Journal of Information Systems Engineering and Management
2025, 10(8s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Conditional Mutual Information Maximization and Restricted Boltzmann Machine for Stock Price Prediction

P. Deepa[1], Dr.B.Murugesakumar[2]
Ph.D Research Scholar, Department of Computer Science, Dr.SNS Rajalakshmi College of Arts and Science (Autonomous), Coimbatore-641049[1]
Associate Professor, Head, Department of Computer Science, Dr.SNS Rajalakshmi College of Arts and Science (Autonomous), Coimbatore-641049[2]

ARTICLE INFO

Received: 19 Oct 2024

Revised: 13 Dec 2024

Accepted: 24 Dec 2024

ABSTRACT

Using various machine learning (ML) methodologies in forecasting the stock prices has lately been successful. Nonetheless, most of these models depend on the narrow set of attributes as input and lacks the sufficient information to offer the estimates of stock market. To enhance the stock price prediction models, conditional mutual information maximizing (CMIM) method is used for data preprocessing and feature selection based on restricted boltzmann machine (RBM). By optimizing their conditional mutual information with the target variable, CMIM assists in selecting the most significant characteristics which reduces the dimensionality and duplication. Following that, the picked features are refined using RBM, ML model is capable of detecting the hidden patterns in data. These approaches use selection algorithms to extract the key features. Hence, improves the performance of stock price prediction models. In terms of prediction accuracy, the experiments results indicate the significantly outperforming standard feature selection strategies. This proposed method outperforms with 97.69% accuracy. Thus, the model offers a solid foundation for financial forecasting.

Keywords: Conditional mutual information maximization, feature selection, machine learning, stock price, restricted boltzmann machine

INTRODUCTION

Stock price prediction has been the primary focus of finance research in various machine learning (ML) [1] and deep learning (DL) [2] approaches are used to understand the complicated market dynamics. This is challenging because of the given volatility of stock prices and their impact on variables such as market sentiment, global events and trading volumes [3]. In predictive modeling, selection of right characteristics from the financial data [4] is essential for reduced over fitting and improved model performance. Stock price forecasting has long been a popular and significant topic in the study of financial econometrics [5]. Standard statistical approaches for time series prediction are constrained due to the complexities of stock market [6]. As a result, ML and many other artificial intelligence (AI) systems including the restricted boltzmann machine (RBM) and conditional mutual information maximization (CMIM) have been developed to predict the stock prices with exceptional accuracy [7].

With CMIM, one focus in selecting the qualities is to predict the target variable. CMIM ensures the selected features are highly relevant to predictive modeling goal by optimizing the conditional mutual information between features and the target [8]. CMIM eliminates the duplicate and unnecessary characteristics by selecting the features which complement one another regarding the aim variable [9]. CMIM likes to choose the diversified and non-redundant features by maximizing the conditional mutual information in enhancing the quality of feature set [10]. Even with little data, CMIM identifies the informative aspects extremely well. CMIM is appropriate for scenarios with restricted data availability because this method extract the relevant features from high-dimensional datasets with relatively small sample sizes by focusing on the conditional mutual information [11].

RBM, a kind of markov random filed (MRF) consists of a visible and hidden layer without linking between the units inside each layer [12]. Because of its high level of representation, RBM has been used in literature, music,

motion data, photography, and other areas. In underactive supervision, RBM uses hidden units known as features to create the discriminative items from complex data sets [13]. Further, RBMs uses deep belief network (DBN) which is very adaptable and capable of learning the low-dimensional discriminative features from high-dimensional and complex data [14]. DBN excels in numerous types of learning tasks including the object recognition and document categorization [15-17]. This methodology aims to overcome the limitations of traditional feature selection approaches by the mutual information-based preprocessing with the powerful feature extraction capabilities of CMIM and RBM. Demonstration is done by the experimental analysis of improved prediction accuracy strategy which results in the solid solution of stock market forecasting.

This paper primarily provides a combination technique for forecasting the stock price using CMIM, data preprocessing, RBM and the deep variant DBN as feature selectors.

The work is organized as follows. Section II covers various review articles. Section III discusses the utilization of RBM as feature extractor and CMIM for data preparation. In Section IV, tests are performed on raw data to determine the predictive efficacy of system on comparison to the existing ones.

LITERATURE REVIEW

Akhtar et al. [18] these authors found that random forest (RF) approach was superior in estimating the market price of inventory depending on several criteria of collecting the ancient information by means of accuracy comparison in well-known algorithms. Because of their knowledge in lot of historical data and statistical pattern testing, stock market agents and traders found the instructions helpful. Finally, in the test set, support vector machine (SVM) model scored 78.7% whereas the RF classifier scored 80.8%. The difficulty demonstrated that ML model had forecasted the inventory charge well than the previous techniques.

Albahli et al. [19] people expressed their own views and thoughts in business, individual and scandal on social media. Twitter was a well-known, cutting-edge social networking site where the users expressed succinctly. Social media communications such as tweets were used to a large extent and forecasted the stock market behavior by the usage of sentiment analysis (SA). Those authors proposed a novel technique in predicting the stock market movements based on SA. The model had been constructed in the combination of recurrent neural network (RNN), extreme learning machine (ELM) classifiers as well as the suggested stock senti word net (SSWN) sentiment lexicon. To estimate the stock market fluctuations, ten ML models were examined using Twitter's and Sentiment140 datasets.

Alotaibi [20] this new stock market prediction algorithm had three foundations: extracting features, selecting the best ones and providing predictions. One received the traits and characteristics such as second-order technical indicator-based (SOTI) characteristics than the others. During prediction step, the obtained characteristics were used to train the classifiers. If the best-predicted results were produced, choosing the most relevant features becomes even more critical. Examining the acquired ones assisted the Red Deer Adopted Wolf Algorithm (RDAGW) in selecting the best features to adopt.

Asadi et al. [21] pre-processed evolutionary levenberg-marquardt neural network (PELMNN) was created by combining the genetic algorithms, feed-forward neural networks, data pre- and post-processing in stock market prediction. These authors scaled the input data using data transformation. Those authors used the stepwise regression to choose the characteristics that significantly affected the stock index and removed the others. As a worldwide search, genetic algorithm developed the neural network weights. This development was done before levenberg-marquardt (LM) algorithm tweaking. Levenberg-Marquardt method used acquired weights as the starting weights for local searches. Final post-processing produced the predicted values and restored the output data to the original scale.

Jagadesh et al. [22] finally, constructing a model for predictive analysis using the noisy and non-linear stock market dataset was evidently difficult. This work used DL architectures to anticipate the closing values of Nifty 50 index for the next trading day. Therefore, the forecasting techniques were improved. To solve the primary challenges in stock market prediction, this study combined the efficient methodologies such as preprocessing, feature selection and classification. Dandelion optimization algorithm (DOA) used feature selection to improve the quality and relevance of input data, where as wavelet transform promoted the data cleaning and noise reduction. Stock market data study showed that the suggested hybrid model combined 3D convolutional neural network (CNN) with gated recurrent

unit (GRU) was successful. Hyper parameter change improved the model performance using the blood coagulation algorithm (BCA), resulted in improved prediction accuracy and resilience.

Javed Awan et al. [23] to forecast the changes in stock prices using Spark big data platform; those authors used many ML models. For their stock price prediction, researchers were turned to Spark machine learning library (MLlib). Using ten different firm data sets, the authors ran ML algorithms. Outfits from the linear regression (LR), RF and generalized LR outperformed those from the decision tree (DT) model. When the data is applied with a texture, naive bayes (NB) and LR had achieved the accuracy rates of around 77% to 80%.

Rasel et al. [24] the study's goal was to determine the best strategy in forecasting the stock market trends and prices using artificial neural networks (ANNs) with specialized operators, well-chosen operator parameters and well defined operators. Using two separate assessment techniques like mean average percentage error (MAPE) and root-mean-square error (RMSE) computations—the suggested models predicted the price and stock market trends with little error. Out of three models, one that forecasted the market price in advance was most accurate. For those in the business world, the suggested model offered improvement in the present methods of stock prices and trends prediction that were notoriously inaccurate.

Paramita & Winata [25] this study investigated the feature selection approaches such as principal component analysis (PCA), information gain (IG) and recursive feature elimination (RFE) in the context of stock market prediction. In comparison to the other methodologies, RFE provided the most accurate market value projection. Reduced data dimension, improved prediction in model's performance and identified significant characteristics in prediction were possible with the aid of RFE. Better prediction was potentially attained via the use of more sophisticated feature selection methods or the mix of several methods performance. Finally, research using real-time data was checked and assessed the prediction model to see the adaptation towards the different market scenarios.

Yuan et al. [26] this research sought to determine the choosing of well integrated stock selection model features and forecasted the future stock prices. Feature selection procedures filtered the feature values and had set the surrounding scene. Specifying the parameters for cross-valuation by temporal sliding window technique, the algorithms contributed in making the stock price trend prediction model applicable to the real-world investment agreements. Experiments showed that the usage of RF in feature selection and stock price trend prediction provided the best results.

Singh et al. [27] this proposed research predicted the real-time stock prices using a trained neural network (NN) model. The experimental method named NN produced the remarkable exact target predictions and model's consistency. Flexibility of NN enabled the swiftly linking of input and output data to forecast the stock market movements. The research indicated that the NN forecasted the financial data in real time. Experimental findings showed that the proposed NN model outperformed the presently used stock market forecasting algorithms with 86% accuracy in training set and 14% loss in testing set. Table 1 compares the data preprocessing and feature selection for stock market datasets.

Table 1 Comparison Table on Data Preprocessing and Feature Selection for Stock Market Datasets

Authors	Key Contribution	Techniques Used	Strengths	Limitations
Macedo et al. [28]	Proposes a mutual information decomposition framework for feature selection.	Decomposed Mutual Information Maximization	Effective for high-dimensional data, enhances computational efficiency.	May require fine-tuning for specific applications.
Zhou et al. [29]	Combines mutual information with correlation coefficients for feature selection.	Mutual Information, Correlation Coefficient	Balances relevance and redundancy in feature selection.	Limited exploration of deep learning integration.

Wu et al. [30]	Introduces a collaborative filtering algorithm using trust information with RBMs.	RBM, Trust Integration	Suitable for recommender systems; handles sparsity effectively.	Limited to collaborative filtering tasks.
Abroshan & Moattar [31]	Proposes a discriminative feature selection algorithm using signed Laplacian RBMs.	Signed Laplacian RBM	Enhances generalization in high-dimensional classification tasks.	Computational overhead in large datasets.
Liang et al. [32]	Utilizes RBMs for stock market trend prediction.	RBM for Temporal Prediction	Effective in time-series analysis and stock prediction.	Lacks generalization to non-financial datasets.
Taherkhani et al. [33]	Introduces Deep-FS, a feature selection algorithm for Deep Boltzmann Machines.	Deep Boltzmann Machines, Feature Selection	Efficient for deep models; improves accuracy in high-dimensional data.	Requires significant computational resources.

2.1 Problems identification

Preprocessing and feature selection using existing ML approaches took longer and unable to handle the difficult data. Data access has been restricted. The accuracy and robustness of stock price prediction algorithms has varied based on the data accessibility. In this study, the recommended RBM and CMIM contribute to the increased process efficiency and speed.

MATERIALS AND METHODS

This section begins with an overview of the dataset collected from Kaggle source. Here, the features are preprocessed and picked from the Indian stock market dataset. The Kaggle dataset are preprocessed using CMIM. Preprocessing using CMIM significantly increased the ML model performance. The primary use of CMIM is the choosing of features from a dataset in order to determine the relevance. When multiple financial indicators and market determinants are present in stock market prediction, CMIM helps to identify the features which have strongest relationship with the objective variable like future stock prices. By stressing these qualities, the model achieves greater efficiency and accuracy. RBM, a kind of generative stochastic NN is capable of effectively learning the complicated patterns from high-dimensional input. It is used to enhance the stock market prediction algorithms. Figure 1 illustrates the Data Preprocessing and Feature Selection procedure using CMIM and RBM.

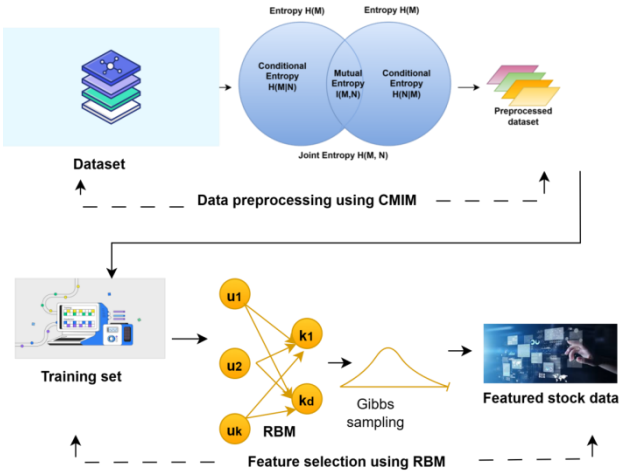

Figure 1 Overall architecture of data preprocessing and feature selection

3.1 Dataset collection

Indian stock market statistics have been accumulated for personal, business and instructional use. This information contains the stock values for NIFTY-50 index which includes the top 50 Indian corporations. The dataset has been provided by the Python package "y-finance" from Yahoo Finance. The dataset mostly covers the period from 3 January 2000 to June 2023.

Kaggle link: https://www.kaggle.com/datasets/rockyjoseph/nifty-50-stock-market-data-2000-2023.

3.2 Data preprocessing using conditional mutual information maximization

Information theory has been extensively used in various sectors which includes Science, Engineering and Commerce (Wang et al. [34]). This maximization is a catch-all term for several research methods working and offers the statistical underpinning of data standardization. Among the other preprocessing approaches, mutual information and information gain are found in this concept. The emphasis is on two discrete random variables, M and N, which are assumed to take discrete values. Entropy *H (M)* is the most fundamental concept in information theory as several features aid in the instant understanding about the information measurement. Equation (1) states the entropy *H (M)*.

$$H(M) = -\sum_{m \in M} p(m) \log p(m) \quad \text{-------- (1)}$$

Entropy is dependent on the chance and has not bared the real values obtained by the incidental variable m. Each logarithm appears in base two. Entropy value is the average number of bits needed to encode or characterize the incidental variable m. An incidental variable maximizes entropy with the stable distribution of values. Equation (2) defines the mutual information *I (M, N)* as the degree of information flow between *M* and *N*.

$$I(M,N) = -\sum_{m,n} p(m,n) log \frac{p(m,n)}{p(m)p(n)} \text{-------- (2)}$$

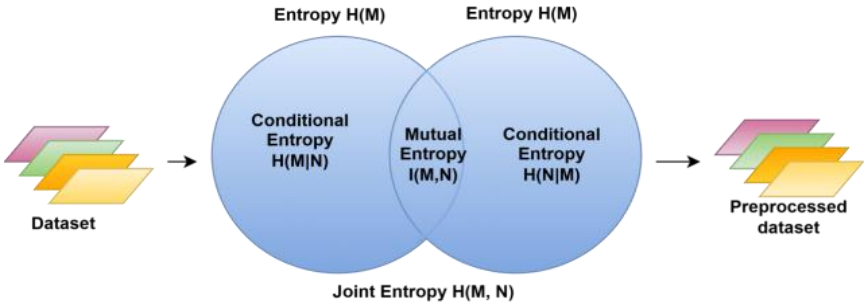

Figure 2 Data preprocessing using maximizing conditional mutual information

In this study, mutual information to preprocess the dataset is seen in Figure 2. Maximizing conditional mutual information (CMI) gives the expected value of two random variables mutual information when the third variable's value is taken into account. CMIM eliminates the duplicate and unnecessary characteristics by selecting the features which complement one another regarding the aim variable. Maximizing CMI allows CMIM to choose several diverse and non-redundant features which boosts the quality of feature subset. Data are preprocessed on Indian equities using the procedures listed below.

To calculate CMI directly,$I\ (P*; C|P_1, \dots, P_l)$ requires the calculation of complex joint probability, which is computationally demanding. To address this difficulty, examination of CMI is aimed by decoupling into simpler forms that are devoid of complicated joint probability. Using l-dimensional forms, such as$I\ (P*; C|P_1, \dots, P_l)$, approximation of $I(P*; C|P_i, \dots, P_j)$ | *{z} l-1)* is tried first. More information reduces uncertainty, resulting in$I\ (P*; C|P_1, \dots, P_l)$ lower than $I(P*; C|P_i, \dots, P_j)$ | *{z} l-1).*$I\ (P*; C|P_1, \dots, P_l)$ is estimated using the lowest value (3).

$$I(P*; C|P_1, \dots, P_l) \approx \min I\ (P*; C|P_i, \dots, P_j) \mid \{z\}\ l\text{-}1) \text{ ------ (3)}$$

CMI in equation (3) is minimized by *l-1* features which are closely related to the feature *P** in the selected stock data. The predictive ability of P* is severely weakened. To prevent this issue, data with min$I\ (P*; V|P_i, \dots, P_j)|\ \{z\}\ l\text{-}1)$ is used as the greatest value. To maximize min$I\ (P*; V|P_i, \dots, P_j)|\ \{z\}\ l\text{-}1,)$, data $P*$ is substantial and influences

minimal on the previously chosen data. Each new data point ensures both the orthogonal to previous ones and informative. This work estimates the form$I\ (P*;V|P_1,\dots,P_l)$ by utilizing triplet form$I(P*;V|P_i)$. Hence, the computation costs are lowered. This simpler method quantifies the information from single feature P_i and forecasts the category using feature P codes. The simpler triplet$I(P*;V|P_i)$ is replaced for the proper form in equation (4). Thereby, the equation yields

$$I\ (P*;V|P_1,\dots,P_l) \approx \min I(P*;V|P_i) \text{ ------ (4)}$$

Choose a feature $P*$ to optimize min$I(P*;V|P_i)$). As a result, the technique is named as Conditional Mutual Information Maximization. To approximate the real CMI, more complicated form such as quadruplet min $I(P*;V|P_i)$similar to the triplet form is used. There are exactly $\frac{l(l-2)}{2}$quadruplets for l selected features; each quadruplet is complex to calculate than a triplet. As a consequence, approximations based on the sophisticated forms inevitably cause severe efficiency issues. Due to the lack of data, even complicated forms seem to be capable of providing the exact estimate. Yet, joint probability is estimated using smoothing techniques. Depending on triplet, the solution is not only efficient but also effective and robust in avoiding the traits from becoming dependent on each another.

Algorithm 1: Conditional mutual information maximization

Input:

- Dataset: $D = \{P_1, P_2, \dots, P_l\}$ where P_i represents the features and l is the number of features.

Steps:

Step 1: Initialize an empty set of selected features, $S = \{\}$

Step2: Calculate the mutual information $I(P*;V)$ between each feature P_i and the target variable$C$.

Step 3: Select the feature P_j with the highest mutual information with V and add it to S.

Step 4: For each remaining feature P_i not in S, calculate the conditional mutual information$I(P*;V)$,where S represent the set of features already selected.

Step 5: Select the feature P_i with the highest Conditional Mutual Information $I(P*;V)$and add it to$S$.

Step 6: Repeat steps 4 and 5 until $|\ S\ | = l$ or until there are no more features to select.

Step 7: Output the subset of selected features S

Output:

- Subset of l features S that maximally correlate with the target variable V

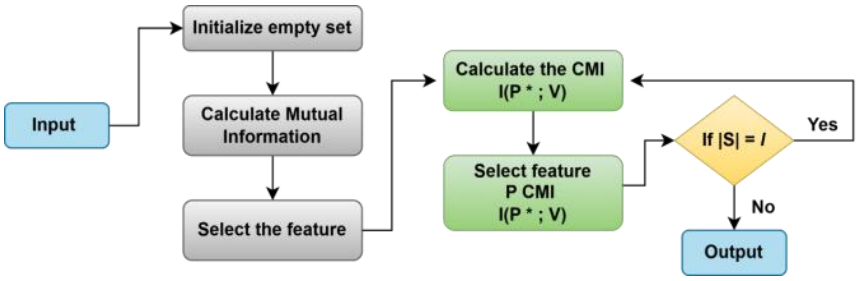

Figure 3 Flowchart for conditional mutual information maximization

Figure 3 and Algorithm 1 demonstrate the selection of CMIM technique to identify the most important features of model by maximizing their knowledge of target variable. This starts by calculating the mutual information for each feature and then chooses the most closely related to the aim. Then, repeatedly selects the additional characteristics by calculating their CMI, ensuring that each new feature delivers unique and non-redundant information. This approach continues until all the characteristics are considered and the desired number of features is selected. It results in the set of qualities with the highest correlation to target variable.

3.3 Feature Selection Using Restricted Boltzmann Machine Algorithm

Restricted boltzmann machine, an energy-based stochastic ANN with unsupervised learning, consists of binary neurons units. Stochastic NN are created by initiating the incidental fluctuations into the entire ANN. Overall feature selection (FS) process using RBM Algorithm is shown in figure 4.

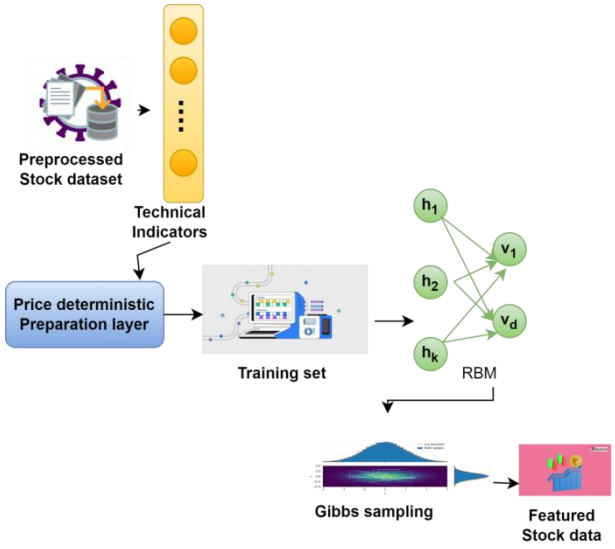

Figure 4 Feature selections using restricted boltzmann machine

Boltzmann Machine contains both visible and hidden neurons units, which replicates the data distribution from visible units as input. Developing a quick learning approach based on Cai et al. [35] offers complications due to the Boltzmann Machine's fully linked network. Restricted Boltzmann Machine allows only without the evident or hidden links. Figure 5 depicts a two-layer typical RBM, with the letters u and k representing the hidden and visible units respectively.

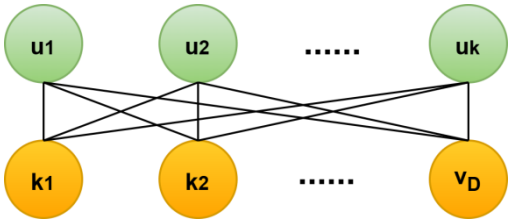

Figure 5 Two layer restricted boltzmann machine

While taking the states of hidden units into account, a generative probabilistic model RBM calculates the likelihood of visible unit's value based on the stock market dataset values. This characteristic allows in recreating the missing visible units. The generative quality of RBM is used to sample from the learning distribution to derive picture textures. Image de-noising applications have also taken the use of RBM generative ability to sample the missing parts in an input image. The proposed FS method describes the strategy for FS using the stock market dataset and RBM's generative nature. Feature Selection needs a set of features which includes the pragmatic information. As a result, RBM generative characteristic eliminates the portions of input data which are irrelevant in practice. The selected final characteristics reduce the network's complexity and include some features.

The constrained shape has not revealed any direct connections between similar components. The energy function in equation (5) represents RBM with D visible nodes $k = (k_1, k_2, \ldots, k_D)$and k hidden nodes $u = (u_1, u_2, \ldots, u_l)$.

$$E(k,u) = -u^T W k - b^T k - c^T u \text{ ----- (5)}$$

The variable W signifies the weights, whereas b and c represent the bias fields acting on the nodes of visible and hidden units. The model's stochastic development is controlled by the joint probability distribution.

$$p(k,u) = \frac{1}{Z} e^{-E(k,u)} \text{ ---------- (6)}$$

In equation (6), the partition function $Z=\sum_{k.u} e^{-E(k,u)}$.Using the samples in target probability distribution, enables to adjust the RBM's internal parameters like weights and biases and represent the key components of stock dataset under consideration. After training, RBM is extracted from the target probability distribution (stock market price data) using a generative model approach. Usage of hidden nodes allows in capturing the relationships which are not easily characterized by pair-wise interactions.

Weights and biases in RBM are often modified during training to fit the training data as closely as possible. Typically, $l = \{W, b, c\}$ [8, 18, 19] model parameters optimize the probability or likelihood $L(\theta \mid v)$ of the training data set X. Maximizing the log-likelihood analytically is a difficult challenge. Calculation of log-likelihood derivative on each iteration is the most computationally costly component of gradient descent approach as represented by single vector v.

$$\frac{\partial}{\partial\theta} logL(\theta|k)) = -\sum_u p(h|k)\frac{\partial E(k,u)}{\partial\theta} + \sum_{k,u} p(k,u)\frac{\partial E(k,u)}{\partial\theta} \quad \text{------- (7)}$$

$$= -\frac{\partial E(k,u)}{\partial\theta}_{data} + \frac{\partial E(k,u)}{\partial\theta}_{model}$$

Considering the training vector, first component of equation (7) is the expectation of the derivative energy function under the hidden nodes' conditional probability distribution. This is sometimes referred to as data-driven expectations. This model's second probability distribution component shows the expected value of $\frac{\partial E}{\partial\theta}$ for all visible and hidden nodes. Calculated estimates for the exposed and hidden nodes are based on exponential time. By selecting the samples from appropriate probability distributions, expectations are estimated using markov chain monte carlo (MCMC) methods. Inclusion of conditional probability reduces both the data-dependent and model-dependent expectations for RBM as the units in same layer are conditionally independent. Equation (8) allows reducing the derivative of log-likelihood regarding the weightW_{ij}.

$$\frac{\partial}{\partial W_{ij}} logL(\theta|k)) = p(u_i = 1|k)k_j - \sum_k p(k)p(u_i = 1|k)k_j \text{------- (8)}$$

Equations (9) and (10) compute log-likelihood derivatives for the bias c_j of j^{th}hidden node and the bias b_i of i^{th} visible node respectively.

$$\frac{\partial}{\partial b_i} logL(\theta|k)) = k_i - \sum_k p(k)k_i \text{ ------ (9)}$$

$$\frac{\partial}{\partial c_j} logL(\theta|k)) = p(u_j = 1|k)k_j - \sum_k p(k)p(u = 1|k) \text{ ------ (10)}$$

The second term i estimated using Gibbs sampling, which is difficult and intractable as the sampling requires summing up of all potential observable vectors.

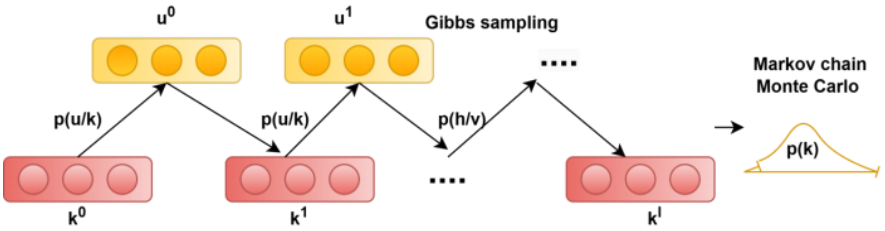

Figure 6 Gibbs Sampling with markov chain monte carlo estimation

The RBM model is constantly managing the model parameters to optimize probability using both persistent contrastive divergence (PCD) and contrastive divergence (CD) techniques. Markov chain monte carlo techniques are used to rate the previously indicated intractable terms, as displayed in figure 6. Gibbs Sampling allows CD to estimate the possibility deriving from RBM. Changes are noted in model parameters when RBM trains often because the MCMC converges on the target system joint probability distribution. Unlike CD, PCD predicts that the model parameters changes significantly at low learning rates by eliminating the requirement to restart the MCMC with training vector $v(o)$ after each parameter change. This generally let MCMC to reach the thermal equilibrium faster.

Algorithm 2: Restricted boltzmann machine

Input: preprocessed data

- Training dataset $D = \{k^1, k^2, \ldots, k^m\}$ consisting of m samples, where $k_i$represents the visible units.
- Number of visible units: n_k
- whole hidden units: n_u
- Study rate: α
- Number of training iterations: $num\_iterations$

Steps:

Step 1: Randomly initialize the weights W and biases b for visible and hidden units.Forward Pass (Positive Phase)

- For each training sample $k_i$in the dataset

Step 2:Compute the probability of activation for hidden units given visible units:

$$P(u_j = 1|k) = \sigma(b_i + \sum_{j=1}^{n_u} W_{ij} u_{ij})$$

Sample visible units v' from the conditional probability distribution$P(k \mid u_i)$.

Step 3: Update Parameters

Compute the positive and negative associations between visible and hidden units:

$$\Delta W = \alpha(< k_i u^T >_{data} - < k_i {u_i}^T >_{recon})$$

Repeat: Repeat steps 2 and 3 for a fixed number of training iterations or until convergence.

Output:

- Trained RBM model parameters: weights W and biases b for visible and hidden units.

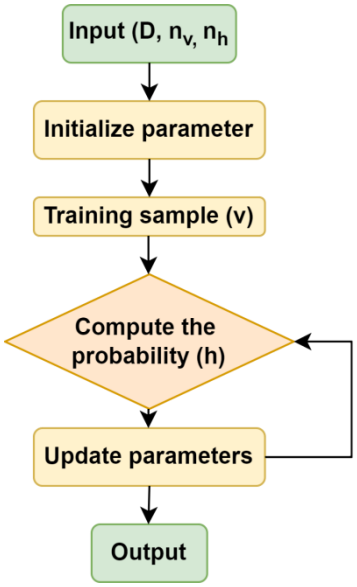

Figure 7 Flow chart of Restricted Boltzmann Machine Algorithm

For unsupervised learning, figure 7 and algorithm 2 shows RBM, a NN model with visible and hidden units. This learning starts by arbitrarily assigning the weights and preconceptions to these units. Beginning possibility of the hidden units is calculated for each training sample using the visible units. Esteem of the hidden units is later used to determine the adjustment of apparent units. Weights are modified by comparing the original and reconstructed data states. This procedure is repeated numerous times to generate the trained RBM with different weights and biases until the model converges.

RESULTS AND DISCUSSIONS

In this paper, Python is used to implement the feature selection. Conditional mutual information maximizing algorithm is used for data preprocessing and RBM for feature selection. Indian stock market datasets are taken for predicting the various performances of existing and proposed algorithms. Compared to the other existing

algorithms, the proposed CMIM and RBM perform well for FS. The most selected features are low, high and volume weighted average price (VWAP).Feature selection process is shown in the following charts.

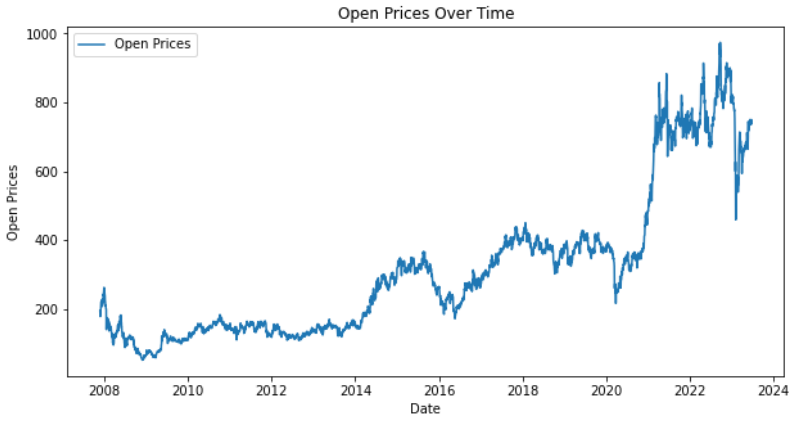

Figure 8 Open prices over time

Figure 8 shows the stock open prices from 2008 to 2024. Starting from 2017, the price rises significantly and peaks around 2022; considerable volatility follows. While the X-axis shows the years from 2008 to 2024,Y-axis indicates the open prices.

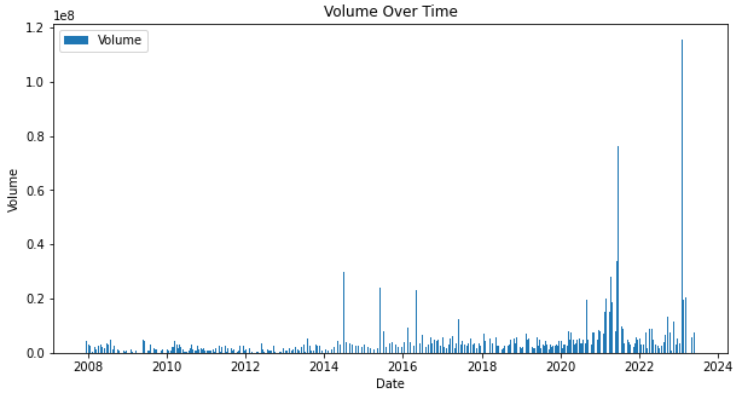

Figure 9 Volume over time

Figure 9 represents the volume of transactions from the year 2008 until 2024. From 2008 to roughly 2020, trade volumes are rather moderate and consistent. Trading volume grows dramatically in the beginning of 2020, peaks around 2022 and 2023. This shows the greater trading at certain periods due to the market volatility or other noteworthy events. The Y-axis indicates the volume, whereas the X-axis shows the years from 2008 to 2024.

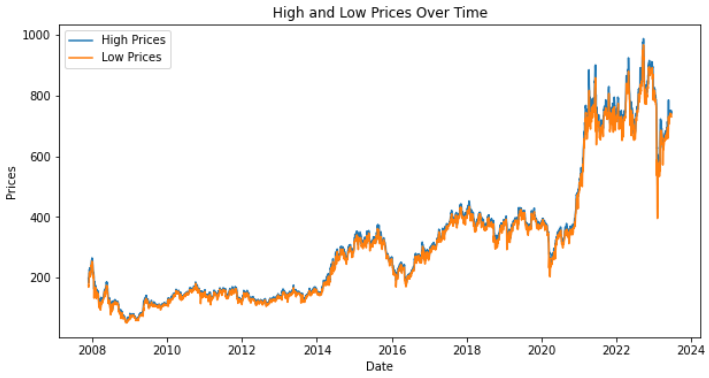

Figure 10 High and low prices over time

Figure 10 displays the high and low stock prices over time from 2008 to 2024. From 2008 to 2016, both the high and low prices gradually increase with periodic fluctuations. After 2020, the difference between high and low prices

becomes more volatile with significant peaks seen in the year 2021–2023 range. This suggests an increase in market fluctuations and possible volatility in the stock prices during those years. The Y-axis indicates the prices and the X-axis shows the years from 2008 to 2024.

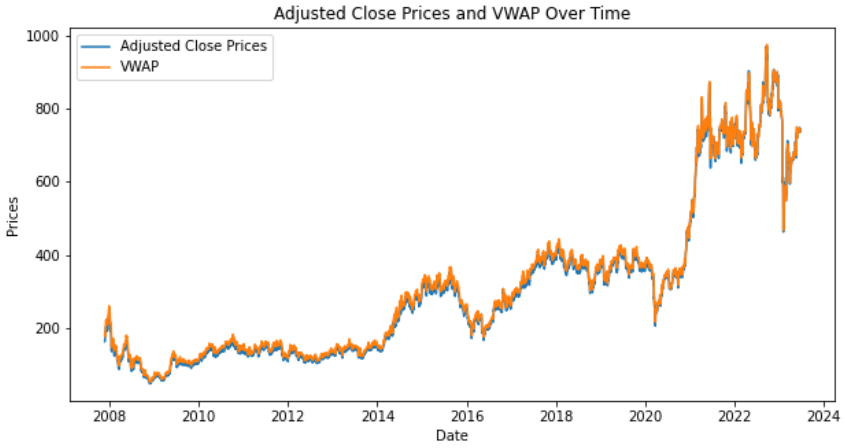

Figure 11 Adjusted close prices and volume weighted average price time

Figure 11 represents the stock from 2008 to 2024 which shows the relationship between adjusted close prices and VWAP. Both the measurements closely track the price movements. Volume weighted average price uses the trading volume to help smooth out the transitory price volatility. Prices climb sharply around 2020, followed by the times of volatility. With dips around 2022-2023, the line demonstrates a positive long-term trend. Years from 2008 to 2024 are shown in the X-axis and Y-axis indicates the prices.

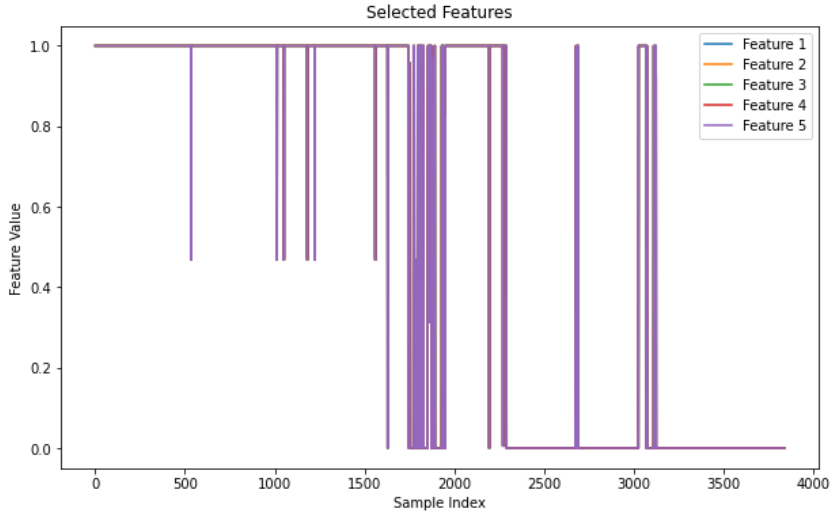

Figure 12 Selected features

Figure 12 represents the sample index on the X-axis and displays the values of five separate features (Feature 1 to Feature 5) across the Y-axis. Each feature behaves differently; some have constant values while the others display discontinuities. Features 4 and 5 prevail; a large proportion of the samples show activity at value 1. Meanwhile, Features 1, 2 and 3 are sparsely distributed across samples and highlight their potential relevance in FS process. The graph highlights the feature relevance variations throughout the dataset.

Table 2 Data preprocessing comparison

Metrics/ Algorithms	**Stock Senti Word Net (SSWN) [2]**	**Pre-Processed Evolutionary Levenberg-Marquardt Neural Networks (PELMNN) [5]**	**Wavelet Transform [12]**	**CMIM**
Accuracy	92.37	93.12	94.67	95.04

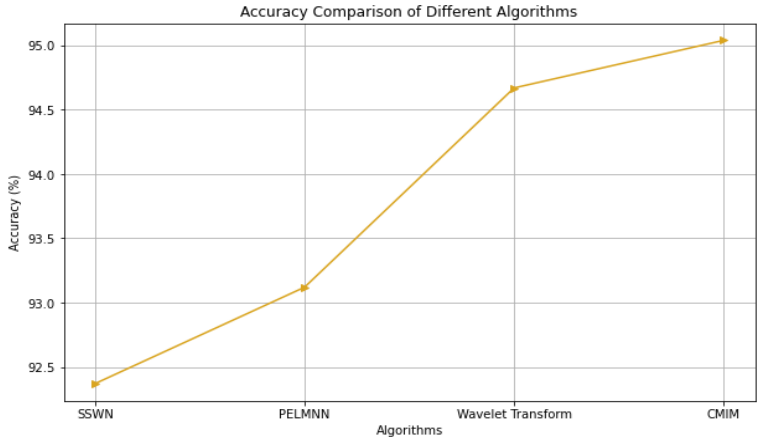

Figure 13 Accuracy comparisons of data preprocessing

Table2 and figure 13 illustrates the accuracy of Data Preprocessing, in which several strategies are used in achieve the certain objective. Stock senti word net has an accuracy of 92.37%, indicating the effectiveness in predicting the stock sentiment. Pre-processed evolutionary levenberg-marquardt neural network has demonstrated the effectiveness of preprocessing approaches by increasing the accuracy to 93.12%. Wavelet Transform approach shows the capabilities in signal processing by increasing the accuracy to 94.67%. Conditional Mutual Information Maximization) algorithm is the most efficient approaches discussed with the best accuracy of 95.04%. This comparison demonstrates the different tactics which provides varying degrees of success in achieving the exact results.

Table 3 Feature Selection comparison

Metrics/ Algorithms	**Accuracy**	**Total number of Feature selection**	**Selected number of feature selection**
Random Forest (RF) [1]	89.00	8	6
Recursive Feature Elimination (RFE) [19]	90.56	8	5
Dandelion Optimization Algorithm (DOA) [12]	91.25	8	4
Stepwise Regression [5]	90.01	8	5
RBM	96.50	8	7

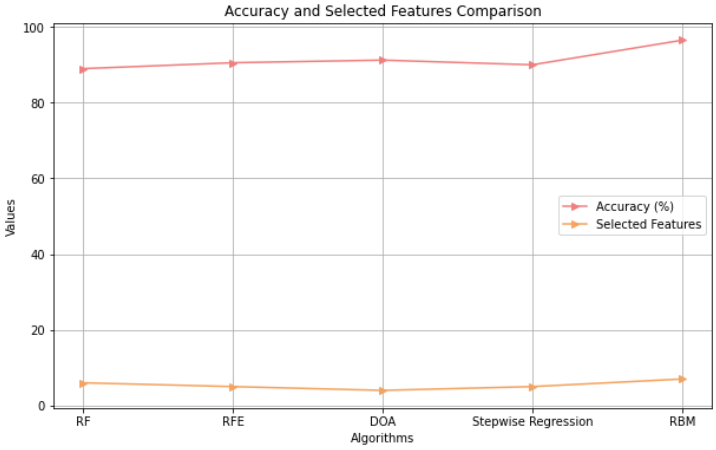

Figure 14 Feature selection comparisons

Along with their FS metrics, table 3 and figure 14 show the comparison of accuracy and FS over several strategies. RF approach achieves 89.00% Accuracy by selecting six out of eight attributes. Recursive feature elimination (RFE)

strategy picks 5 of the same 8 features with an Accuracy of 90.56%, outperforming the other methods. With an accuracy of 91.25%, DOA demonstrates the greater improvement using four characteristics. Meanwhile, stepwise regression (SR) achieves an accuracy of 90.01% by selecting five of the eight criteria. Finally, when seven characteristics are chosen from the same total, RBM outperforms he others with 96.50% Accuracy. This comparison highlights the picking of closely accuracy and the number of features by using various approaches.

Table 4 Performance Comparison of Preprocessing and Feature Selection

Algorithms/ metrics		**Accuracy %**	**Precision %**	**Recall %**	**F-measure %**
Without preprocessing and feature selection	DT [13]	93.02	92.97	92.97	92.91
	RF[1]	93.38	93.11	93.11	93.04
	SVM [23]	93.64	93.38	93.40	93.32
With preprocessing	DT [13]	93.89	93.52	93.55	93.46
	RF [1]	94.02	93.86	93.92	93.84
	SVM [23]	94.31	94.10	94.15	94.03
	CMIM	94.51	94.38	94.40	94.34
With feature selection	DT [13]	94.76	94.56	94.65	94.58
	RF [1]	94.97	94.85	94.85	94.79
	SVM [23]	95.34	95.16	95.16	95.05
	RBM	95.74	95.68	95.72	95.67
With preprocessing and feature selection	DT [13]	95.85	95.72	95.78	95.71
	RF [1]	96.42	96.27	96.30	96.27
	SVM [23]	96.59	96.56	96.58	96.50
	CMIM	96.87	96.85	96.86	96.80
	RBM	97.69	97.36	97.36	97.12

Table 4 shows the comparison of performance measurements like Accuracy, Precision, Recall and F-measure of many approaches in the varied preprocessing and FS situations. Without preprocessing or FS, DT, RF and SVM achieve the accuracies of 93.02%, 93.38% and 93.64% respectively. Preprocessing has improved these statistics somewhat, with SVM achieving 94.31% accuracy. SVM accuracy has increased to 95.34% via FS. Working together, preprocessing and FS yields the best results whereas RBM algorithm achieves the impressive 97.69% accuracy. This suggests that several ML algorithms significantly enhance their performance in classification situations by using preprocessing and FS approaches.

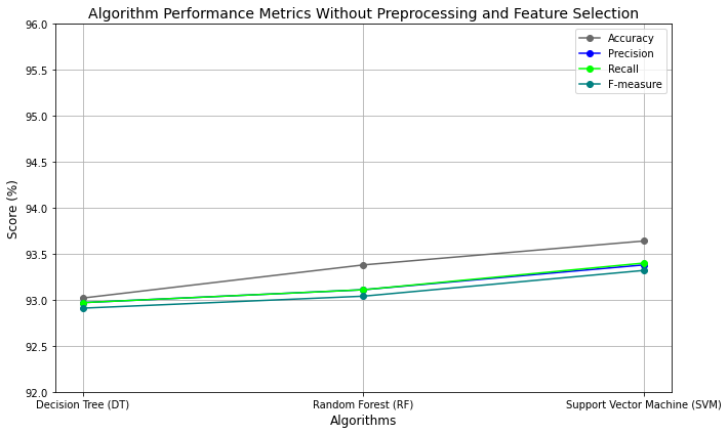

Figure 15 Without preprocessing and feature selection comparison

Figure 15 shows the existing algorithms performance metrics without Preprocessing and FS. Without preprocessing and FS, the efficiency is too low. In this chart, the X-axis shows the DT, RF, SVM algorithms and the Y-axis shows the score values.

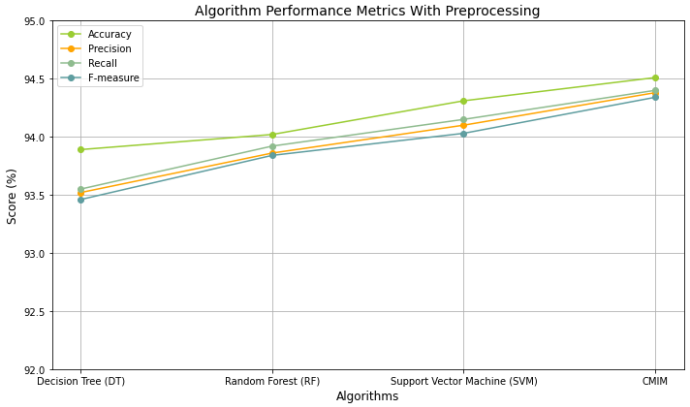

Figure 16 Preprocessing comparison

Figure 16 shows the accuracy, precision, recall and f-measure with preprocessing. In this paper, the Proposed CMIM model is used for preprocessing the data. CMIM performs well in preprocessing than the other algorithms. In this chart, the X-axis shows the DT, RF, SVM, CMIM algorithms and the Y-axis shows the score in percentage.

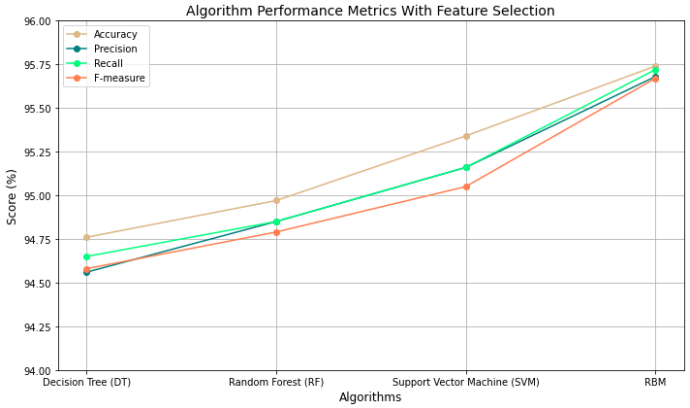

Figure 17 Feature selection comparisons

Figure 17 shows the accuracy, precision, recall and f-measure with FS comparison. In this paper, the proposed RBM method is used for preprocessing the data. Restricted Boltzmann Machine model performs well in FS contrasting to the previous algorithms. In this chart, X-axis shows the DT, RF, SVM, RBM algorithms and the Y-axis shows the score percentage.

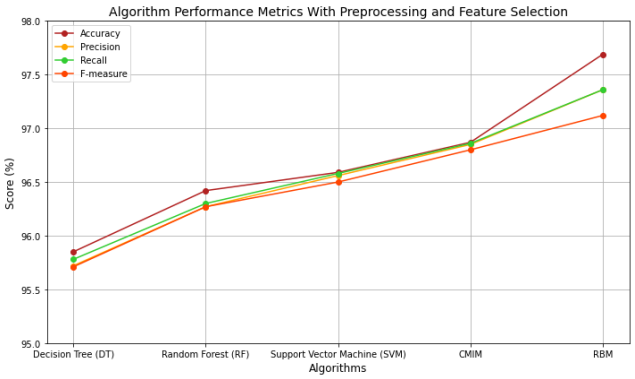

Figure 18 Preprocessing and feature selection comparison

Figure 18 shows the accuracy, precision, recall and f-measure with preprocessing and FS. In this paper, Proposed CMIM and RBM are used for Preprocessing and FS in the given dataset. Compared to the previous algorithms, CMIM and RBM perform well in preprocessing and FS process. In this chart, X-axis shows the various DT, RF, SVM, CMIM, RBM algorithms and the y-axis shows the score percentage.

CONCLUSION

This paper presents Restricted Boltzmann Machines and Conditional Mutual Information Maximization to predict stock prices that aids in understanding the complex market dynamics. Restricted Boltzmann Machines exceeds in detecting hidden patterns and learning the distributions of underlying data, but CMIM identifies the relevant features by maximizing the information shared between input variables and outputs. Together, these strategies form a powerful prediction framework that alters to market complexity. This hybrid technique improves prediction accuracy of 97.69% by integrating ML capabilities and feature relevance. Empirical evidence show that CMIM combined with RBMs outperforms the traditional models in two ways: providing improved generalization and low overfitting. It lays the groundwork for future advances in stock market prediction by utilizing novel ML approaches.

Appendix

S. No	Abbreviation	Description
1	AI	Artificial Intelligence
2	ANN	Artificial Neural Networks
3	BCA	Blood Coagulation Algorithm
4	CD	Contrastive Divergence
5	CMI	Conditional Mutual Information
6	CMIM	Conditional Mutual Information Maximization
7	CNN	Convolutional Neural Network
8	DBN	Deep Belief Network
9	DL	Deep Learning
10	DT	Decision Tree
11	DOA	Dandelion optimization algorithm
12	FS	Feature Selection
13	GRU	Gated Recurrent Unit
14	LM	Levenberg-Marquardt
15	LR	Linear Regression
16	ELM	Extreme Learning Machine
17	MAPE	Mean Average Percentage Error

18	MCMC	Markov Chain Monte Carlo
19	ML	Machine Learning
20	MLlib	Machine Learning library
21	MRF	Markov Random Filed
22	NB	Naive Bayes
23	NN	Neural Network
24	PCA	Principal Component Analysis
25	PCD	Persistent Contrastive Divergence
26	PELMNN	Pre-processed Evolutionary Levenberg-Marquardt Neural Network
27	RBM	Restricted Boltzmann Machine
28	RDAGW	Red Deer Adopted Wolf
29	RF	Random Forest
30	RFE	Recursive Feature Elimination
31	RMSE	Root-Mean-Square Error
32	RNN	Recurrent Neural Network
33	SA	Sentiment Analysis
34	SOTI	Second-Order Technical Indicator
35	SR	Stepwise Regression
36	SSWN	Stock Senti Word Net
37	SVM	Support Vector Machine
38	VWAP	volume Weighted Average Price

Acknowledgement

None

Conflicts of interest

Authors have no conflicts of interest to declare.

REFERENCE

[1]. Altinbas H, Biskin OT. Selecting macroeconomic influencers on stock markets by using feature selection algorithms. Procedia Economics and Finance. 2015;30:22–9.

[2]. Barak S, Dahooie JH, Tichý T. Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese Candlestick. Expert Syst Appl. 2015;42(23):9221–35.

[3]. Chaudhari K, Thakkar A. Neural network systems with an integrated coefficient of variation-based feature selection for stock price and trend prediction. Expert Syst Appl. 2023;219:119527.

[4]. Chen S, Zhou C. Stock prediction based on genetic algorithm feature selection and long short-term memory neural network. IEEE Access. 2020;9:9066–72.

[5]. Gündüz H, Çataltepe Z, Yaslan Y. Stock daily return prediction using expanded features and feature selection. Turk J Electr Eng Comput Sci. 2017;25(6):4829–40.

[6]. Htun HH, Biehl M, Petkov N. Survey of feature selection and extraction techniques for stock market prediction. Financial Innov. 2023;9(1):26.

[7]. Kapinus M, Liashenko K, Ljepava N, Liashenko L, Danylov V. Predicting stock market trends with Python. Grail Sci. 2024;(40):109–16.

[8]. Khedr AE, Yaseen N. Predicting stock market behavior using data mining technique and news sentiment analysis. Int J Intell Syst Appl. 2017;9(7):22.

[9]. Liu Y, Chen Y, Wu S, Peng G, Lv B. Composite leading search index: a preprocessing method of internet search data for stock trends prediction. Ann Oper Res. 2015;234:77–94.

[10]. Meesad P, Rasel RI. Predicting stock market price using support vector regression. 2013 International Conference on Informatics, Electronics and Vision (ICIEV); 2013 May; IEEE. p. 1–6.

[11]. Nti IK, Adekoya AF, Weyori BA. A comprehensive evaluation of ensemble learning for stock-market prediction. J Big Data. 2020;7(1):20.

[12]. Shen J, Shafiq MO. Short-term stock market price trend prediction using a comprehensive deep learning system. J Big Data. 2020;7:1–33.

[13]. Reddy VKS, Sai K. Stock market prediction using machine learning. Int Res J Eng Technol. 2018;5(10):1033–5.

[14]. Torres EP, Torres EA, Hernández-Álvarez M, Yoo SG. Emotion recognition related to stock trading using machine learning algorithms with feature selection. IEEE Access. 2020;8:199719–32.

[15]. Weng B, Ahmed MA, Megahed FM. Stock market one-day ahead movement prediction using disparate data sources. Expert Syst Appl. 2017;79:153–63.

[16]. Zhang X, Hu Y, Xie K, Wang S, Ngai EWT, Liu M. A causal feature selection algorithm for stock prediction modeling. Neurocomputing. 2014;142:48–59.

[17]. Zhong X, Enke D. Predicting the daily return direction of the stock market using hybrid machine learning algorithms. Financial Innov. 2019;5(1):1–20.

[18]. Akhtar MM, Zamani AS, Khan S, Shatat ASA, Dilshad S, Samdani F. Stock market prediction based on statistical data using machine learning algorithms. J King Saud Univ Sci. 2022;34(4):101940.

[19]. Albahli S, Irtaza A, Nazir T, Mehmood A, Alkhalifah A, Albattah W. A machine learning method for prediction of stock market using real-time twitter data. Electronics. 2022;11(20):3414.

[20]. Alotaibi SS. Ensemble technique with optimal feature selection for Saudi stock market prediction: a novel hybrid red deer-grey algorithm. IEEE Access. 2021;9:64929–44.

[21]. Asadi S, Hadavandi E, Mehmanpazir F, Nakhostin MM. Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. Knowl Based Syst. 2012;35:245–58.

[22]. Jagadesh BN, RajaSekhar Reddy NV, Udayaraju P, Damera VK, Vatambeti R, Jagadeesh MS, et al. Enhanced stock market forecasting using dandelion optimization-driven 3D-CNN-GRU classification. Sci Rep. 2024;14(1):20908.

[23]. Javed Awan M, Mohd Rahim MS, Nobanee H, Munawar A, Yasin A, Zain AM. Social media and stock market prediction: a big data approach. Comput Mater Continua. 2021;67(2):2569–83.

[24]. Rasel RI, Sultana N, Hasan N. Financial instability analysis using ANN and feature selection technique: application to stock market price prediction. 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET); 2016 Oct; IEEE. p. 1–4.

[25]. Paramita AS, Winata SV. A comparative study of feature selection techniques in machine learning for predicting stock market trends. J Appl Data Sci. 2023;4(3):147–62.
[26]. Yuan X, Yuan J, Jiang T, Ain QU. Integrated long-term stock selection models based on feature selection and machine learning algorithms for China stock market. IEEE Access. 2020;8:22672–85.
[27]. Singh K, Tiwari R, Johri P, Elngar AA. Feature selection and hyper-parameter tuning technique using neural network for stock market prediction. J Inf Technol Manag. 2020;12(Special Issue):89–108.
[28]. Macedo F, Valadas R, Carrasquinha E, Oliveira MR, Pacheco A. Feature selection using decomposed mutual information maximization. Neurocomputing. 2022;513:215–32.
[29]. Zhou H, Wang X, Zhu R. Feature selection based on mutual information with correlation coefficient. Appl Intell. 2022;52(5):5457–74.
[30]. Wu X, Yuan X, Duan C, Wu J. A novel collaborative filtering algorithm of machine learning by integrating restricted Boltzmann machine and trust information. Neural Comput Appl. 2019;31(9):4685–92.
[31]. Abroshan Y, Moattar MH. Discriminative feature selection using signed Laplacian restricted Boltzmann machine for speed and generalization improvement of high dimensional data classification. Appl Soft Comput. 2024;153:111274.
[32]. Liang Q, Rong W, Zhang J, Liu J, Xiong Z. Restricted Boltzmann machine based stock market trend prediction. 2017 International Joint Conference on Neural Networks (IJCNN); 2017 May; IEEE. p. 1380–7.
[33]. Taherkhani A, Cosma G, McGinnity TM. Deep-FS: A feature selection algorithm for Deep Boltzmann Machines. Neurocomputing. 2018;322:22–37.
[34]. Wang G, Lochovsky FH. Feature selection with conditional mutual information maximin in text categorization. Proceedings of the thirteenth ACM international conference on Information and knowledge management; 2004 Nov; ACM. p. 342–9.
[35]. Cai X, Hu S, Lin X. Feature extraction using Restricted Boltzmann Machine for stock price prediction. 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE); 2012; IEEE. p. 80–3.

Deepa P, MSc, M.Phil, Ph.D Research scholar,Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore. She has around 5 years of experience as Assistant professor in various colleges. She obtained her MSc from Anna University, Chennai, Tamilnadu, and M.Phil in Bharathiar University, Coimbatore, Tamilnadu. She has published papers in Web of Science and UGC CARE. She has attended National and International conferences and presented papers. Her are of interest is Data Mining and Data Analysis.

Dr.B.MurugesakumarMCA, M.Phil., Ph.D., working as Associate Professor and Head, Department of Computer Science at Dr.SNS Rajalakshmi College of Arts and Science. He had over 20+year experience in teaching core computer science subjects for Under Graduate and Post Graduate students. He guided 7M.Phil Scholars. He is guiding 4Ph.D Scholars. He had received many Best Faculty Award, Best Motivator Award, Amazing Gem Award, Bharath EducationExcellence Award. He had published 15+ Research article in Scopus Indexed, Web of Science and UGC CARE. He had presented many research papers in various National and International Conferences. He also acted as subject expert in BOS Member, DC member and Resource Person in various colleges. He had published many patents. He also received many Online Certifications from courseera, udemy, IBM, edX, Kotlin and NPTEL. He has received gold medal in Java NPTEL Course.