Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/120543

HKSVM-DSS: Novel Machine Learning-Based Approach for Decision Support System in Stock Market

M. S. I. Alsalamah

Department of Management Information System and Production Management, College of Business and Economies, Qassim University, Qassim, Saudi Arabia

Received: 10 Mar. 2023, Revised: 22 April. 2023, Accepted: 24 April. 2023.
Published online: 1 May 2023.

Abstract: The stock market serves as an attractive investment venue that draws interest from a broad cross section of people. At the same time, while it continues to be a substantial source of income, it is frequently seen as one of the riskiest investing options due to the fundamental characteristics of the financial industry and several other elements that frequently escape the notice of inexperienced investors. No one can accurately forecast how well a stock will behave in the times to come, although several factors can aid in stock analysis. To determine the ideal moment to buy stocks and the specific stocks to buy, a decision support system (DSS) that incorporates market patterns, economic analyses, and tactics is thus, urgently needed. This study uses machine learning (ML) approaches to handle various issues presented by the assessment of market data. So, using the hyper-tree kernel-adaptive support vector machine (HKSVM) technique, this study introduces an automatic stock DSS to anticipate the top and bottom stock prices in the forthcoming years. The Z-score normalization method is first used in raw trading statistics to retrieve the data without repeated or redundant information. Then, by using the Latent Dirichlet Allocation (LDA) approach, feature extraction is carried out. By offering a reliable and automatic framework for research on stock trading data, the experimental findings and comparisons proved good interpretability and prediction effectiveness for the suggested HKSVM approach.

Keywords: Decision support system (DSS), stock market, trading, machine learning (ML), hyper-tree kernel-adaptive support vector machine (HKSVM).

1 Introduction

The internet is a key marketing tool that can be utilized to increase brand recognition, draw in customers, cultivate trust, and provide online collaboration and communication tools [1]. Due to a variety of factors, Saudi stock market traders frequently confront significant difficulties [2], but primarily, Saudi stock markets have historically been more volatile than those of other industrialized nations like the US, with more variations in return. This makes predictability of stock rates difficult. Chaouachi and Salim [3] applied a cointegration approach called Autoregressive Distributed Lag (ARDL), and Al-Suhaibani and Kryzanowski [4] applied the vector autoregressive (VAR) model to assess the information content in the Saudi stock markets. They reached the conclusion that there is a need for applying machine learning to predict the top and bottom stock prices in the forthcoming years.

Huge wealth is traded on stock exchanges around the world. The market value for worldwide equities exceeded $85 trillion in 2019. Traders have looked for ways to learn more about the companies listed on the market for as long as there have been markets to increase their investment returns. Investors once focused on their competence to recognize market trends, but this is no longer possible due to the scale of the markets and the rapidity at which trades are completed. Though a lot can be learned from simple financial data, but in current history, investment firms have increasingly turned to machine intelligence systems to sift through vast volumes of real-time equities and growth figures for changes [5]. The primary goal of machine learning in the finance sector is to lower the frequency of false positives that a human wouldn't have expected while carrying out stock transactions.

Several traders try to increase their profits or prevent inflation from affecting their savings by investing in the stock market. Unfortunately, the current financial information systems (FIS) place an undue emphasis on disseminating data like the most recent stock price or financial news. They might be given basic stock trading heuristics. For instance, if the 7-day moving average curve crosses and is above the 50-day moving average, a buy signal is generated. However, they still lack the analytical tools necessary to conduct complex analyses.

For investors of all kinds, the stock market is increasingly a preferred investing venue. The most recent information is provided by the current financial information systems. The new architecture for financial information systems is suggested in this study. Investors may use it specifically to develop financial models for predicting stock prices and

*Corresponding author e-mail: mslamh@qu.edu.sa

indexes. In a virtual trading platform on which the performance of the financial models may be measured.

Due to the possibility of financial benefit, the stock market has retained a significant place in the fields of finance, engineering, and mathematics. It is regarded as the top investment outlet since a sizable volume of cash is transacted here. However, there hasn't been agreement on whether the Efficient Market Hypothesis (EMH), predicting is unnecessary to ensure the market works well [6]. In recent years, the typical person's interest in the stock market has grown exponentially. As traders act on the sector to make a gain over their duration, it is not unexpected that assets worth billions of dollars are traded on stock markets each day. A commercial entity, such as a private or corporate investor, might continuously generate larger risk-adjusted returns than the market if they were able to predict the market's response with accuracy [7]. This encourages the development of precise stock market prediction models using machine learning and computational intelligence techniques. Several surveys have been done to attempt to forecast how stock values will change in the future. On the one hand, some individuals subscribe to the EMH and assert that it is impossible to forecast stock values. On the other hand, there is some research that has demonstrated that it is possible to accurately forecast stock values with a reasonably good degree of accuracy provided the appropriate models are used [8]. This study has even concentrated on choosing acceptable functional forms, forecasting approaches, and variables to utilize in its models. In this context, an innovative method of stock price forecasting is based on an approach that takes the form of a time series decomposition of the data series of stock markets.

The stock market has developed into a significant investing activity on a global scale, and large numbers of small individual traders have been drawn by the expansion of the internet and online stock trading [9]. Many inexperienced individual investors find that stock market investments are above their understanding and that it is time to seek the assistance of reputable financial experts. It would be ideal to have an executive decision device that might assist stockholders by giving accurate asset values. In asset prices, decision support tools are crucial for maintaining the quality of investment. Investors strike a balance both greed, the desire to maximize the gain from an investment, and anxiety, the possibility of suffering a significant loss from a poor investment decision. Investors may now measure and analyze risk and diversification because of advancements in financial theory and the growth of data sources.

The researcher proposes a financial DSS that can accurately anticipate stock prices and analyze stock financial health to automatically identify top-performing stocks for risk diversification and high returns relative to market returns. The financial DSS, based on ML architecture, sets up the connection between basic financial variables and the overall market, determines the essential value of the stock, and conducts a thorough financial health and risk analysis based on automatic stock DSS to predict the top and bottom stock prices in the years ahead. Further, the study carries out extensive literature survey, explains suggested technique, the experimental result and finally, the conclusion.

- To get the data without repetitive or superfluous information, basic trading statistics first employ the Z-score equalization procedure.
- The strategy known as Latent Dirichlet Allocation (LDA) is used to carry out feature extraction.
- The hyper-tree kernel-adaptive support vector machine (HKSVM) method is described.

1.1 Previous studies

It is necessary to address the shortcomings of fundamental, technical, and new modelling techniques and this consciousness has spurred several academics to investigate novel stock price prediction strategies. Consequently, more inventive techniques are being used for stock value predictions, and a new type of collective intelligence has arisen. For stock market forecasting and research, some methods make use of supervised machine learning algorithms [10]. This study provides a thorough examination of 57 selected studies and covers the most often cited papers, authors, and co-citation rates. In a comprehensive analysis and review study, more than 50 research articles on stock market forecasting were employed as a result. They essentially classified the selected studies by providing a full explanation of the applicable prediction approaches, the year they were published, the performance metrics, and the software tools [11]. Additionally provided were a detailed description and classification of the stock market forecasting model.

To estimate stock values, several experimental and other studies have been carried out. This is a quick summary of some significant recent developments in this field of study. The effectiveness of prediction performance is analyzed and evaluated using a variety of methods in this assessment [12]. To forecast the trend of the time series data of the MSCI Taiwan Index, techniques including Backpropagation Neural Network (BPNN), LSTM, and Forecast are taken into account.

In another study, by employing information from many international stock markets, an artificial neural network to forecast daily stock exchange index returns was created [13]. The goal is to encourage lucrative trading. Untransformed data inputs are used in a system that relies on univariate neural networks that forecast short-term indexes of the stock market performance.

Another study trains a CNN model to increase the robustness and accuracy of predictions. CNNs have been demonstrated to be incredibly successful in modelling difficult computer vision and image processing issues [14]. This study investigated CNN's predictive capabilities using a complicated multivariate time series of NIFTY 50 index values. In multi-step time series forecasting with multivariate input data with sub-models, they have a distinct sub-CNN model for each of the four input variables [15].

SVM and R Forest Classifiers are used in another method because they are simple to use and can handle a wide range of sorts of characteristics. A type of differencing classifier is the SVM Classifier. Supervised literacy, which is classed as educational facts, is used by SVM. The hyperplanes are what categorize the new data collection. These could be supervised literacy trends that adhere to the relevant bracketing and the purpose is to take competences set of norms from the SVM classifier's settings [16].

Major stock markets such as the New York Stock Exchange, or NYSE, the London Stock Exchange, also known as the NASDAQ, and the Karachi Stock Exchange have studied regression-based approaches. Apple, Microsoft, and Google, the three top corporations, were also considered in an evaluation. The New York Stock Exchange (NYSE) and Nigerian Stock Exchange (NSE) both employ the autoregressive integrated moving average (ARIMA) for short-term decisions and according to Ayala et al. [17], Amman Stock Exchange (ASE) also uses it.

Several scholars define the subject of stock price decisions as a problem for which, besides technical and fundamental analysis, time series forecasting is used. This method employs two sorts of tactics: conventional processes and machine learning techniques [18]. Smoothing, statistical analysis, and regression-based procedures are only a few examples of conventional tactics.

Another strategy combines signal refinement measurements with machine learning methods. As a result, they consider previous methods that combined these disciplines to study financial markets or associated data. A novel method for determining time series fractal dimension is discussed. The difficulty of the response variable under study is frequently used to improve machine learning techniques or to acquire a deeper understanding of the dynamics of the time series information. Additionally, to create predictions for things like the dollar/peso currency value, this method is integrated with neural networks and fuzzy concepts [19].

The latest developments in machine learning (ML) have given rise to fresh, creative, and practical methods. Also, they have made it possible for the creation of sophisticated methods that seek to determine the likelihood of a firm failure by connections between various financial data types and a firm's future financial standing. The literature has employed a variety of ML algorithms and approaches for this problem, including Support Vector Machine (SVM), enhancing methods, analysis of variance, and machine learning [20].

Koukaras et al. [21] provide a strategy for forecasting stock movement using data from StockTwits and SA on Twitter. Originally, they use APIs for extracting tweets and StockTwits market tweets. Following that, they use TextBlob and VADER as well as SA on text. A rule-based SA tool with an SM data lexicon is called VADER. It determines whether an emotion is favorable or negative in addition to returning positive, neutral, or negative values. On the other side, TextBlob is a SA library that gives each tweet opinion words.

The stock market can be described as either regular or irregular depending on the conditions. The term extreme value theory (EVT) refers to a method that may detect extreme changes. The purpose of Napitupulu and Mohamed [22] was to measure the hazards of investing in the stock market.

They use methods tools such as linear regression (LR) and tree-based machine learning (ML) to determine the monthly participant fair values of European stocks based on a total of 21 accounting factors. When LR and ML models are compared, they find that there is a significant amount of variation in the relevance of the factors. When they looked at trading strategies that were based on deviations from fair values, they discovered that ML strategies had much more significant risk outcomes than their basic LR equivalents.

2 Methodologies

It is crucial to develop a decision support system (DSS) that considers market trends, economic assessments, and trading strategies into consideration to determine the ideal time to buy stocks and the particular stocks to buy. Novel machine learning (ML) methods can manage a wide range of obstacles created by the evaluation of market data to identify the most efficient flexible trading strategy. The flowchart for the suggested technique is shown in Figure 1. The stages that make up the entire framework are: pre-processing, feature extraction, DSS, and performance evaluation.

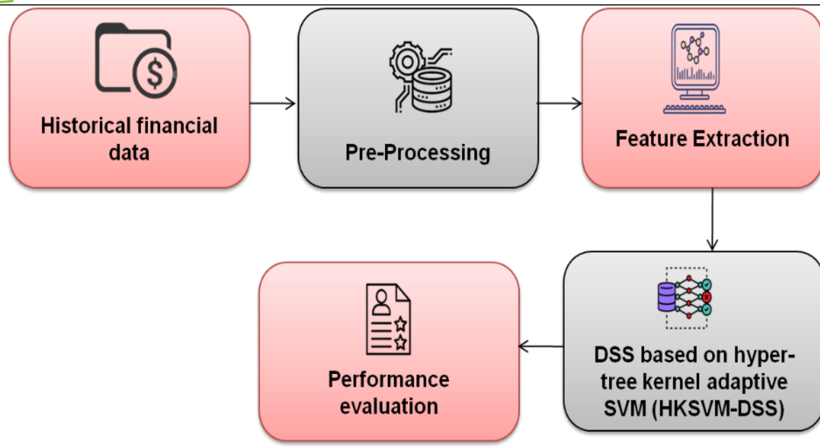

Fig. 1: Flowchart for the suggested technique

2.1 Dataset

This study uses the dataset Trading History, which contains 324,530 transaction observations over 5 months, together with a feature set of 278 factors (trade characteristics) (see Table 1). The dataset is a private dataset from ATS Corporation, and it shows operations from a recent year across several markets [23].

Table 1: Dataset description

Domain	Continuous	Ordinal	Categorical
Client ID	26	113	16
Order status	6	32	5
Instrument definition	-	20	6
Instrument trading Data	7	22	7
Others	29	2	-
Total	58	189	34

2.2 Pre-processing

For pre-processing, we utilize Z-score normalization technique. Pre-processing is a technique used for preparing raw data to be used by a machine learning model. This process is known as data pre-processing.

The obtained data are first pre-processed using Z-score normalization. The Z-score normalization generates a normalized score by analyzing the original data and using its mean and standard deviation. Equation (1) shows that it is feasible to standardize the unstructured data using the z-score variable.

$$K_n' = \frac{K_n - \bar{F}}{std(F)} \tag{1}$$

Where,

K_n' Shows the standardized Z-score values.

Z-scores have been normalized and shown inK_n'.

K_n Identifies row F of the lth column where the value occurs.

$$std(F) = \sqrt{\frac{1}{(O-1)} \sum_{n=1}^{n} (K_n - \bar{F})^2} \tag{2}$$

$$\bar{F} = \frac{1}{n} \sum_{n=1}^{n} K_n \text{ or mean value} \tag{3}$$

The Z-score technique can be used for each row because each row's values are identical, resulting in standard data with a standard deviation of 0. The Min-Max normalization process is analogous to the usage of the z-score to generate a scale from 0 to 1.

2.3 LDA based Feature extraction

Feature extraction includes transforming raw data into numerical features that may be processed to maintain the original data set's content. Features may be extracted using LDA.

i. Latent Dirichlet Allocation

The latent subject Z in a set of document D can be found using the unsupervised method of LDA. Each document is displayed as a probability distribution over topics using LDA, where each topic is a probability distribution over all the words in a particular vocab. Both θ and φ have previous distributions with hyperparameters ∝ and β. Equations (4) and (5) may be used to extract a topic z_{d_i} for each word w_{d_i} in document d, and then return a word w_{d_i}. Repeat steps (6) and (7) N times to produce a document d, where N is the size of document d.

$$\theta_d \sim Dirichlet(\propto) \quad z_{d_i} \sim Multinomial(\theta_d) \tag{4}$$

$$\varphi_z \sim Dirichlet(\beta) \quad w_{d_i} \sim Multinonial\left(\varphi_{z_{d_i}}\right) \tag{5}$$

Using Gibbs Sampling, θ, and φ are the latent topics in documents that may be discovered, and a topic percentage distribution can be predicted for every new document.

2.4 Decision support system based on machine learning

We offer an automatic stock DSS to predict the top and bottom stock values in the upcoming years using the hyper-tree kernel-adaptive support vector machine (HKSVM) method.

The existence of unclassifiable regions can be eliminated using decision-tree-based SVM for multi-class challenges, which also has a stronger generalization capability than the traditional approach. Categorization performance is strongly correlated with the tree-like structure, which in turn, correlates to the split of the feature space (FS).

The formation of the decision tree (DT) is the training process: Each node, except the leaf nodes, represents a hyperplane that divides one category from all the others. One must begin by finding out which hyperplane represents the root of the DT. The hyper-plane separating more than 2 classes at the node linked to the top node must be found, and the procedure must be continued until only one class remains in the separated region. Once the training procedure is complete, the FS will be partitioned into $k-1$ hyper-planes, and there will be no unclassifiable regions. At each node in the tree, the input x is used to determine the value of the decision function, which in turn influences the node below it in a manner determined by the sign of the result of the decision surface at the current node. Conducting this till a leaf node is reached causes input x to be assigned the class that corresponds to the node.

With such a close relationship between the DT's structure and the classifier's classification performance, it is crucial to investigate how to establish the DT's structure because of the training images to achieve the least potential classification error. The overall recognition performance declines if the highest node of the DT has poor performance. Thus, the DT 's leaf nodes should be used to divide up more easily distinguishable groups.

Nevertheless, if the distribution of the classes is disregarded, Euclidean or Mahalanobis distance between class centers, which is used in many separability metrics, may give an inaccurate indication of the degree of separation between classes. The most common measure of separability is the Euclidean distance (ED). The separability among classes shows that the ED between the centers of the two classes is not necessarily accurate for the separability between classes.

Even though the EDs between the centers of the three classes are the same, class k is more separable. Hence, the class distribution plays a role in the between-class separability measure. Including the classes' distribution into the formulation of class separability measure, this research will result in a metric that more precisely indicates class separation.

Let's assume $X_i, i = 1, \dots., k$ are collections of data used for training that belong to class i for a k-class issue.

Let sm_{ij} be the separability metric for $classes\ i$ and j respectively:

$$sm_{ij} = \frac{d_{ij}}{(\sigma_i + \sigma_j)} \tag{6}$$

d_{ij} is the ED between the centers of classes i and j, where $j, i, j = 1, \dots, k, d_{ij} = \left\|c_i - c_j\right\|$

$c_i = \frac{1}{n_i} \sum_{x \in X_i} x$, where c_i is the center of class i according to the data used for training. The sample number for class i is n_i.

A measure of the class distribution, σ_i is the class variance, defined as $\sigma_i = \frac{1}{n_i - 1}\sum_{x \in X_i}\|x - c_i\|$.

Class i and j do not overlap if and only if $sm_{ij} \geq 1$, and classes i and j do overlap if and only if $sm_{ij} < 1$. The larger the sm_{ij} , the more clearly defined are classes i and j, as shown by the formula for sm_{ij} .

Class i separability measure, denoted by sm_i, can be thought of as the smallest of those separating i from the other classes.

$$sm_i = \min_{\substack{j = 1, \dots, k \\ j \neq i}} sm_{ij} \tag{7}$$

In other words, if class i has a high separability measure, then it is possible to distinguish it from the other classes. As a result, the class with the highest separability measure is the class that can be split the most easily:

$$s = \operatorname*{arg\,max}_{i = 1, \dots, k} sm_i \tag{8}$$

Input vector is used to define the separability measure sm_{ij} mentioned above. The input space is transformed into the high-dimensional FS to enhance the separability. Data that is inseparable in the input space can become separable in the FS with the right choice of kernel.

Assume the kernel function $K(.,.)$, the FS as H, and the mapping Φ.

Samples x_1 and x_2 are inputs; after being mapped into FS as H, the ED between them is calculated as follows.

$$d^H(x_1, x_2) = \sqrt{K(x_1, x_1) - 2K(x_1, x_2) + K(x_2, x_2)} \tag{9}$$

Assume that $m_\Phi = \frac{1}{n}\sum_{i=1}^{n} \Phi_{(x_1)}$ is the centroid of the class in the FS as H.

This is the number of class-specific samples, denoted by n. Another expression must be found such that the functional form of the mapping $\Phi_{(x_1)}$ is not required.

Training samples for two classes, $\{x_1, x_2, \dots . x_{n_1}\}$, and $\{x_1', x_2', \dots, x_n'\}$ are mapped into FS. The class centers in H, m_Φ andm_Φ' of FS as H are respectively H and m. In FS, the distance between m_Φ and $m_\Phi'$is denoted as $d^H(m_\Phi, m_\Phi')$

$$d^H(m_\Phi, m_\Phi') = \sqrt{\frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{i=1}^{n_1} K(x_i, x_j) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{i=1}^{n_2} K(x_i, x_j) + \frac{2}{n_2^2}\sum_{i=1}^{n_2}\sum_{i=1}^{n_2} K(x_i', x_i')} \tag{10}$$

Let $d^H(x, m_\Phi)$ be the distance among training images x and class center m_Φ in FS as H, then for the training samples$\{x_1, x_2, \dots . x_n\}$ of a given class,

$$d^H(x, m_\Phi) = \sqrt{K(x, x) - \frac{2}{n}\sum_{i=1}^{n} K(x, x_i)} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K(x_i, x_j) \tag{11}$$

As a result, we can construct the separability measure across classes i and j in FS as H as follows:

$$sm_{ij}^H = \frac{d^H(m_\Phi^i, m_\Phi^j)}{\sigma_i^H + \sigma_j^H} \tag{12}$$

A class's variance in FS is denoted by $\sigma^H = \frac{1}{n-1}\sum_{i=1}^{n} d^H(x_i, m_\Phi)$.

Using the newly specified separability metric, a DT will be constructed.

Consider a DT in which each node corresponds to a hyper-plane that separates one class from the others. The number of hyper-planes to be determined for a problem with k classes is $k-1$, hence the DT contains $k-1$ nodes other than leaf nodes. An algorithm for HKSVM may be obtained by incorporating the between-class separability measure in FS during the DT creation process, which is based on the premise that the most simply separated classes are segregated first. Algorithm 1 indicates the pseudocode for HKSVM-DSS. Assuming X_i , $i = 1, \dots, k$ are active training data sets that belong to class i they would make up the set of active training data X.

Algorithm 1: Pseudocode for HKSVM-DSS

Stage 1: Compute the separability measure in FS $sm_{ij}^H, i, j = 1, \dots k$ and then create a matrix of separability measures.

$$SM^H = \begin{bmatrix} Inf & sm_{12}^H & sm_{13}^H & \cdots & sm_{1,k-1}^H & sm_{1,k}^H \\ sm_{21}^H & Inf & sm_{23}^H & \cdots & sm_{2,k-1}^H & sm_{2,k}^H \\ sm_{31}^H & sm_{32}^H & Inf & \cdots & sm_{3,k-1}^H & sm_{3,k}^H \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ sm_{k-1,1}^H & sm_{k-1,2}^H & sm_{k-1,3}^H & \cdots & Inf & sm_{k-1,k}^H \\ sm_{k,1}^H & sm_{k,2}^H & sm_{k,3}^H & \cdots & sm_{k,k-1}^H & Inf \end{bmatrix}$$

This symmetric matrix has the values $sm_{ii}^H, i = 1, \dots, k$ set to Inf for the simplicity of future computation. Inf denotes that the associated class does not need to be further separated. It is necessary to determine the number of hyperplanes, as indicated by the set variables $\leftarrow\ k$ and t.

Stage 2: Decide which class i 0 may be split the simplest:

$$i_0 = \underset{i=1,\dots k}{argmax} sm_i^H$$

The separability metric of class i wheresm_i^H, reflects how distinct class i is from all other classes.

$$sm_i^H = \underset{\substack{i=1,\dots,k \\ j\neq i}}{min}\, sm_i^H$$

Define $sm_{i_0j}^H \leftarrow Inf, for\ j = 1, \dots, k,$ and $sm_{ii_0}^H \leftarrow Inf, for\ j = 1, \dots, k.$

Stage 3: Utilizing X_{i0} and $X - X_{i0}$ as the trained data, determine the hyperplane $f_{i0,\bar{\imath}0}$ that distinguishes class i_0 from the other classes.

This DT currently has $f_{i0,\bar{\imath}0}$ as its root node.

Stage 4: Modify the active training data set X.

$X \leftarrow X - X_{i_0}$

$t \leftarrow t - 1$

Step 5: Continue if. t>1; else, terminate.

3 Results

The existing methods are Long short-term memory (LSTM), Z-score and multi-layer perceptron artificial neural network (MLP-ANN), attention-based bidirectional long short-term memory network (A-BLSTM) and our proposed method is hyper-tree kernel-adaptive support vector machine (HKSVM-DSS). Using certain performance indicators, the suggested model's efficiency may be assessed. Different criteria can be used to assess a system's success in machine learning.

The ratio of accurately predicted samples tested to all expected test data is called precision. Figure 2 depicts the comparison of precision in the existing and proposed approach. The recommended HKSVM-DSS is 93.65% compared to the current techniques LSTM (Banik et al., [24]), Z-score and MLP-ANN [25], and A-BLSTM [26], which has 90.75 %, 91.65 %, and 91.75 %, respectively. The results demonstrate that compared to current approaches like LSTM, Z-score and MLP-ANN [27], and A-BLSTM [28], the suggested method (HKSVM-DSS) provides a high degree of precision in automatic stock DSS when compared to current practices.

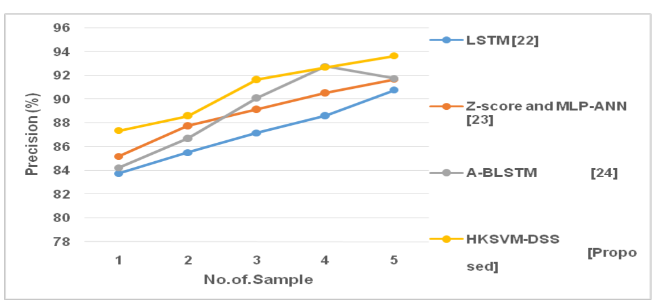

Fig. 2: Precision Results

The F1-score is calculated by combining the weights from the precision and recall scores. Figure 3 displays the performance of the F1-score. The suggested techniques HKSVM-DSS gets 92 % of the F1-score, whereas LSTM [29], Z-score and MLP-ANN [30], and A-BLSTM [31], only get 89.90 %, 90.15 %, and 90.95 %, respectively. Compared to conventional approaches, HKSVM-DSS techniques are greater in F1-score. This illustrates that our suggested approach is the most efficient way to automatically stock DSS.

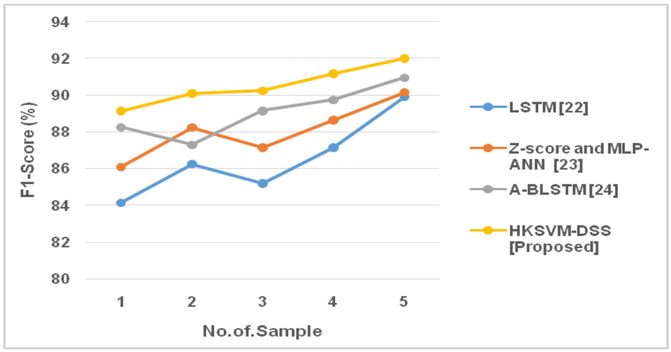

Fig. 3: F1-score Results

The recall is the proportion of true positive samples predicted out of all true positive samples for a given class. The comparison of the recall between the existing approach and the proposed method is shown in Figure 4. The suggested HKSVM-DSS has 92.50 %, which is higher than the current approaches LSTM [32], Z-score and MLP-ANN [33], and A-BLSTM [34], which have 89.15 %, 88.75 %, and 89.25 %, respectively. HKSVM-DSS strategies have a greater Recall rate than conventional approaches to automatic stock DSS.

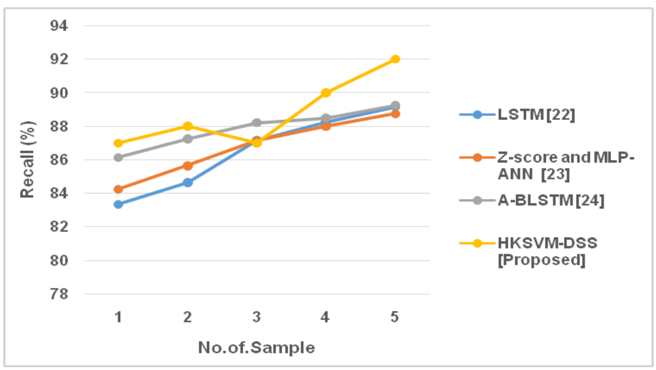

Fig. 4: Recall Results

The proportion of accurately anticipated events to every predicted instance is known as accuracy. The comparison of accuracy using recommended and conventional approaches are shown in Figure 5. In contrast with the current techniques LSTM [35], Z-score and MLP-ANN [36], and A-BLSTM [37], which have accuracy rates of 87.1 %, 88.15 %, and 89.95 %, respectively, the proposed HKSVM-DSS has a 95.5 % accuracy rate. The findings show that the recommended technique, HKSVM-DSS, has a high degree of accuracy for automatic stock-based DSS when compared to current methods.

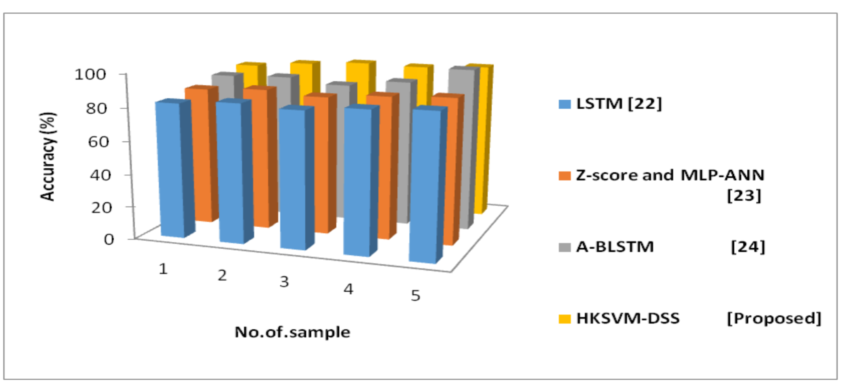

Fig. 5: Accuracy Results

MAE is described as the average of the absolute values of the prediction errors. The comparison of MAE using recommended and conventional approaches are shown in Figure 6. The suggested HKSVM-DSS has a 1.75 % error rate compared to the existing approaches LSTM [38], Z-score and MLP-ANN [39], and A-BLSTM [40] , which have MAE values of 3.50 %, 3.12 %, and 2.85 %, respectively. The results demonstrate that, in comparison to existing techniques, the suggested technique, HKSVM-DSS, has a lower error rate for automatic stock-based DSS.

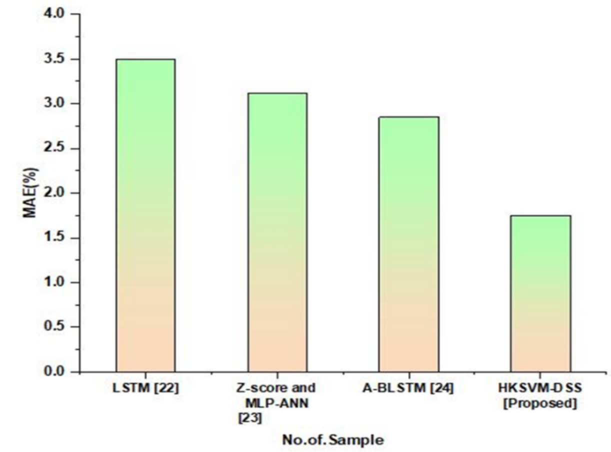

Fig. 6: MAE Results

Root Mean Square Error (RMSE) is a typical method for assessing the error of a predictive model for quantitative data. The comparison of RMSE using recommended and conventional approaches are shown in Figure 7. In contrast with the current techniques LSTM [41], Z-score and MLP-ANN [42] and A-BLSTM [43] which have accuracy rates of 3.75 %, 2.72 %, and 2.50 %, respectively, the proposed HKSVM-DSS has a 1.50 % of error rate. The findings show that the recommended technique, HKSVM-DSS, has a low root mean square error rate to automatic stock DSS when compared to current methods.

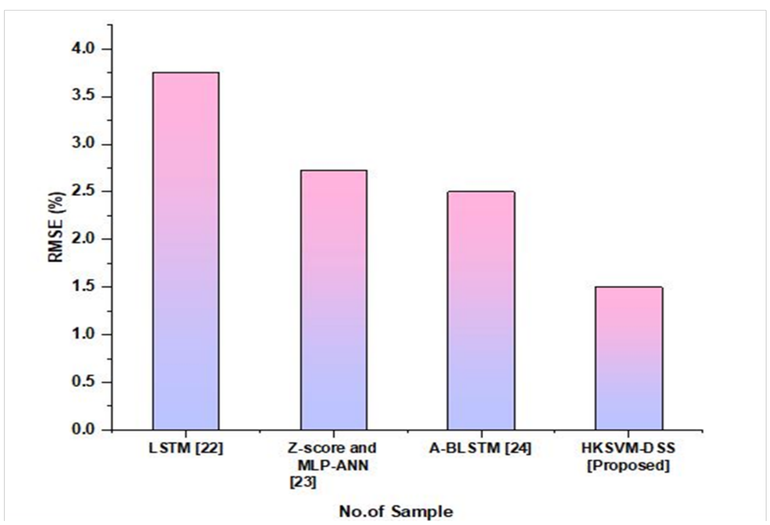

Fig. 7: RMSE results

4 Discussions

When attempting to anticipate an inherently unpredictable event, an LSTM model will perform poorly. By definition, randomness is unpredictable and cannot be consistently predicted [44]. Due to its complete connectivity, MLP has an excessive number of parameters. NN requires very large amounts of prior data. The optimum NN design structure is still unclear [45]. With complicated systems, the outcome and accuracy may diminish. The statistical relevance of the finding is essential.: Extremely careful data design is required and systematically assessed [46]. BLSTM is a significantly slower model, and the training process for it takes significantly more time. It is an extremely time-consuming system that is also challenging to operate in parallel [47]. To overcome these limitations, we utilize HKSVM-DSS for the automatic stock DSS.

5 Conclusions

This study presents ML techniques that can manage a variety of challenges provided by the evaluation of market data to identify the most efficient flexible trading plan. To predict the peak and bottom stock values in the upcoming years, this study develops an automatic stock DSS utilizing the hyper-tree kernel-adaptive support vector machine (HKSVM) method. To recover the data without redundant or repetitive information, the Z-score normalization approach is initially applied to raw trading data. Then, feature extraction is performed using the latent Dirichlet allocation (LDA) technique. The experimental results and comparisons demonstrated strong interpretability and prediction efficacy for the recommended HKSVM technique by providing a trustworthy and automated framework for research on stock trading data. In the future, further technical analysis may be investigated.

6 Recommendations

Based on the findings of the study, the following are recommended:

1. Approaches other than the Z-score normalization approach should be tried to isolate the most effective tool before subjecting data to HKSVM.

2. Sociological assessment should be made beforehand to forecast the long-term effects of applying the HKSVM model.

3. Predictive analysis using HKSVM has been used in other fields such as image recognition. Results from those fields should be carefully analyzed before introducing the model in the stock markets.

4. HKSVM should be tested across stock exchanges to conclusively establish its prediction potential.

Conflicts of Interest Statement

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

[1] B. S. Arasu, B.J.B Seelan., and N. Thamaraiselvan, A machine learning-based approach to enhancing social media marketing. *Computers & Electrical Engineering, **86***, 106723(2020).

[2] M. M. Akhtar, A.S. Zamani, S. Khan, A.S.A. Shatat, S Dilshad and F. Samdani. Stock market prediction based on statistical data using machine learning algorithms. *Journal of King Saud University-Science, **34***(**4**),101940(2022). DOI:10.1016/j.jksus.2022.101940

[3] M. Chaouachi and C. Slim, Current covid-19 impact on Saudi stock market: Evidence from an ARDL model. *Available at SSRN 3636333*(2020).

[4] M. Al-Suhaibani and L. Kryzanowski. The information content of orders on the Saudi stock market. *Journal of Financial Research, **23***(**2**), 145-156(2000).

[5] T. J. Strader, J.J. Rozycki, T.H Root and Y.H.J Huang. Machine learning stock market prediction studies: review and research directions. *Journal of International Technology and Information Management, **28***(**4**),63-83(2020).

[6] S.M. Idrees, M.A Alam. and P. Agarwal. A prediction approach for stock market volatility based on time series data. *IEEE Access, 7,* 17287-17298(2019).

[7] F Zhang. Extreme learning machine for stock price prediction. *The International Journal of Electrical Engineering & Education*, .0020720920984675(2021).

[8] E. Hoseinzade and S. Haratizadeh. CNNPred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications, **129***, 273–285(2019).

[9] S. Patalay and M.R Bandlamudi. Decision support system for stock portfolio selection using artificial intelligence and machine learning. *Ingénierie des Systèmes d Inf. **26***(**1**), 87-93(2021).

[10] N. Rouf, M.B. Malik, T Arif., S Sharma., S. Singh, S Aich, and H.C Kim. Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics, **10***(**21**), 2717(2021).

[11] B. M. Henrique, V.A Sobreiro and H Kimura. Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications,* ***124***, 226-251(2019).

[12] W. X, P.C Fang. W.R. Lan, H.C. Lin Chang, H.Y. Chang, Y.H. Wang. Combine Facebook prophet and LSTM with BPNN forecasting financial markets: The morgan Taiwan index, *in: Proceedings - 2019 International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 20199*(2019).

[13] X. Zhong and D. Enke, Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation,* ***5*****(1)**,1-20(2019).

[14] A. Geron, *Hands-on Machine Learning with Scikit-Learn Keras & Tensorflow*, O'Reilly Publications, USA, (2019).

[15] J. Brownlee. *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery (2017).

[16] M. M. Akhtar, A.S. Zamani, S. Khan, A.S.A. Shatat, S Dilshad and F. Samdani. Stock market prediction based on statistical data using machine learning algorithms. *Journal of King Saud University-Science,* ***34(4)*****,**101940(2022). DOI:10.1016/j.jksus.2022.101940

[17] J. Ayala, M. García-Torres, J. L. V. Noguera, F. Gómez-Vela and F. Divina. Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge-Based Systems,* ***225***, 107119(2021).

[18] S. Raubitzek and T. Neubauer. An exploratory study on the complexity and machine learning predictability of stock market data. *Entropy, 24*(3), 332(2022).

[19] J. Ayala, M. García-Torres, J. L. V. Noguera, F. Gómez-Vela and F. Divina. Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge-Based Systems*, **225**, 107119(2021).

[20] G. Lombardo, M. Pellegrino, G. Adosoglou, S. Cagnoni, P.M Pardalos and A Poggi. Machine Learning for Bankruptcy Prediction in the American Stock Market: Dataset and Benchmarks. *Future Internet,* ***14*****(8)**, 244(2022).

[21] P. Koukaras, C Nousi and C. Tjortjis. Stock market prediction using microblogging sentiment analysis and machine learning. *Telecom,* ***3*****(2)**, 358-378 (2022).

[22] H. Napitupulu and N. Mohamed. A conceptual model of investment-risk prediction in the stock market using extreme value theory with machine learning: a semisystematicsemi systematic literature review. *Risks,* ***11*****(3)**, 60(2023).

[23] R. Rosati, L. Romeo, C.A. Goday, T Menga and E. Frontoni. Machine learning in capital markets: a decision support system for outcome analysis. *IEEE Access,* ***8***, 109080-109091(2020).

[24] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems,* ***239***, 107994(2022).

[25] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems,* ***159***, 113814(2022).

[26] J. Long, Z. Chen, W He, T. Wu. and J. Ren. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing,* ***91***, 106205(2020).

[27] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems, 159*, 113814(2022).

[28] J. Long, Z. Chen, W He, T. Wu. and J. Ren. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing,* ***91***, 106205(2020).

[29] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems, 239*, 107994(2022).

[30] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems, 159*, 113814(2022).

[31] J. Long, Z. Chen, W He, T. Wu. and J. Ren. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing, **91***, 106205(2020).

[32] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems, **239***, 107994(2022).

[33] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems, **159***, 113814(2022).

[34] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems, 239*, 107994(2022).

[35] J. Long, Z. Chen, W He, T. Wu. and J. Ren. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing, **91***, 106205(2020).

[36] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems, **239***, 107994(2022).

[37] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems, **159***, 113814(2022).

[38] J. Long, Z. Chen, W He, T. Wu. and J. Ren. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing, 91*, 106205(2020).

[39] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems, **239***, 107994(2022).

[40] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems,* **159**, 113814(2022).

[41] J. Long, Z. Chen, W He, T. Wu. and J. Ren. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing*, 91, 106205(2020).

[42] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems, 239*, 107994(2022).

[43] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems, 159*, 113814(2022).

[44] S. Banik, N. Sharma, M. Mangla, S.N Mohanty.and S. Shitharth. LSTM based decision support system for swing trading in stock market. *Knowledge-Based Systems, **239***, 107994(2022).

[45] U. Khan, F. Aadil, M.A. Ghazanfar, S. Khan, N. Metawa, K. Muhammad, I Mehmood and Y. Nam, A robust regression-based stock exchange forecasting and determination of correlation between stock markets. *Sustainability, **10***(**10**), 3702(2018).

[46] D. Wu, X. Ma and D.L. Olson, Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems, **159***, 113814(2022).

[47] J. Long, Z. Chen, W He, T. Wu. and J. Ren. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing, **91***, 106205(2020).