Comprehensive Analysis of Stock Price Dynamics Using Ensemble Machine Learning Models for Enhanced Prediction Accuracy

Vinamra Nayak[1,*]
[1]Associate Professor, GGITS, Jabalpur, M.P., India
Emails: vinamranayak@ggits.org

Abstract

Stock price prediction is an important component of the financial analysis because the results influence the increase in economic growth and investment. This work aims to develop an ensemble SL technique that consists of mainly PCA, PSO, and SVM to achieve better prediction. Hence, through PCA, large numbers of stocked data dimensions are compressed without compromising on the crucial feature of data set. The problem of parameter selection for non-linear datasets is handled by using a bio-inspired optimization technique known as PSO in order to optimize the SVM hyperparameters. As the core accurate predictor model, the SVM employs the Radial Basis Function to provide the substantial regression capacity for sophisticated financial data sets. The ensemble framework was used with actual stock price data and the information set into training and testing sets. The acknowledgement of probable manifold values indicated that the proposed approach is more accurate than conventional approaches, with an accuracy rate of 95.5 %, when benchmarked using RMSE or MAE. In particular, the forecasts of stock prices by integrating PCA for feature reduction and PSO for parameter tuning with SVM regression is a notable improvement. The proposed methodology can be easily applied to scale for financial analytics since it manages to solve for the issues of noisy and non-linear high dimensional data.

Keywords: PCA; PSO; SVM; Stock Price Prediction; Financial Data Analysis; Regression Models. Machine Learning in Finance; Stock Market Trends; Dimensionality Reduction

1. Introduction

Predicting stock prices has latterly remained a universally popular issue in the subfields of investment and computer science. It is important for investors, analyst and policy maker to have an efficient tool through which informed decisions from trading strategies to risk management and market trends can be made. However, [1] the volatility and nonlinearly of the financial markets makes stock price prediction inherently difficult. Those factors affecting price levels are numerous, starting with macroeconomic parameters, and including firm-specific information, mood on the market and even geopolitics.

One of the important challenges to face when dealing with stock price prediction is the fact that one needs to capture the complex [2] interactions between all these factors while at the same time the data is normally very noisy and high dimensional. Of course, traditional statistical techniques for Time Series analysis can be useful for some of these jobs but cannot capture such complications. For instance, ARIMA and simple linear regression models use stochastically stationary and linearity, which might not be the case on the [3] financial markets. These limitations have opened way for better and advanced methods based on the concept of machine learning (ML) and computational intelligence.

Artificial intelligence has changed the way that we can think of stock price prediction. Since they are able to tap into [4] algorithms that are designed to read data and find patters, ML models are more accurate and timelier. In contrast to classical statistical models, it is possible to note that in the ML models' work, there is no need to

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

make specific assumptions about the distribution of the sample underlying data, which provides a high level of flexibility in real [5] financial markets, characterized by significant nonlinearity.

The variety of Machine Learning approaches makes it possible to use different method to the problem of stock price prediction. Neural networks, decision trees, SVM are the supervised learning reproductions applied for historical information to build up the relation with the future price movement. At the [6] same time, some of the useful methods of un-supervision learning are clustering and techniques for dimensionality reduction. Some of the methods used in reinforcement learning approaches also starting to adopt especially when the goal is to have an optimum of decisions over time like in Algorithmic [7] trading.

However, since standalone ML models can have limitations especially when working with a large number of features. [8] Most often such datasets are characterized by numerous variables, such as historical prices, volume, technical indicators, and sentiment analysis. These features may be some or all of them, and may contain either duplicate or irrelevant information that can only confuse the model and thus lead to reduced precision. In addition, the MLA [9] is very sensitive to the configuration of the parameters and it takes time to set the best parameter that can give the best result.

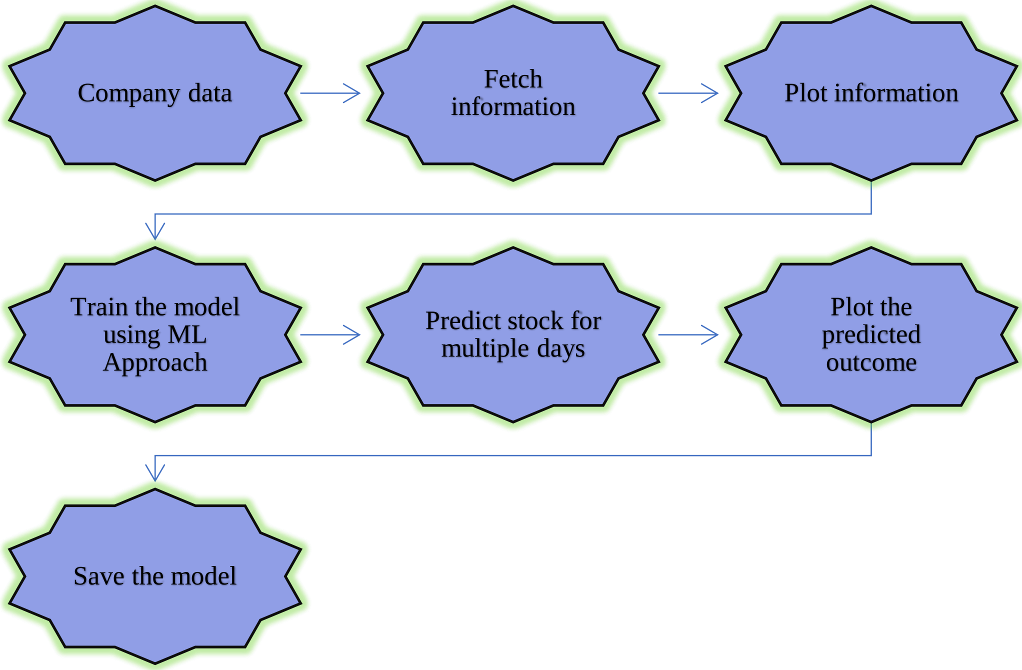

Figure 1. Illustrating the dynamics of stock price estimate using machine learning

The unpredictable nature of financial data introduces several challenges that make stock price prediction an intricate task:

- Volatility and Nonlinearity: Functions in financial markets are far from stable due to many factors which cause high fluctuations and nonlinearity, such as political events, economic policies, and overall sentiment among investors.
- High Dimensionality: Many times, the financial datasets have a large number of variables including stock price, trading volume, macroeconomic factors, and social mood. It also means that the kind of feature space we need to work with such [10] datasets have to involve reducing dimensionality underpinning the analysis.
- Noisy Data: This research shows that financial information is time varying, noisy, contains outliers and possibly contains missing values and these are elements that affect the superiority of the data negatively and hence the routine of the developed predictive replicas.
- Overfitting: One of the issues linked to a large number of attributes is over fitting of the machine learning models which means that apart [11] from learning about the data set on which it is trained, the model has higher probability of not being able to learn about the data set on which it will be tested.
- Real-Time Prediction: Models work in high-frequency trading where data prediction must be done in real-time which makes the process exact and fast.

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

To overcome these challenges, the advanced methodologies include data preprocessing, feature selection and optimization methods. [12] For example, operational intelligence techniques like principal component analysis (PCA) simplifies the dataset by only selecting the most important attributes. Model tuning involves optimizing all model parameters, where there are known techniques like GA and PSO that tune all the model parameters, making sure machine learning model is both linearly accurate and efficient.

Furthermore, other research has also demonstrated that combining two or several machine learning methods can slightly enhance the predictive performance. These methods employ the spirit of an ensemble where the strengths [13] of these models are harnessed while at the same time, minimizing the weaknesses of the models. For instance, using regression models together with an optimization approach can solve the issues such as parameter sensitivity and noisy information.

It should be renowned that the problem of information pre-processing takes an [14] especially important place in contemporary approaches to stock price prediction. Normalization and scaling help to overcome the problem of the fact that features with large ranges dominate the model, while missing value imputation helps overcome the absence of data.

2. Related Work

The forecasting of stock prices is a complex and complex process because of the fluctuation and the non linearity characteristics of stock prices. In the years past, various methodologies have being employed by researchers to enhance the [15] efficiency of the forecasting methodologies they employed, starting from simple statistical models to the advanced, more complex; machine learning. Furthermore, the analysis of related work primarily shows the innovation in the model development area, especially the identification of the potential for integrating heuristics and optimization-based approaches in model development to yield better-performing models.

Conventional Statistical Methods

The basic approach in forecasting stock prices remain ARIMA, ARCH, and [16] GARCH. These techniques aims at disaggregating time series data to enable one identify various trends. There is the reason why these models are useful for portraying linear problems, but when dealing with non-linear and dynamic financial problems, these models do not offer a workable solution in very volatile markets. This shortcoming prompted scholars to seek other techniques that could be responsive to the messiness that obtains in and around financiers.

Switch to machine learning models

After constraints with statistical models were realized, the claim of ML in forecasting of the stock market began. ANN, SVM, and BPNN were some of the more [17] effective strategies for capturing the incidental relationships within the financial datasets due to the non-linear nature of these relationship.

Artificial Neural Networks (ANNs): ANNs mimic the operational of the brain and are hence capable of discerning heretofore unrecognized patterns in [18] financial records. The earlier work showed that they could enhance anticipation of stocks' time trends by efficiently managing non-linear interactions. However, they culmed into large datasets and heavy computation, which presented problems for them.

Emergence of Hybrid Models

The weaknesses of isolated machine learning models led researchers to the use of multi-technique models that would tap on different strengths. [19] Combined models have introduced extra methodologies to stock price prediction and this has transformed the area.

- ANN and Genetic Algorithms (GA): Another study adopted a class of hybrid ANNs and GAs to enhance the selection of the features to increase the efficiency of the stock index prediction. These models were helpful in enriching the [20] learning process by filtering out the attributes and getting rid of noise.
- SVM and Optimization Algorithms: Integration with optimization algorithms which includes PSO, ACO, and TLBO has proved very effective. These approaches responded to the question of selecting parameters in the context of supporting the performance of the models based [21] on the SVM strategy.
- Dimensionality Reduction and Regression Models: Analysis of high dimensional financial data presents some uniqueness in the process of training predictive models. Other techniques include PCA, KPCA and

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

ICA has been incorporated with regression models depending on SVM and SVR to improve [22] the concert of the models by compressing the feature space.

- Self-Organizing Feature Maps (SOFM) and SVR: The usage of SOFM in conjunction with SVR has demonstrated potential while being used in financial forecast. The input data is categorized and sorted by the clustering method SOFM The regression function is accomplished by SVR. [23] This configuration has been known to provide the best training speed and prediction rates.

Numerous studies have compared the performance of different approaches to understand their strengths and limitations:

- Performance Metrics: Accuracy of the models has been known to be measured by means of MAPE, RMSE and Correlation Coefficients.
- Hybrid Models vs. Standalone Methods: It is found out that the hybrid models perform restored than the individual techniques with favour to accuracy and stability. For example, it is established that SVM combined with PSO resulted in improved performance as compared to SVM only in various applications such as parameter [24] optimization.
- Feature Selection: The accuracy of the models that make use of superior feature selection approach such as SOFM and PCA is high because these approaches select only relevant features while ignoring others considered irrelevant.

This paper also identifies research gaps and provides opportunities of studies in the following:
Despite significant advancements, [25] several challenges remain in the field of stock price prediction:

- Real-Time Adaptability: Most traditional models are not suitable for real-time adaptation because of their inadequate performance in complex financial markets.
- Cross-Market Applicability: There is scope to learn in a broader manner as to how these models can be generalized across different markets.
- Optimal Hybridization: Where we still find working in this area challenging is to determine which combination of the machine learning method, dimensionality reduction, and optimization method is optimal or suitable for the specific financial datasets.

This has also been evident in the transition from the application of purely statistical techniques to machine learning, and a combination of both techniques in the study of stock price prediction. It is worth noticing that researchers introduced the most complicated algorithms, as well as taking advantage of different techniques, the forecast accuracy was significantly increased. The further development of this field is dedicated to the management of existing gaps, in terms of real-time data processing, interdisciplinary applications of this approach and the optimization of the hybrid model. Such developments promise to deliver better, stronger, and more certain instruments aimed at understanding the intricacies of the world of finance.

3. Objective of the Research

The research objectives of the thesis are addressed by strategies presented in

- Enhanced Predictive Accuracy: To improve the current work on stock price prediction consuming powerful ML techniques and optimization algorithms that will improve accuracy and dependability of the prototypical in the unpredictable stock market.
- Dimensionality Reduction: To make the information processing quicker and easier by applying the procedure of Principal Component Analysis (PCA) that decreases data dimensionality, but preserves crucial data.
- Optimization of Model Parameters: To enhance PSO utilization in tuning the SVM parameters seeking to optimize performance for different and non-linear datasets more efficiently

4. Motivation

The fluctuation and uncertainties in financial markets require highly sophisticated stock prices prediction methods. Conventional approaches cannot properly express non-linear behaviours and high-dimensional space. This research seeks to fill these gaps, as it utilises both state of the art ML and optimization procedures to produce accurate and relevant information for focal stakeholders in financial sectors.

5. Proposed Work

The proposed framework incorporates the process of PCA, PSO, and SVM to complement increased stock price predictability. It is important because PCA is a technique that simplifies a dataset by first eliminating most of

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

the noise and then reducing the number of features that are important to the analysis. SVM has been adopted in this research due to its effectiveness when processing linearly indivisible data sets, since PSO, a bio-inspired optimization algorithm, further optimizes the SVM hyperparameters for enhanced performance. SVM acts as the algorithm used in the designs on the website since it can model nonlinear data, using a kernel function to process financial datasets. This combination copes with the problems like high dimensionality of the data, noise, and parameter sensitivity, making the proposed solution effective for accurate stock prices prediction.

5.1 Principal Component Analysis (PCA)

PCA is a technique that forms a significant technique of reducing large sets of sample points and variable quantity while preserving the maximum available relevant information. It helps in converting a dataset into new coordinate system in which first axis is called as principal component which enlightens the largest modification, second axis is the principal component that explains the second largest alteration and so on. In PCA, one seeks linear amalgamations of the unique features and forms new components which are orthogonal, and that carries the maximum variance. This makes the computations easier and brings about visualization and easy analysis on high dimensional data. PCA is commonly used in data pre-processing to remove noise, and many other preprocessing techniques that improve the effectiveness of machine learning algorithms.

Standardize the Dataset: Normalize the data where the values are adjusted to fit mean and variance to zero and one respectively to increase the model's efficiency.

$$D = \frac{O-\mu}{\sigma} \qquad (1)$$

Covariance Matrix: Estimate the co- variance matrix in order to understand correlations between features.

$$\Sigma = \frac{1}{m} D^T D \qquad (2)$$

Eigen Decomposition: Find out the eigenvalues: λ of the covariance matrix and their corresponding eigenvectors: u. Each component is characterized by its variance, eigenvalues, and the direction these components point to, shown by eigenvectors.

$$\sum u = \lambda u \qquad (3)$$

Feature Selection: Select a number k eigenvalues and their respective eigenvectors bigger than others and keep just k eigenvectors, then they are principal components.

Data Transformation: So, we can project the information on a fresh area with smaller dimensions:

$$B = DU_k \qquad (4)$$

Here D = standardized dataset, O = original information, μ = mean, σ = standard deviation, U_k = matrix loop of top k eigenvector.

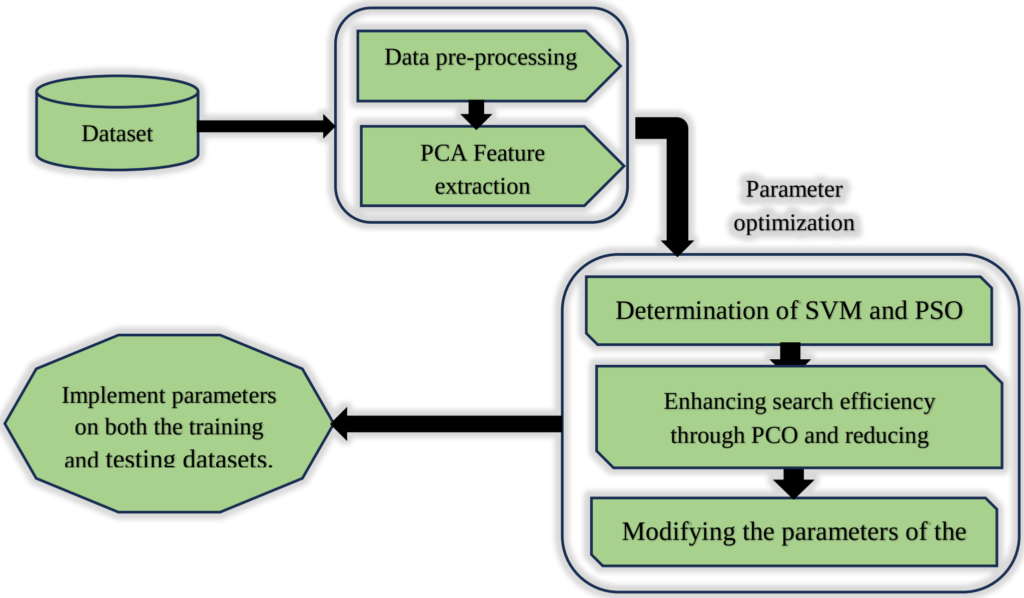

Figure 2. An integrated method utilizing PCA, PSO and SVM

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

The suggested framework includes three elements: We will also implement three state-of-the-art algorithms namely PCA, SVM, and PSO. The approach involves, using PCA to extract important features from the dataset, the use of SVM for predictions, and the use of PSO for tuning of SVM parameters. The model performance equation is influenced strongly by the kernel type, the regularization parameter, ε-insensitive loss used in SVM. The given model uses RBF as its kernel function because works based on the availability of nonlinear information sets. For regression analysis, PCA helps in the solution, and features are taken to the SVM where PSO optimizes.

5.2 Particle Swarm Optimization (PSO)

PSO is a heuristic optimization technique like where a host of particles in the search space swarm as do birds or fish. In PSO, a population of candidate solutions named particles search for a better solution in the search space by updating position constructed on own involvement and also on the knowledge of neighborhood particles. Every particle has a velocity calculated in parallel with personal best and global best positions of the swarm. PSO, especially unique variant, is suitable for solving diverse nonlinear and large optimization tasks or those with multi-dimensional search space and no need for gradient evaluation.

Initialization: Start the process with a swarm of particles that are randomly positioned in the hyperparameter space; and randomly initialized velocities ai and ui.

Fitness Evaluation: Select that particle from the population which can receives a higher fitness value using a fitness function. For SVM, this can be the Root Mean Square Error (RMSE):

Update Personal Best (pi) and Global Best (g): Every single particle then adjusts its current position (pi) to the solution according to fitness, and the swarm located the best solution (g).

Update Velocity and Position: Up the particle velocity and position by using:

$$u_i = wu_i + c_1 s_1 (q_i - a_i) + c_2 s_2 (g - a_i) \quad (5)$$

$$a_i = a_i + u_i \quad (6)$$

Termination: Continue, running the optimization for iterating maximum iterations or until the global best fitness is attained.

Here w = inertia weight, c1 = cognitive acceleration coefficients, c2 = social acceleration coefficients, s1 and s2 = random numbers in [0,1].

5.3 Support Vector Machine (SVM)

SVM is also one of the maximum used powerful tool under the category of supervised learning in classification as well as in regression. Voting for an SVM, this one's goal is to find a hyperplane that will try to maximize the distance between points fitting to dissimilar classes in a high-dimensional feature space. In regression problems, SVM employs Support Vector Regression (SVR) that gives the error requirement within a certain range. SVM has great performance especially in the nonlinear problem using kernel functions like the RBF kernel, to map the data into the higher dimensional intergalactic of which linear separation of the data is possible.

Kernel Function: In this case, the RBF kernel is applied for data with nonlinear relationships.

$$I(a, a^1) = \exp(-\gamma \|a - a'\|^2) \quad (7)$$

Objective Function: To achieve a happy medium between complexity of models and prediction error, SVM aims to reduce the following objective function:

$$\frac{1}{2}\|w\|^2 + C \sum_{k=1}^{m} (\xi_k + \xi_k^*) \quad (8)$$

Constraints: SVM applies certain constrains to make the error in the predictions to be within a region ϵ

$$b_i - (w \cdot \phi(a_i) + y) \leq \in + \xi_k \quad (9)$$

$$(w \cdot \phi(a_i) + y) - b_i \leq \epsilon + \xi_k^* \quad (10)$$

$$\xi_k, \xi_k^* \geq 0 \quad (11)$$

Regression Function: As a result of optimization, the last estimation for an independent data point

x is given by

$$\hat{b} = w \cdot \phi(a) + y \quad (12)$$

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

Here γ = adjusts the impact of each training points, C = penalty parameteer, y = bias, w = weight vector, ϕ(x) = kernal transformed input, ϵ = loss parameter, ξ_k, ξ_k^* = slack variables for tolerance, $\hat{b}$ = predicted stock price.

Algorithm for PCO-PSO-SVM

Input: Dataset E with features a and labels b. Constraints for PCA(i), PSO (w, c1, c2) and SVM Data Preprocessing Normalize a Handle missing values and outliers Dimensionality Reduction (PCA): Compute U_i (top i eigenvectors). Transform a to b using U_i. Iinitialize PSO: Select random values a_i and u_i wanted in the hyperparameter space. Optimize SVM Parameters: For every iteration: Evaluate fitness of every particle using RMSE Update q_i and g Update u_i and a_i Train SVM: Then, the results are attained by applying optimized C, γ, and ϵ for the SVM model with the reduced dataset. Test Model: The final model built using optimized SVM best hyperparameters.

Consequently, the proposed PCA-PSO-SVM approach enhances the previous stock price prediction by PCA-PSO and PSO-SVM, by eliminating multiline affecting data, doing hyperparameters' optimization and taking SVM advantages of predicting nonlinearity. It is because this integrated framework improves the forecast breadth, depth, and adaptability so that the availability of this solution is highly suitable for modeling price movements in intricate stock markets.

6. Result

The outcome of the work shows the high accuracy of the proposed PCA-PSO-SVM framework for forecasting stock prices. In particular, the PCA is chosen for dimensionality reduction, PSO for hyperparameters optimization, and SVM-regression for modelling given high-dimensional and noisy financial information. Regarding the assessment, RMSE and MAPE tests are used within the assessment formula to check the model's prediction capability. The findings demonstrate that developed predictive representations provide much better prediction efficiency than simple statistical models and even single application of ML algorithms.

Accuracy (ACC): The ratio between the correct forecast of the stock prices and the total forecasted stock prices. It means the bigger figure gives a better performing model as compared to the lower figure.

$$Accuracy = \frac{Correct\ Positive + Correct\ Negative}{Correct\ Positive + Correct\ Negative + Incorrect\ Positive + Incorrect\ Negative} \quad (13)$$

Precision (Pre): Measuring how accurate the positive predictions are, the ratio of correct positive predictions to the whole number of positive predictions.

$$Precision = \frac{Correct\ Positive}{Correct\ Positive + Incorrect\ Positive} \quad (14)$$

Recall (Rec): The proportion of correctly identified positives of the actual sum of positives that shows the model's performance in defining positive stock trends.

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

$$Recall = \frac{Correct\ Positive}{Correct\ Positive + Incorrect\ Negative} \quad (15)$$

Sensitivity (Sen): The yield of the proposed model: accuracy of detecting positive stock price movement (analogous to the yield).

$$Sensitivity = \frac{Correct\ Positive}{Correct\ Positive + Incorrect\ Negative} \quad (16)$$

Specificity (Spe): Non-positive cases: the capacity for correctly forecasting a decline in stock price.

$$Specificity = \frac{Correct\ Negative}{Correct\ Negative + Incorrect\ Positive} \quad (17)$$

Here

Correct Positive = Days when the stock prices rose and the model correctly anticipated the rise.

Correct Negative = Those situations in which the days in which the prices went down and the model was right to go down on this stock.

Incorrect Positive = Such days as where the model marked the wrong prediction, by pointing to a higher stock price.

Incorrect Negative = cases where the model predicted a lack of a rise when in real life a stock price rose.

Mean Absolute Error (MAE): MAE, on the other hand, progresses to average the magnitude of the errors in any way, without reference to their direction. It provided an absolute value of mean error which is easy for comprehension.

$$MAE = \frac{1}{m} \sum_{k=1}^{m} \left| b_i - \hat{b}_i \right| \quad (18)$$

Root Mean Square Error (RMSE): RMSE is the statistical method determining the square root of the normal of squares on the difference amongst the forecasted values and the actual values. It provides more emphasis in large deviations because of squaring the deviations; thus, this technique is sensitive to outliers.

$$RMSE = \sqrt{\frac{1}{m} \sum_{k=1}^{m} \left(b_i - \hat{b}_i \right)^2} \quad (19)$$

Mean Absolute Percentage Error (MAPE): MAPE calculates mean percentage complete deviations from the actual values to the predicted ones. It presents the errors it to actual value because there are times the measured error can be misleading depending on the scale of the dataset used to arrive at the models.

$$MAPE = \frac{100}{m} \sum_{k=1}^{m} \left| \frac{b_i - \hat{b}_i}{b_i} \right| \quad (20)$$

Here m = total number of information points, b_i = actual stock price at kth observation, $\hat{b}_i$ = Predicted stock price at kth observation.

Table 1: Performance comparison of existing approach with suggested approach

Approach	**Accuracy %**	**Precision %**	**Recall %**	**Sensitivity %**	**Specificity %**
SVM	84.5	83.2	84.1	84.5	84.3
PSO-SVM	86.5	85.8	86.11	86.2	85.9
PCA	83.2	82.1	82.3	82.1	83.1
PCA-SVM-PSO	95.5	95.1	95.3	95.2	95.1

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

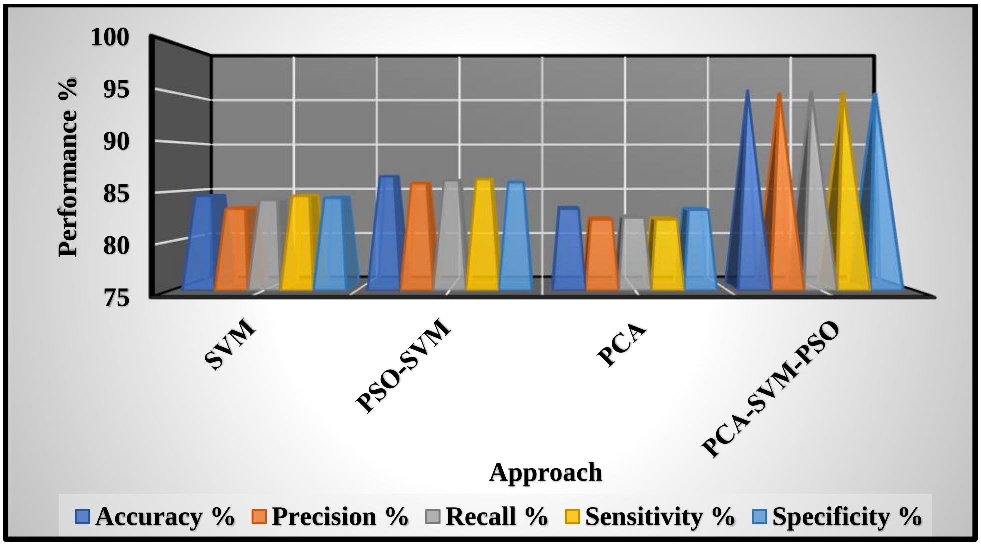

Figure 3. Graphical representation of Performance comparison of existing approach with suggested approach

The findings of the proposed approach using PCA-SVM-PSO are compared to other related methods to stress their usefulness in predicting the stock prices. They are proved to yield higher accuracy of 95.5%, compared with the accuracies of the standalone SVM (84.5%), PSO-SVM (86.5%) and PCA (83.2%). The value of the precision, recall, sensitivity, and specificity of PCA-SVM-PSO is all above 95%, because the machine need to minimize both incorrect positives and incorrect negatives for the prediction of the movements in stock price. By combining PCA in dimensionality reduction, PSO in hyperparameter optimization, and SVM in regression, the study presents a stable framework. Other approaches fail to do this to the best of their ability because they do not handle data complexity and problems with parameter tuning.

Table 2: Comparison of Training phase error rates for existing approach with suggested approach

Approach	**MAE %**	**RMSE %**	**MAPE %**
SVM	3.5	6.2	1.11
PSO-SVM	2.89	5.8	0.95
PCA	3.1	6.1	0.98
PCA-SVM-PSO	2.01	3.89	0.88

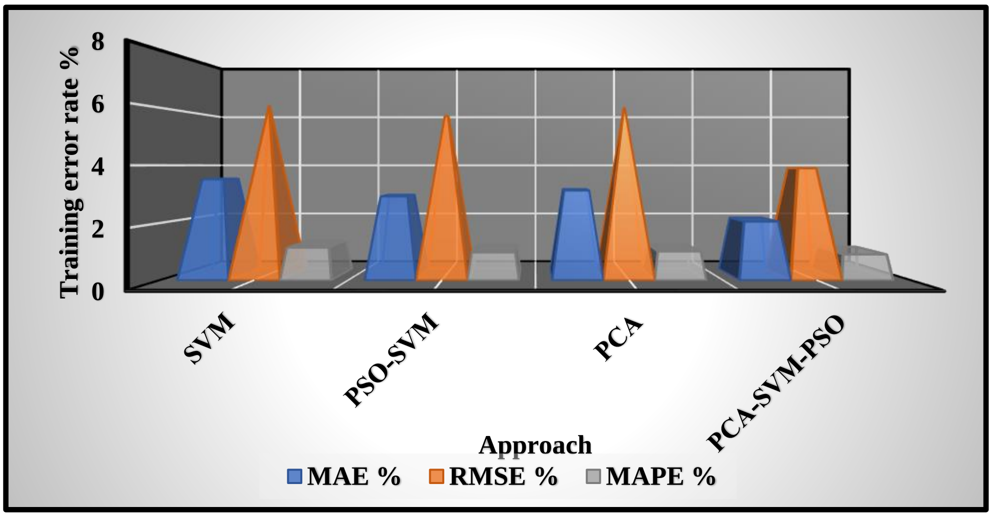

Figure 4. Graphical representation of Comparison of Training phase error rates for existing approach with suggested approach

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

The results of the training errors demonstrate that the PCA-SVM-PSO approach yields the lowest errors among the compared techniques in terms of given indicators: MAE 2.01%; RMSE 3.89%; and MAPE 0.88%. These results show its higher effectiveness than currently used models in reducing prediction error. PSO-SVM, with MAE of 2.89% and RMSE of 5.8%, performs better than standalone SVM (MAE: 3.5) The proposed technique of the study resulted in better values for both MAE and RMSE compared to those of the benchmark statistical methods of the study such as multiple linear regression (MLR) with an MAE of 5.0% and a RMSE of 6.2%) and Principal Component Analysis (PCA) with an MAE of 3.1% and a RMSE of 6.1%. From the methodological point of view, the PCA-SVM-PSO is a winner as it uses PCA's data dimensionality reduction, SVM's ability to solve robust regression problems, and PSO, which optimally defines SVM hyperparameters, and in this combination, the training error rates are reduced to a minimum in comparison with other similar techniques.

Table 3: Comparison of Testing phase error rates for existing approach with suggested approach

Approach	MAE %	RMSE %	MAPE %
SVM	3.3	7.3	0.84
PSO-SVM	2.97	6.8	0.79
PCA	3.1	7.1	0.81
PCA-SVM-PSO	1.99	3.88	0.59

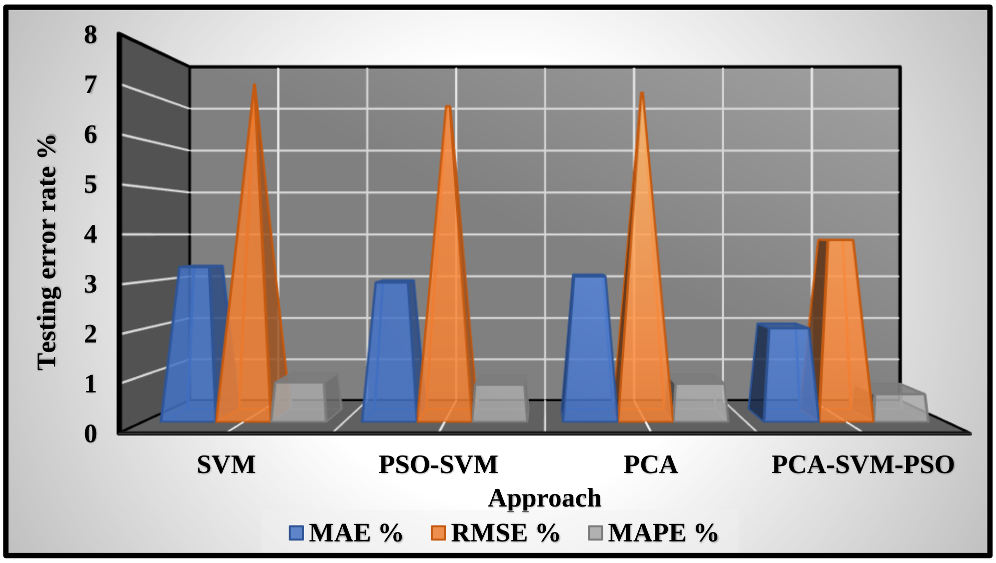

Figure 5. Graphical representation of Comparison of Testing phase error rates for existing approach with suggested approach

The testing error rate comparison highlights the superior performance of the PCA-SVM-PSO approach, which achieves the lowest errors across all metrics (MAE: 1. Distribution comparison; MAE: 1.92%, RMSE: 3.88 %, MAPE: 0.59%. This shows its high level of applicability to data it has not learned from in its process known as "out-of-sample estimation". The standalone SVM method shows higher errors (MAE: 3.3 While PSO-SVM is (MAE: 2.97%, RMSE: 6.8%, MAPE: 0.79%); PCA (MAE: 3.1%, RMSE: 7.1%, MAPE: 0.81%) shows moderate enhancements. The integration of the three stages of PCA for dimensionality reduction, SVM — for regression, and PSO for selecting the hyperparameters leads to the development of the highly efficient PCA-SVM-PSO approach that prevents significant errors during testing.

7. Conclusion and Future Enhancement

Fluctuations in stock prices cannot be easily explained due to the variety of various global, national and market factors. It is always a herculean task to make a forecast where variables are unexpectedly changeable in such a dynamic field. The use of conventional techniques of prediction entails some limitations, especially when dealing with hidden and nonlinear associations between variables as well as increased dimensions of the financial data. In this context, the use of machine learning which supports high- level optimization techniques provide a good

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

solution. This study goals to present a full investigation on stock price fluctuation by establishing a new ensemble model which consists of PCA, PSO, and SVM. Dimensionality reduction from PCA makes it possible to work with large numbers of features and original data dimensions in high datasets with some ease. The hyperparameters of SVM are tuned by PSO to get the best predictive efficiency of the designed model. SVM as the main predictive engine can capture the nonlinearity of the relationships; thereby, making the entire framework more resilient to different forms of noise and volatilities inherent in stock market data. The PCA-PSO-SVM framework hypothesized outperformed the proposed model accuracy and computational time of stock prices. By reducing data hardness, filtering the noise and handling parameter sensitivity, this combined forecasting approach improves the accuracy of prediction by a wide margin. The present paper adds to the existing literature on machine learning targeted at the finance domain by offering a stable, high-performing approach to stock price prediction and more generally, financial analysis.

There are potential enhancements as follows: the addition of more powerful optimization algorithms as the genetic algorithms, the use of real-time data with sentiment analysis and deeper architectures like LSTMs for dependencies on temporal variables. The application to other financial assets, like cryptocurrencies and commodities, might help improve the flexibility and fit of the framework to an even greater extent.

References

[1] L. Nemes and A. Kiss, "Prediction of stock values changes using sentiment analysis of stock news headlines", Journal of Information and Telecommunication, vol. 5, no. 3, pp. 375-394, Feb. 2021.

[2] Hayder Sabah Salih,Fatema Akbar Mohamed, Fusion-based Diversified Model for Internet of Vehicles: Leveraging Artificial Intelligence in Cloud Computing, Journal of Fusion: Practice and Applications, Vol. 12 , No. 2 , (2023) : 54-69 (Doi : https://doi.org/10.54216/FPA.120205)

[3] Zeena N. Al-kateeb,Dhuha Basheer Abdullah, A Smart Architecture Leveraging Fog Computing Fusion and Ensemble Learning for Prediction of Gestational Diabetes, Fusion: Practice and Applications, Vol. 12 , No. 2 , (2023) : 70-87 (Doi : https://doi.org/10.54216/FPA.120206)

[4] V. Gupta and M. Ahmad, ''Stock price trend prediction with long shortterm memory neural networks,'' Int. J. Comput. Intell. Stud., vol. 8, no. 4, p. 289, 2019, doi: 10.1504/ijcistudies.2019.10025266.

[5] X. Zhang, S. Qu, J. Huang, B. Fang and P. Yu, "Stock Market Prediction via Multi-Source Multiple Instance Learning", IEEE Access, vol. 6, pp. 50720-50728, 2018.

[6] M. Elsharkawy, I.S. Farahat, "A Proposed Predictive Model for Business Telemarketing Information Management, Journal of Cybersecurity and Information Management", Vol. 9, No. 1, (2021): 27-39 (Doi: https://doi.org/10.54216/JCIM.090103)

[7] L. Wang, F. Ma, J. Liu and L. Yang, "Forecasting stock price volatility: New evidence from the GARCH-MIDAS model", Int. J. Forecasting, vol. 36, no. 2, pp. 684-694, Apr. 2020.

[8] L. Yang and T. Zhai, ''Research on sentiment tendency analysis of video reviews based on sentiment dictionary,'' Netw. Secur. Technol. Appl., vol. 255, no. 3, pp. 53–56, 2022.

[9] A. E. Khedr, S. Salama and N. Yaseen, "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis", International Journal of Intelligent Systems and Applications, vol. 9, no. 7, pp. 22-30, Jul. 2017.

[10] M. Nazar Dawood, M. Ayad Alkhafaji, A. Hussian, H. Alaa Diame, N. Ali Hussien, S. Yassine, V. Rajinikanth, "Developing a Risk Management System with an Optimistic Predictive Approach and Business Decision-Making", Journal of Intelligent Systems and Internet of Things, Vol. 9, No. 2, (2023): 51-64 (Doi: https://doi.org/10.54216/JISIoT.090204)

[11] A. Sariga , J. Uthayakumar, Type 2 Fuzzy Logic based Unequal Clustering algorithm for multi-hop wireless sensor networks, International Journal of Wireless and Ad Hoc Communication, Vol. 1 , No. 1 , (2020) : 33-46 (Doi : https://doi.org/10.54216/IJWAC.010102)

[12] Irina V. Pustokhina, Blockchain technology in the international supply chains, International Journal of Wireless and Ad Hoc Communication, Vol. 1 , No. 1 , (2020) : 16-25 (Doi : https://doi.org/10.54216/IJWAC.010103)

[13] Parvesh K , Tharun C , Prakash M, Apparel Recommendation Engine Using Inverse Document Frequency and Weighted Average Word2vec, Journal of Cognitive Human-Computer Interaction, Vol. 1 , No. 2 , (2021) : 46-56 (Doi : DOI: https://doi.org/10.54216/JCHCI.010201)

[14] Vijay K, Collaborating The Textual Reviews Of The Merchandise and Foretelling The Rating Supported Social Sentiment, Journal of Cognitive Human-Computer Interaction, Vol. 1 , No. 2 , (2021) : 63 - 72 (Doi : DOI: https://doi.org/10.54216/JCHCI.010203)

[15] Y. Trichilli, M. B. Abbes and A. Masmoudi, "Predicting the effect of googling investor sentiment on Islamic stock market returns: A five-state hidden Markov model", Int. J. Islamic Middle Eastern Finance Manage., vol. 13, no. 2, pp. 165-193, Feb. 2020.

DOI: https://doi.org/10.54216/AJBOR.120103

Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024

[16] V. Roy et al., “Network Physical Address Based Encryption Technique Using Digital Logic”, International Journal of Scientific & Technology Research, Vol. 9, No. 4, 2020, Pp no.- 3119-3122.

[17] S. V. Kolasani and R. Assaf, "Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks", Journal of Data Analysis and Information Processing, vol. 08, no. 04, pp. 309-319, 2020.

[18] Yutao Han , Ibrahim M. EL-Hasnony , Wenbo Cai, Dragonfly Algorithm with Gated Recurrent Unit for Cybersecurity in Social Networking, Journal of Cybersecurity and Information Management, Vol. 0 , No. 2 , (2019) : 75-88 (Doi : https://doi.org/10.54216/JCIM.000107)

[19] Esraa Mohamed, The Relationship between Artificial Intelligence and Internet of Things: A quick review, Journal of Cybersecurity and Information Management, Vol. 1 , No. 1 , (2020) : 30-34 (Doi : https://doi.org/10.54216/JCIM.010101)

[20] M. M. Rounaghi and F. N. Zadeh, ‘‘Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model,’’ Phys. A, Stat. Mech. Appl., vol. 456, pp. 10–21, Aug. 2016, doi: 10.1016/j.physa.2016.03.006.

[21] M. Vijh, D. Chandola, V. A. Tikkiwal and A. Kumar, "Stock Closing Price Prediction using Machine Learning Techniques", Procedia Computer Science, vol. 167, pp. 599-606, 2020.

[22] Y. Ji, A. W. Liew, and L. Yang, ‘‘A novel improved particle swarm optimization with long-short term memory hybrid model for stock indices forecast,’’ IEEE Access, vol. 9, pp. 23660–23671, 2021, doi: 10.1109/ACCESS.2021.3056713.

[23] G. Ataman and S. Kahraman, "Stock market prediction in Brics countries using linear regression and artificial neural network hybrid models", Singap. Econ. Rev., vol. 67, no. 2, pp. 635-653, 2022.

[24] V. Roy. "An Improved Image Encryption Consuming Fusion Transmutation and Edge Operator." Journal of Cybersecurity and Information Management, Vol. 8, No. 1, 2021, PP. 42-52.

[25] H. Liu and Y. Hou, ‘‘Application of Bayesian neural network in prediction of stock time series,’’ Comput. Eng. Appl., vol. 55, no. 12, pp. 225–229, 2019.

DOI: https://doi.org/10.54216/AJBOR.120103
Received: June 10, 2024 Revised: September 25, 2024 Accepted: December 10, 2024