and extended Frobenius norm. The correspondence matrix between the two clustering methods shows a different structure, with an adjusted Rand measure value of 0.03 between the two clusterings. This indicates no high correlation between the two, as seen in the matrix. In the case of DTW on price series, we take weekly samples to reduce the search time for silhouette coefficients that indicate adequate coefficient values and the number of groups. In comparison, the entire process would take almost three weeks to complete. Fourier execution lacks clarity, as it finds the maxima of interest at low group values. We can test different points of view by examining the correspondence matrix between the two clustering methods, which also shows the selection of varying circuit characteristics. Despite the groups being more concentrated than in previous cases, the observed behavior aligned with the trial findings. We calculate the adjusted Rand measure for these two sets and find a value of 0.3, higher than the set of financial ratios.

After getting these four groupings, we move on to the next task, namely final value prediction. We ran seven models in ARIMA and another six in artificial neural networks, representing simple cases with their action data and multiple with each group's action data with closing prices only for ARIMA, given equipment availability and computational costs. One of the problems we faced in running this model was the time required for each run, which ranged from four to six days. In addition to the PC we used in the tests, this forced us to use three virtual machines with CPUs to run ARIMA in parallel and three with TPUs for models based on neural networks, both on Google Cloud and finally via one Research Cloud. This way, we can reduce the execution time of all models to less than ten days. Considering that the sMAPE and MASE values for the artificial neural network are lower than those obtained, and for the artificial neural network even lower than those obtained, the fit achieved with the artificial neural network is better than ARIMA, in contrast to what was observed in the trials. The possible reason for this behavior is that the more exogenous data is detrimental, the more ARIMA cannot process it well, whereas models with neural networks have a greater capacity to process such data efficiently.

With these predictions, we re-ran the developed investment algorithm and obtained results that again improved to no better than those obtained by chance. Again, we can observe a slight bias in favor of neural networks and more transactions executed in neural networks. However, this quarter's results were similar to those of the neural network trial, indicating a decrease and improvements in the application. In both cases, these results remain positive compared to other traditional methods. The test observed a very similar earnings behavior, capitalizing on the stock market's sharp decline following the war's start to generate additional profits. The pilot detected no significant difference in average performance between ARIMA and artificial neural networks. Another intriguing effect is the acquisition of artificial neural network characteristics with better returns in the medium-value company segment, but not to the extreme. This may be because, in the case of large companies, there is a lot of information available, thereby reducing the capacity of these methods to exploit different associations or data sources efficiently, while in the small company segment, they have a considerable sensitivity to unforeseen events. It can be captured or modeled adequately.

In ARIMA, this effect is not very pronounced, with comparable performance in the mid-and low-value segments declining in the high-value segment. We also performed a separate analysis of opens by size and industry type, obtaining the following results for artificial neural networks and buy-and-hold comparisons: for neuron networks, we noticed that the effects of recorded energy losses were more concentrated in the medium to large segments, while in the other segments, gains were good or very good for technology, communications, and consumer discretionary. However, the application maintains a more balanced approach and does not highlight energy issues; instead, it acknowledges that the segment recorded the best forecast. Compare this to the exceptional performance in buy and hold, which achieved energy and loss returns in most of the remaining industries, except communications, where a small segment experienced significant improvement. Finally, the performance histograms of artificial neural networks and ARIMA will be examined. We see that they adjust roughly to a standard curve, although in the case of this method, the results reach a peak.

These findings have important implications for modeling and prediction. Too much information can be counterproductive, suggesting that in some situations, selecting relevant information and focusing on a particular action can be more beneficial than using all available information. Additionally, forecasting models, especially neural networks, tend to be better at generating correct signals for higher-yielding transactions but are wrong on low-yielding days. Neural networks are also more likely to drive operations than ARIMA models, indicating that investment strategies driven by neural networks tend to find more opportunities than strategies driven by the method. These findings demonstrate the importance of selecting the correct information and focus in forecasting modeling and that more complex approaches do not always produce better results.

4. Conclusion

Predicted results provide an investment strategy that compares the results obtained with what happens, resulting in generally positive profits. This research supports two critical hypotheses. First, deep learning techniques, such as artificial neural networks, are usually superior, especially in price, to traditional models, as seen in the trial period. Artificial neural networks perform better than ARIMA. The capacity of artificial neural networks to manage non-linearities in financial data explains this. Second, many experiments demonstrate that not all data can be utilized using predicting techniques. Compared to using unique population data, the aggregation of stock information increases the power of estimations and enables enormous profits by providing pertinent information that supplements the individual data of a given company. Here, prices (or returns) combined with data from quarterly accounting reporting offers insight into the projection model. However, gathering data from financial reports or the price or performance of each stock in the sample will impair the artificial neural network's ability to predict the future. These results demonstrate that improved estimations do not always result from more significant data. Forecast accuracy and performance on the test set will likely be negatively impacted by noise and competing signals introduced by data from unrelated stocks. Put another way, by including unrelated data,

the information about the action supplied by the group of which it is a member is obscured. As such, the forecasts more closely align with the market's average behavior than with the performance of a specific stock.
These findings provide valuable insight into the importance of quality over quantity in forecasting. Using relevant related data can produce better results than relying on all available data. The clustering approach also proved to be a valuable pre-processing step in uncovering hidden information. Furthermore, this research validates the idea of adaptive markets, where investors can profit from identifying temporal deviations in the behavior of similar stocks. However, as mentioned, this study has certain limitations in the selection of data and sample used, which may limit the generalizability of the findings. Therefore, future research could expand the testing period and consider different sectors and company sizes to evaluate the method's sensitivity. In addition, integrating alternative data, such as sentiment analysis or macroeconomic data, can increase the accuracy of predictions. Plans to develop improved ARIMA or neural network models and real-time applications may also open up new opportunities in market analysis and prediction. Overall, your research contributes to understanding the importance of selecting the correct data and appropriate approaches in market analysis.

References

[1] Permatasari, P. D., & Agustina Widodo. (2023). Comparative Analysis Of Stock Liquiditi Level, Price Share And Stock Returns Before And After Stock Split On The Compass 100 Index Year 2017-2021. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *9*(3), 774–782. https://doi.org/10.35870/jemsi.v9i3.1165

[2] I. N. G. A. Astawa, I. G. N. B. Caturbawa, E. Rudiastari, M. L. Radhitya and N. K. D. Hariyanti, "Convolutional Neural Network Method Implementation for License Plate Recognition in Android," *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, Makassar, Indonesia, 2018, pp. 176-179, doi: 10.1109/EIConCIT.2018.8878658.

[3] Tannady, H., Nuzulya Pahlawi, Moch Arif Hernawan, Deddy Novie Citra Arta, & Sri Dewi Yusuf. (2023). Role of Stock Performance as an Intervening Variable in a Relationship Between Profitability, Leverage, Growth and Company Value. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *9*(2), 220–225. https://doi.org/10.35870/jemsi.v9i2.1016

[4] R. T. R. L. Bau, A. Hermila, I. Farman, L. Muhammad Hidayat and S. Salim, "AI Perspectives in Education: A BERT-based Exploration of Informatics Students' Attitudes to ChatGPT," *2023 9th International Conference on Education and Technology (ICET)*, Malang, Indonesia, 2023, pp. 36-41, doi: 10.1109/ICET59790.2023.10435040.

[5] Amrullah, A. A., Tantoni, A., Hamdani, N., Bau, R. T. R., Ahsan, M. R., & Utami, E. (2016). Reviewatas Analisis Sentimen Pada Twitter Sebagai Representasi Opini Publik Terhadap Bakal Calon Pemimpin.

[6] Rokhmawati, D., Desman Serius Nazara, Sutrisno, Muhammad Irvan, & Deddy Novie Citra Arta. (2023). The Role of Corporate Governance as a Moderating Variable in Relationship of Determinant Factors Stock Return. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *9*(1), 86–92. https://doi.org/10.35870/jemsi.v9i1.917

[7] N. K. D. Hariyanti, I. K. Pasek and N. W. Sumetri, "Technology Acceptance Analysis Of Planning And Budgeting System Of Bali State Polytechnic," *2018 International Conference on Applied Science and Technology (iCAST)*, Manado, Indonesia, 2018, pp. 622-626, doi: 10.1109/iCAST1.2018.8751602.

[8] Pramudito, D. K., Sumarni, A. T., Astuti, E. D., & Aditi, B. (2023). The Influence of User Trust and Experience On User Satisfaction Of E-Commerce Applications During Transactions in Mini Markets Using Delon and McLean Method. *Jurnal Sistim Informasi dan Teknologi*, 1-7.

[9] Supri, B., Rudianto, Abdurohim, Badriatul Mawadah, & Helmi Ali. (2023). Asian Stock Index Price Prediction Analysis Using Comparison of Split Data Training and Data Testing. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *9*(4), 1403–1408. https://doi.org/10.35870/jemsi.v9i4.1339

[10] Hermiyetti, H., Rina Apriliani, Muhammad Hery Santoso, Murwani Wulansari, & Agoes Hadi Poernomo. (2023). The Role of Good Corporate Governance as a Moderating Variable in Relationship Between Solvency, Company Size, Liquidity and Stock Price. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *9*(4), 1247–1253. https://doi.org/10.35870/jemsi.v9i4.1285

[11] Munir, M., Hamzah, H., Hakim, L., Kusnadi, I. H., & Magdalena, M. (2023). The Effectiveness Analysis of Student Creativity Increasing Program Using Grimm's Rainbow Media. *Journal on Education*, *6*(1), 5973-5981.

[12] Nurbakti, R., Putri, L. G., Kurniawan, D. R., Judijanto, L., & Arifin, M. S. (2023). Analysis of The Influence of Digital Promotions, Perception of Ease, Transparency and Service Features Toward Public Intention to Use Digital Bank in Money Saving. *Jurnal Sistim Informasi dan Teknologi*, 79-84.

[13] Sudirjo, F., Nurbakti, R., Nugroho, I. H., Ekaputra, A., & Utami, E. Y. (2023). Analysis of The Influence of Technology Perceived Quality, Digital Advertising and Technology Digital Experiental Marketing on

Customers Intention to Purchase at Marketplace Platform Tokopedia. *Jurnal Informasi Dan Teknologi*, 285-290.

[14] Lidyah, R., Maya Indriani Yacob Impak, Dede Rustaman, Loso Judijanto, & Musran Munizu. (2023). Analysis Of Discretionary Accrual, Differences In Profits, Cash Flow And Fiscal Profits On Profits Persistence Of National Manufacturing Companies Listed In Indonesian Stock Exchange. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *9*(6), 2510–2514. https://doi.org/10.35870/jemsi.v9i6.1660

[15] Kusnadi, I. H. (2021, January). Tourism destination development strategy ready-to-visit post pandemi Covid-19 in Sariater Subang Jawa Barat. In *Proceeding The First International Conference on Government Education Management and Tourism* (Vol. 1, No. 1, pp. 222-235).

[16] Liow, F. E. R., Imam Hanafi, Loso Judijanto, Tanti Widia Nurdiani, & Musran Munizu. (2023). Analysis Of The Influence Of Financial, Environmental Performance And Corporate Governance On Performance Of Corporate Social Responsibility Of Energy And Mineral Companies Listed In Indonesian Stock Exchange. *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *9*(6), 2621–2626. https://doi.org/10.35870/jemsi.v9i6.1691