
Saba Zahid¹*, Hassan Mujtaba Nawaz Saleem²

¹Ph.D. Scholar, Institute of Business Management & Administrative Sciences, The Islamia University of Bahawalpur.
²Associate Professor, Institute of Business Management & Administrative Sciences, The Islamia University of Bahawalpur.

*Corresponding author: s21bmgmt4m01034@iub.edu.pk

Abstract

Predicting the stock prices’ movement has been considered a tedious task due to the stock market's volatile behavior. This study is an attempt to predict stock prices based on 15 trading factors by applying deep learning algorithm support vector machine (SVM) on daily price data of three stocks of COVID-19, collected from secondary sources and processed through Python software. The finding of this study suggests that the linear kernel gives 53%, the polynomial kernel gives 46% and the radical basis function (RBF) kernel gives 62% validation accuracy, which shows the RBF kernel predictive ability is highest than others during high volatility. The current study contributes to the stock return literature-originated machine learning algorithms during the unprecedented market condition. The finding of this study is helpful for investors and traders who calculate stock return in their portfolio diversified decisions, regulators, and policymakers in the formulation of their regulations & implementation decisions while the market condition is unprecedented. Thus, the support vector machine (SVM) algorithm offers an accurate prediction of stock return to investors who invest during high volatility.


Introduction:

Sudden stop in a large steam of activities of national and global economies due to a respiratory syndrome known as covid-19 (Shafi, Liu et al. 2020). Enterprises experience a significant impact
of this outbreak (Dyduck, Chudziński et al. 2021). Develop and under developing both economies suffering in a bad situation by imposing lockdown (Alsamhi, Al-Ofairi et al. 2022). UNIDO (2020) state SMES in Pakistan have negative impact of covid-19. However, impact density varies on the basis of firm size, sector, nature of business and the technology utilization. In Pakistan 90% business show decline in their revenue, while 2% firm show positive attitude respect to their profit. Pakistan economy that suffering already in a recovery position hit badly due to covid-19 pandemic (Amir, Farooq et al. 2022).

Predicting the movement of stock prices has been the focus of the study by experts in a broad variety of disciplines, including economics, history, finance, mathematics, and computer technology. The stock market’s volatility makes it challenging to use conventional statistical methods, such as time-series analysis or regression (Gupta, Nel et al. 2023). Many proprietary information models have been developed by banking firms and traders to outperform the market, but few of them have been consistently successful (Kolte, Roy et al. 2023). However, the allure of stock forecasting lies in the fact that even a modest gain of a few percent might result in a million-dollar windfall for these organizations.

Stochastic linear time-series models, such as the ARIMA, have long been the primary focus of many forecasts (Y.A. L. B. Gianluca Bontempi 2015). However, no-linear models such as ARCH generally have lower prediction errors due to the volatility underlying the movements of stocks and other commodities, making liner strategies unsatisfactory (Zhang, Shan et al. 2009). In recent years, analysts have looked to big information and machine learning methods from the area of computer science to better predict stock prices. These make use of computing capacity to develop mathematical and statistical hypotheses. To “figure out” the answer to a problem, machine learning algorithms analyze existing data. Algorithms and high-frequency trading procedures employed by financial firms are also based on big information and machine learning approaches.

In this paper, we use Support Vector Machines, a machine learning technique known as SVM with their three different kernels to check the accuracy of price prediction during high volatility. Specifically, we want to utilize SVM at time t to forecast whether the price of a certain stock will be greater or lower on day t + 1. The time t represents the time in which market conditions around the world, resulted in a global recession of Covid-19 because the economic impact of this pandemic is still unknown. In this study fifteen factors are used to predict stock prices, these fifteen factors are further divided into four main categories, which include volume and price factor, valuation factor, risk factors, and scale factors. Pakistan stock market’s three stocks are analyzed. We placed
the algorithms on fifteen variables, of three stocks including the food, personal care products, automobile assembler, and pharmaceutical industry’s daily data. These values are determined by utilizing the final prices of each stock each day between 2020 and 2022. In this research, we investigate whether or not this past information can aid in predicting the future course of prices while the market suffers from high volatility. In addition to this analysis, the accuracy of 3 different kernels of SVM is named polynomial kernel, linear kernel, and RBF kernel. According to the Efficient Markets Hypothesis (EMH), prices should fluctuate at random in response to information in the market.

**Background information:**

**Stock Market Efficiency:**

The market efficiency hypothesis, a popular topic of economic study, states that stock prices already represent all relevant information and are, therefore, unexpected (Kang, Zong et al. 2023). The EMH states that stock prices will act randomly as they react to fresh information. They are unpredictable if they solely react to fresh information because predictable movement would indicate that information was not represented by market rates, the fact that shares move randomly is evidence of market efficiency (Chen 2023, Sheth and Shah 2023).

The weak, semi-strong, and strong forms of this theory exist. The majority of studies have found support for the somewhat strong interpretation. This version asserts that stock prices accurately represent all publicly accessible information; nevertheless, non-public data may be exploited to make biased predictions of future earnings, and because of this, insider trading rules are so strict (Chhajer, Shah et al. 2022). The EMH is challenged, however, by a few features of the market. Anomalies in the market describe these situations. The short-term momentum in stock prices was found by Jagadeesh and Titman. Recently increasing stock keep increasing while decreasing ones keep decreasing. The pattern would seem to imply that stock values in the future are somewhat predictable, which would go counter to the EMH (Jegadeesh and Titman 1993). Seasonal patterns may be seen in the stock market as well. Through analysis of data spanning decades, Jacobsen and Zhang discovered that trading techniques may take advantage of seasonal patterns of high cold returns and low summertime returns to generate profits (Ben Jacobsen 2012, Jacobsen 2012, Jacobsen and Zhang 2014). If the EMH were accurate, then it would be impossible to forecast the direction in which future stock prices with a probability higher than 50%. Simply put, one should not have any more luck predicting whether prices will rise or fall in the future than they would be using a random number generator (Yang and Ma 2022, Zhang and Wu 2022, Kurani,
Doshi et al. 2023). However, as shown, machine learning approaches may use momentum as well as other price patterns to anticipate price direction with higher than 50% accuracy, as seen by the research covered therein (Krollner, Vanstone et al. 2010, Chen 2022, Khashanah and Shao 2022).

**General Machine Learning:**
Both traditional and modern machine learning methods fall into one of two broad categories. The first kind is called supervised learning, and it requires training data in the form of examples that have been labeled to show what the desired outcome should be for each feature set included in the example. A dataset’s characteristics and labels are sent to algorithms as training examples, and the algorithm is tasked with using this information to predict the labels for a second dataset (test data)(Tiphimmala 2014, Illa, Parvathala et al. 2022). In contrast, unsupervised learning relies on unlabeled samples in the feature set. As a rule, algorithms will attempt to “cluster” your data into meaningful categories.

Regression and classification issues are subsets of supervised learning. In classification problems, the output is discretely labeled as one of many possible categories, but in regression problems, it takes on a continuous range of values. In this study, we frame the issue of predicting stock prices as one of categorization. To forecast whether a stock’s price will be higher or lower than the current day’s price days from now, the feature set of its recent price fluctuations and momentum is combined by calculating the stock return and making standardization of this return by comparing actual return.

**Previous Research:**
Artificial neural networks have been the primary focus of the study in the field of machine learning for predicting (ANN) (Krollner, Vanstone et al. 2010, Al-Najjar 2022). Artificial neural networks (ANNs) simulate the structure of the nervous system by using a network of interconnected nodes that stand in for individual neurons (input layer, processing layer, and output layer, etc.). the input is fed into the ANN, and then the output is computed using the weights that have been assigned to the various connections. As it learns, the computer will adjust the weights based on the patterns it finds in the training data. When the data does not contain drastic shifts, as shown by Kar, ANNs may achieve high accuracy (Bregu, Shosha et al., Kar 1990, Kar 1990, de Oliveira, Zárate et al. 2011, Fagner Andrade de Oliveira 2011, A. Victor Devadoss 2013). Although Patel and yalamalle both agree that ANNs may achieve prediction accuracy of just over 50%, they warn that the nonlinear and temporal changes in stock market data make it challenging for even the most sophisticated methods to make accurate forecasts (Patel and Yalamalle 2014).
Recent research on the topic has employed another technology known as support vector machines in addition to or as an option for ANNs. SVMs may commit classification mistakes inside training data to reduce total error across test data, while ANNs are models that strive to decrease classification errors within the training data. A key benefit of SVMs is that it finds a global optimum, while the neural network may only discover a local optimum (Kim 2003, Krollner, Vanstone et al. 2010, Khashanah and Shao 2022, Kurani, Doshi et al. 2023).

By using a support vector machine (SVM) model, Kim achieved a prediction accuracy of up to 57% (using test data outputs), which is much over the 50% criterion. According to the findings of survey research on stock prediction done by the Shah, SVM is the most effective machine learning model. Kim’s conclusion is in line with his 60% success record in making predictions. Since most current research has integrated SVMs, this is the approach we employ in our study (Lv, Guo et al. 2022, Sheth and Shah 2023).

**Support Vector Machines:**

For binary classification, Support Vector Machines are among the top options. As a result, they establish a dividing line between the two sets of data, with most of the points in one set lying on one side of the line and most of the points in the second lying on the other. Think of a characteristic vector \( x = (X_1, \ldots, X_n) \) with \( n \) dimensions (Milke 2022, Koranteng 2023). A linear border (hyperplane) may be defined as follows:

\[
\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n = \beta_0 + \sum_{i=1}^{n} \beta_i x_i = 0
\]

The other the total of the items in one set will be more than zero, while the sum of the elements in the other set will be less than zero. Labeled instances have \( y \), where \( y \) is the label. We use the notation \( y = 1, 1 \) to organize things.

\[
Y = \beta_0 + \sum \alpha_i y_i x(i) * x
\]

The recast the hyperplane equation, we may use inner products. The best hyperplane is the one where we have the greatest possible separation between the plane and any given point. The margin represents this extra space. When dividing the information, the decision boundary hyperplane (MMH) is optimal. Since this may not be a perfect separation, we may include error variables 1…n and constrain their total to be less than some limit B. The most important consideration when
choosing a hyperplane is its proximity to the border. The points themselves are called support vectors, and the classification hyperplane that assigns each support vector to one of two categories is called a support vector classifier (SVC).

While the idea behind the SVC is similar to that of ordinary least squares (OLS), another well-known linear regression model, the two models maximize different values (Ghimire, Guéguen et al. 2022). The OLS calculates the sum of squares of the residuals, or the distances from each piece of data to the fit line. On the other hand, the SVC is concerned only with the support vectors, and it uses the inner product to optimize the distance from the hyperplane to the support vector. Additionally, in SVC, the inner products are weighted by their labels, but in OLS, the weighting is the square of residuals. As a result, SVC and OLS are two distinct strategies for tackling the issue.

The restriction of SVCs to linear bounds is a major limitation. To address this problem, SVMs use non-linear kernel functions to translate the inputs to a higher-dimensional space, where they may be classified linearly. Classifications that are linear in the higher dimensional space will be non-linear in the lower dimensional space. The SVM ramps the input to higher dimensions by substituting a more generic kernel function $K$ for the inner product. So, for a support vector machine, $y = 0 + i y_i K(x(i), x)$ describes this relationship.

**Linear kernel:**

A linear kernel is a feature space corresponding with its original feature space. In linear kernels no shadowy involved in n-dimensional space, they direct translate Euclidian distances. Linear kernel support vector machine performs well only on very simple problems. The algorithm performance is worse due to the more indecipherable boundaries between class clusters.
Polynomial kernel:
The polynomial regression feature space represents the polynomial feature space corresponding; both are the same. In a polynomial kernel each x feature is copied over to x², x³.........., xⁿ up to the degree, and then draw the feature space by separating the hyperplanes. Concerning their polynomial nature, they draw them in two dimensions. The polynomial kernel of the support vector classifier has unique degree parameters. The higher the degree the resulting space is more overfitted. Hyperplanes remain continuous and boundless while generating space and separating them.

RBF kernel:
The most popular kernel choice in support vector machines is the RBF kernel. RBF is known as the radical basis function. This function is used to approximate the other functions and is used default in sklearn. The kernel is translated in complex math but basically, they draw the most complex boundaries that one’s available to Sklearn. Kernel density estimation is through behind kernel smoothing which includes KDE plotting that is popular in EDA machine learning.
Model Creation and Evaluation Method:
This study uses SVM by focusing on its three different kernels for price forecasting. This study aims to find the best kernel that is used SVM model to see an opportunity to predict stock price by applying the historical stock data through the pandemic data and subsequent recovery period.

The Framework of Stock Prediction:

Data Collection and Time Frame:
Economic conditions greatly deteriorated during covid-19. Unemployment increases that create a downturn in the Labor market, which enhances the depression of the great pandemic. This study wants to analyze the historical factors that predict stock prices while the market is unstable. This study also wants to see which kernel of the SVM model performs well during high volatility. The study used historical data on covid-19 by focusing on 3 stocks from 3 different sectors. Focusing on three different sectors as opposed to a single sector allows us to test the model on different nature companies making our results relatively standardized. Automobile assembler (Atlas Honda limited), food and personal care (Fuji Foods limited), and pharmaceuticals (abbot laboratories limited) sector stock included in this study. Stock price data were obtained from the Pakistan stock exchange website.

SVM model:
Three kernel functions of SVM is used in this study, to check the accuracy of stock prediction, which include linear kernel, polynomial kernel, and radical kernel.

The function of the linear kernel in SVM.
$K(x, x_k) = \sum_{j=1}^{n} (x_{ij}x_{kj})$

The function of a polynomial kernel in SVM.

$K(x, x_k) = (\sum_{j=1}^{n} a + x_{ij}x_{ij})^b$

The function of a radical kernel in SVM.

$K(x_i, x_k) = \exp(-\frac{1}{\delta^2} \sum_{j=1}^{n} (x_{ij} - x_{kj})^2)$

The Python library that is used for this SVM model is numpy and pandas.

**Feature selection:**

This study uses 15 trading factors that are mainly divided into four broader categories, which include volume and price factors, valuation factors, risk factors, and scale factors. (Table 1)

**Method**

Table 1 describes how each feature is calculated. To calculate the trading, feature this study looks at everyday trading data from 2020 through March 2022.

<table>
<thead>
<tr>
<th>Trading Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Previous day closing price</td>
</tr>
<tr>
<td>▪ Trading volume</td>
</tr>
<tr>
<td>▪ Opening price</td>
</tr>
<tr>
<td>▪ Average price</td>
</tr>
<tr>
<td>▪ Closing price</td>
</tr>
<tr>
<td>▪ Lowest price</td>
</tr>
<tr>
<td>▪ Turnover</td>
</tr>
<tr>
<td>▪ Highest price</td>
</tr>
<tr>
<td>▪ P/B ratio</td>
</tr>
<tr>
<td>▪ P/E ratio</td>
</tr>
<tr>
<td>▪ P/S ratio</td>
</tr>
<tr>
<td>▪ Up &amp; Downs of Rs</td>
</tr>
<tr>
<td>▪ Turnover Rate</td>
</tr>
<tr>
<td>▪ Total Market Capitalization</td>
</tr>
<tr>
<td>▪ Total Equity</td>
</tr>
</tbody>
</table>

Table 1
Results:

The training accuracy and validation accuracy of three stocks concerning three different kernels of the SVM model is shown in Table 2.

<table>
<thead>
<tr>
<th>Stocks</th>
<th>Training Accuracy</th>
<th>Validation Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear Kernel</td>
<td>Polynomial Kernel</td>
</tr>
<tr>
<td>ABOT</td>
<td>0.582</td>
<td>0.584</td>
</tr>
<tr>
<td>FFL</td>
<td>0.367</td>
<td>0.407</td>
</tr>
<tr>
<td>HCAR</td>
<td>0.467</td>
<td>0.408</td>
</tr>
</tbody>
</table>

Table 2

The training accuracy of ABOT stock is 58% and validation accuracy is 62% by using a linear kernel of the SVM model. The training accuracy of ABOT stock is 58% and validation accuracy is 56% by using a polynomial kernel of the SVM model. The training accuracy of ABOT stock is 62% and validation accuracy is also 62% by using the radicle kernel of the SVM model. That means the radical kernel of SVM is more competent to predict the volatility during unstable market condition. The density of all 15 factors respect to the ABOT model show in appendix figure 1. The individual factor predictive capability of ABOT each factor shown in appendix figure 2. Training accuracy of all factors respect to three kernels of ABOT train algo shown in appendix 3. The feature engineering that is used in train and test model respect to factors pairing to evaluate the predictive accuracy of ABOT shown in appendix 4. The price movement of ABOT stock is shown in the below figure 1.

Figure 1
The training accuracy of FFL stock is 36% and validation accuracy is 44% by using a linear kernel of the SVM model. The training accuracy of FFL stock is 40% and validation accuracy is 38% by using a polynomial kernel of the SVM model. The training accuracy of FFL stock is 38% and validation accuracy is 50% by using the radicle kernel of the SVM model. That means the radical kernel of SVM is more competent to predict the volatility during unstable market condition. The density of all 15 factors respect to the FFL model is shown in appendix figure 5. The individual factor predictive capability of FFL is each factor shown in appendix figure 6. Training accuracy of all factors respect to three kernels of FFL train algo shown in appendix 7. The feature engineering that is used in train and test model respect to factors pairing to evaluate the predictive accuracy of FFL shown in appendix 8. The price movement of FFL stock is shown in the below figure 2.

![FFL Close price.](image)

**Figure 2**

The training accuracy of HCAR stock is 46% and validation accuracy is 41% by using a linear kernel of the SVM model. The training accuracy of HCAR stock is 40% and validation accuracy is 41% by using a polynomial kernel of the SVM model. The training accuracy of HCAR stock is 62% and validation accuracy is also 52% by using the radicle kernel of the SVM model. That means the radical kernel of SVM is more competent to predict the volatility during unstable market condition. The density of all 15 factors respecting the HCAR model show in appendix figure 9. The individual factor predictive capability of HCAR each factor shown in appendix figure 10. Training accuracy of all factors respect to three kernels of HCAR train algo shown in appendix 11. The feature engineering that is used in train and test model respect to factors pairing to evaluate the predictive accuracy of HCAR shown in appendix 12. The price movement of HCAR stock is shown in the below figure 3.
Discussion:
The innovation of the stock selection model proposed in this study by using fifteen trading factors as a predictive factor to evaluate the forecasted price of stock while the market condition is highly volatile. Trading factors contain the historical information of stocks daily. To prove the effectiveness of the SVM kernel, the proposed stock selection model compared three kernels of SVM. The finding of this study suggests that the linear kernel gives 53%, the polynomial kernel gives 46% and the radical basis function (RBF) kernel gives 62% validation accuracy. The results show that the radical kernel of SVM that adopted a model of machine learning used in this study performs better than Linear and Polly kernels. This study also supports the idea that machine learning model plays an important role in stock selection while the market condition is unprecedented. Additionally, historic data is an optimized predictor of stock forecasting by using machine learning during high volatility.

Future Work:
This study used daily data, but the market swings its mood every minute or second, next study by looking at intra-day trading data enhances the accuracy of the model by considering intra-day data to create robust that capitalized on sudden changes.

Machine learning model accuracy greatly depends on feature selection. In this study, we consider trading factors, and future work may consider economic factors like inflation, and GDP growth rate to enhance the accuracy.

Another scope of future growth is to apply the model in other sectors. In this study, we focus on only 3 sectors and individual stock from each sector. We can apply our model to other sectors and check the model’s predictive ability.
Appendix:

ABOT (Abbot Laboratories (Pakistan) Limited – Pharmaceuticals):

Appendix Figure 1

Appendix Figure 2
Appendix Figure 3

Appendix Figure 4

FFL (Fauji Foods Limited – Food & Personal Care Products):

Appendix Figure 5

Appendix Figure 6

113

Appendix Figure 9

Appendix Figure 10
References:


Bregu, E., et al. "THE FUTURE OF SECURITIES MARKET IN ALBANIA."


