Maqsut Narikbayev University

International School of Economics

MASTER'S DISSERTATION

«Application of Machine Learning algorithms for AIX Index movements predictions»

7M04124 - *«Finance»*

May 2024

Written by: Yerzhigit Ichshanov

Supervisor: Yerzhan Tokbolat

Astana, 2024

Abstract

There have been many efforts to predict stock market trends by numerous scholars across a wide range of areas utilizing different methods and applications. This thesis examines in detail how advanced machine learning techniques can be used to predict the indices of the Astana International Exchange (AIX). As a relatively new but highly impactful stock exchange in Central Asia, the AIX is important in the regional financial landscape. The accurate forecasting of AIX index trends is critical for stakeholders such as investors, fund managers, and policy makers, as it allows them to take informed decisions and obtain better understanding of market fluctuations. The paper evaluates the performance of different machine learning models in projecting the trend performance of AIX indices and investigates the consequences of these predictions for investment strategies and wider market movements. For the learning and validation of machine learning models, the research uses historical data on AIX index share prices for the period of 2023 year. The study shows that all used models may give sufficient outcomes in forecasting the dynamic of AIX index. The findings demonstrate that in comparison with other models, ARIMA can become useful instrument for forecasting price dynamic on AIX and can be used by investors and traders to take well-grounded conclusions in their investment strategies.

Key words: Machine Learning, AIX Index, Financial Prediction, ARIMA Model

Table of Contents

List of tables

List of illustrations

Introduction

The Astana International Exchange (AIX) is a major player in the global financial ecosystem, acting as a major trading hub for a wide range of financial instruments and securities in Kazakhstan. Being able to predict the fluctuations of the AIX index is essential for a number of market participants, including investors, traders and financial analysts. It has a critical role to play in strategic decision-making and risk management approaches that are relevant to Kazakhstan's financial landscape. Although traditional financial modelling methodologies allow for in-depth analysis of market trends, their ability to capture the complex, non-linear characteristics common to financial data can be slightly limited. In this regard, artificial intelligence (or AI) is one of the core solutions that taking place around the world. AI and, specifically, its constituent Machine Learning (or ML) has a great deal to contribute to the global financial world, as this area is the most accessible and easy to use with a wealth of information. It is therefore not unexpected that both AI scholars and the economic world are expressing interest in using this approaches and ML techniques for a different applications, ranging from financial risk mitigation to stock price forecasting.

Over the recent past, the use of machine learning (ML) algorithms has been increasingly acknowledged as a powerful instrument for improving forecasting accuracy in financial markets. Machine learning techniques can identify complex patterns and relationships hidden in large and complex data sets and thus help to predict market behavior more accurately. Among the many machine learning techniques available, Linear Regression, Time Series analysis and Support Vector Machines (SVMs) have proven to be particularly effective in the specific context of predicting the movement of the AIX index.

Linear Regression plays a fundamental role in predictive modeling by identifying a linear relationship between the explanatory variables and the response variable. In the particular situation of forecasting the movement of the AIX Index, Linear Regression helps to identify linear relationships between various elements such as historical values of the AIX Index, trading volumes, and key economic indicators. By identifying the coefficients in the regression model, it is possible to understand the magnitude and direction of the effect of each explanatory variable on the AIX index [1, 3].

Time series analysis methodologies are carefully developed to evaluate sequences of data in chronological progression, making them extremely efficient for predicting trends in the AIX Index. Cutting-edge models such as the Autoregressive Integrated Moving Average (ARIMA) and Seasonal Decomposition of Time Series (STL) are skilled at identifying and interpreting temporal correlations, emerging trends, and cyclical patterns in AIX Index datasets. By carefully analyzing the historical values of the AIX Index and identifying inherent temporal patterns, Time Series Analysis facilitates the creation of sophisticated models. These models can expertly track and predict both the short-term fluctuations and long-term development of the AIX Index [2, 4].

Support Vector Machines (SVMs) are complex controlled learning models known for their ability to represent challenging data relationships. The main goal of SVMs is to determine an optimal hyperplane that clearly separates different classes or provides accurate continuous predictions. In the area of forecasting AIX Index trend prediction, SVM is especially good at navigating high-dimensional feature spaces and identifying nonlinear correlations between the predictor variables and the AIX Index. By applying the kernel trick, SVM extends the input features into a higher-dimensional domain, allowing for the creation of non-linear decision planes, thereby significantly improving the accuracy of predictive outcomes [5, 6].

In this investigation, our objective is to assess the performance of distinct machine learning approaches—specifically Linear Regression, Time Series Analysis, and Support Vector Machines—in accurately forecasting the fluctuations of the AIX Index. Through the application of historical data pertaining to the AIX Index, in conjunction with relevant predictive factors, we intend to develop and critically appraise models that are equipped to reflect the evolving dynamics of the financial sector in Kazakhstan. Through empirical analysis and methodological rigor, we endeavor to shed light on the strengths, limitations, and comparative performance of these machine learning techniques in forecasting AIX Index movements, thereby contributing to the broader discourse on computational finance and predictive modeling in emerging markets.

Literature Review

The application of machine learning in finance is a rapidly growing area of research, with numerous studies exploring various aspects of this integration.

Recently, the application of machine learning in the financial sector was studied by Rashid Husain and R. Vohra (2017), where the authors surveyed existing and prospective applications of machine learning in finance, discussing how Big Data technologies can be applied in this sector. They highlight the leading paths for machine learning applications in finance [15].

Other researchers Hamed Ghoddusi, Germán G. Creamer, and Nima Rafizadeh (2018) critically review the literature on machine learning applications in energy economics and finance. They discuss applications in areas like energy price prediction, demand forecasting, and risk management, noting the popularity of SVM, ANN, and GAs in these studies [16].

Sophie Emerson et al. (2019) focus on machine learning applications in investment processes, such as return forecasting and risk modeling. They evaluate current literature, identifying key themes and technologies in quantitative investing [17]. While, Derek Snow (2019) focuses on the use of machine learning for financial event prediction, such as earnings surprises and bankruptcy predictions, highlighting the advancements over traditional linear models [18].

Based on the above works, the use of Machine Learning (ML) techniques in the field of finance assumes an important role. Whereas, by implementing ML methods and techniques, AIX index forecasting can be utilized. The challenge of forecasting the Astana International Exchange (AIX) index is of primary concern to stakeholders such as investors and policy makers. The incorporation of machine learning algorithms into analyzing the dynamics of financial markets has become a growing trend in several past years. Various machine learning techniques including Linear Regression, Time Series Analysis, and Support Vector Machines (SVM), provide some interesting possibilities to perform efficient forecasting of AIX index movements.

Through their study on the property market in Astana, Zhantileuov et al. (2023) conducted a comparative analysis of various regression algorithms that involved linear regression, decision trees, random forest, XGBoosting and CatBoosting. It has been shown by the research that although leading edge methods such as XGBoosting and CatBoosting provide high precision, linear regression, despite its comparatively poorer performance, has proven to be more effective in summarizing data across various scenarios [7].

From their research work, Patel et al. (2015) used a number of machine learning algorithms, involving Support Vector Machines (SVM), to forecast stock fluctuations and price indices in the Indian market. It was shown by their study that algorithms such as random forest show increased performance when paired with certain forms of data representation, thereby emphasizing the key role of data preprocessing in improving the accuracy of predictive models [8].

The other research study by Patel et al. (2015) have designed a novel two-stage fusion approach that combines techniques such as Support Vector Machines (SVM) in a machine learning framework focused on predicting stock market indices. By their study, they emphasize the great potential of fusion models in improving the precision of financial market predictions [9].

In a 2012 research, Hanias et al. examined the implementation of neural networks for predicting the performance of the Athens Stock Exchange index, proposing a viable approach that can be adjusted for forecasting the AIX index. Continuing this line of research, Ayyıldız (2023) studied a number of machine learning techniques involving decision trees, random forests, SVMs and neural networks for their performance in predicting stock market indices. Hi Through his study, he showed that artificial neural networks, along with logistic regression and SVM, have an accuracy exceeding 70% [10]. Such results are consistent with the potential of neural networks in stock market forecasting illustrated by Hanias et al. (2012) [11]. In addition, the paper of Güresen et al. (2011) on evaluating neural network models for forecasting NASDAQ Stock Exchange index prices gives insights that may be useful in developing models for the AIX index [12]. All collectively, both of these works highlight the crucial importance of neural networks in improving the accuracy of stock market forecasts, which has considerable implications for the AIX index.

The most recent research by Sagadatova et al. (2023) used time series prediction techniques to forecast electricity consumption patterns in Astana. The application approach and results of this study suggest a fundamental framework that can be effectively used to forecast changes in the AIX index [13].

The academic study on forecasting the Astana International Exchange Index makes extensive use of various predictive methodologies, with a special attention to machine learning algorithms. These studies provide important findings and strategies suitable for precise prediction of the AIX index, hence delivering important recommendations for investors and policy makers in their decision-making processes.

Data Collection and Preprocessing

The process of gathering and preparing data is a pivotal aspect in examining the Astana International Exchange (AIX) Index. It is a vital step that consists of a thorough search for the required financial information and its careful preparation for analysis. This diligence is essential to guarantee the precision and efficiency of the subsequent forecast modelling.

In the perspective of an AIX index, the data collection process usually involves obtaining both historical and current financial information. It involves a number of critical financial indicators such as stock prices, trading volume, total market value of companies and other important economic indicators.

There are a wide range of financial indicators covered by the AIX Index data. It is very important to collect not only real-time data for operational assessment, but also a historical data, which is crucial for analyzing trends and creating forecasting models.

On the website of the Astana International Exchange (AIX), one can find stock quotes of various Kazakhstan-based companies, including but not limited to KEGOC, Air Astana, Kaspi Bank, and Halyk Bank [14].

Model Selection and Training

This part of the research focuses on critically analyzing a number of machine learning approaches, such as Linear Regression, Time Series Analysis, and Support Vector Machines. The purpose is to identify and contrast their distinct features in terms of prediction accuracy.

Linear Regression

One of the widely used machine learning methods is linear regression method, which can be used for analyzing and modelling relationships between different variables in finance. It is very useful to make some analysis of financial indicators based on single or more variables.

Regression analysis is one of the most widely used statistical methods. It is employed to construct a mathematical model based on experimental data.

Regression analysis is employed to formulate a mathematical model of an object or phenomenon from experimental or observational data. These models depict precise mathematical correlations between the behavior of an object or the traits of a phenomenon and the factors that affect them.

Regression analysis is a statistical technique used to model a target variable based on independent predictors. This method is widely employed for prediction purposes and for exploring causal relationships between variables. The choice of regression model varies according to the number of independent variables and the nature of the relationship between the independent and dependent variables.

Simple linear regression is a form of regression analysis where there is one independent variable, and a linear relationship exists between this independent variable (x) and the dependent variable (y). In this approach, we aim to fit a line that most accurately represents the data points. This line is described by the linear equation The line can be modelled based on this linear equation: $y = a0 + a1 * x$.

To determine the unknown parameters of the regression equation (a0 and a1), the least squares method is commonly employed. This technique offers unbiased estimates of the parameters.

The primary objective of the linear regression algorithm is to identify the optimal values for parameters a0 and a1. To facilitate understanding of this algorithm, it is essential to explore two key concepts integral to linear regression.

The main formula of linear regression with only one independent variable X can be written to find the dependent variable Y. The formula looks as following:

$$Y = a_0 + a_1 X + \beta$$

- Y is the dependent variable (e.g., stock price)
- X is the independent variable (e.g., economics indicator)
- a_0 is the y-intercept of the regression line
- a_1 is the slope of the regression line, representing the change in Y for a one-unit change in X.
- β is the error term, accounting for the variability in Y not explained by X.

Multiple regression method is enhanced form of the previous model. When there are several independent variables, the multiple regression might be used to obtain the dependent variables. The dependent variables for different independent variables can be found by using the below form:

$$Y = a_0 + a_1 X_1 + \ldots + a_n X_n + \beta$$

In some cases, these models show superiority over more complex models, especially in scenarios characterized by a limited number of training examples or sparse data sets [19]. Linear regression is popular because of its simplicity and easiness during the The popularity of linear regression is due to its simplicity and ease of interpretation. This type of simple machine is vital in assessing the correlation between dependent and independent variables. Moreover, it is widely recognized in statistical models for predictive analysis [20].

Linear regression is easy interpretable due to its direct correlation between dependent and independent variables. According to research of Montgovery et al., this quality allows to understand relationships of the data stronger [41]. Furthermore, the simplicity of the linear regression helps people to use this method to the various fields without any working experience with machine learning techniques [42]. In addition, linear regression model is computationally efficient. It requires less computational resources compared to other machine learning instruments. These above facts speed up the computation process and enables efficient handling of huge database [43].

Linear regression has many advantages. Some of them have been discussed already. However, there are some disadvantages of this type of method. The first limitation of linear regression is the linear correlation between dependent and independent variables. According to Fox (2016), in real-life situations interrelationships between function and variables are not linear. It leads to biased and imprecise projections. It happens when the independent and dependent variables have non-linear patterns [44]. Furthermore, the limited flexibility makes linear regression inefficient when the data sets have high level of nonlinearity [3].

As it was discussed before the linear regression method has notable advantages. It is interpretable, simple and computationally efficient. Its performance is dependent on compliance with underlying assumptions. Nevertheless, there might be occurred some limitations of this method when the data is complicated and has nonlinear interrelation among the variables. It potentially reduces the precision of forecasting future values.

Time Series Analysis

Based on the study of various machine learning methodologies, the following study focuses on the importance and practical utility of time series analysis. This analytical technique is key to subjects as diverse as finance, economics, business administration and environmental science. Time series analysis implies the careful investigation and interpretation of data gathered at regular periods of time. Its main usefulness comes from its capacity to predict future patterns and events based on trends identified in previous data.

A time series is a sequence of data points recorded or monitored during successive time frames. However, this data can take many shapes and forms, such as daily stock market prices, monthly sales figures, annual gross domestic product (GDP) statistics, or hourly weather readings. A characteristic of a time series is that it is arranged chronologically.

Fundamental Components of Time Series

In general, time series data are decomposed into four key components::

- The original key component is the tendency, which is the steady directional movement of a series over a significant time interval, which can appear as upward, downward, or steady patterns.
- The next element relates to seasonalisation, which refers to recurring patterns that occur at regular periods of time. An instance of this kind of seasonality would be variations such as an increase in ice-cream production during the hotter months.

- Other components include elements of cyclicality, defined by variations that happen at different periods of time and are frequently dependent on wider economic or contextual forces.
- Time series data sets typically contain elements of stochasticity, especially when affected by noise or unpredictable outside factors. Although the main purpose of time series analysis is to identify and simulate the patterns behind the data, it also recognizes the existence of random variations that make a contribution to the total volatility found in the series. It is important that the statistical methods used in time series analysis frequently include techniques to distinguish among systematic patterns and stochastic fluctuations in a data set.

Methodological Approaches in Time Series Analysis

Descriptive Techniques:

- Time series data frequently contain elements of chance or stochasticity, especially because of the existence of background noise or unpredictable external effects. Although the primary purpose of time series analysis is to detect and simulate patterns in the data, it also acknowledges the presence of random movements that influence the total volatility found in the data series. The statistical techniques applied in time series analysis often include methods that allow the distinction of systematic patterns and random variations in a data set.
- Splitting the series into parts to identify and estimate tendency, seasonal and non-regular variables separately.

Statistical Forecasting Models:

- The first model is Autoregressive (AR). The future values can be predicted according to the previous past values.
- In the Moving Average (MA) Models, the models exploit the association between an observation and the residual errors resulting from the application of a moving average model to lagged observations.
- There are well-known models which is called Autoregressive Integrated Moving Average (ARIMA). This method combines Autoregressive (AR) and Moving Average (MA) models and incorporates differencing to transform the time series into a stationary state.
- ARIMA models have similar models with some improvements. These models are called Seasonal ARIMA (SARIMA) models. There are included seasonal elements.
- One of the statistical models is Vector Autoregression (VAR) Model**.** In this model, multiple time series variables are written in one line equation.

Machine Learning Techniques:

- There are some machine learning models such as Random Forests, Gradient Boosting Machines, where complex non-linear dependencies are captured in the data.
- Other deep learning techniques such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks can be used when large time series data sets are available.

Time series analysis provides many advantages. For one thing, it is crucial in uncovering the dynamics of time-ordered data sets, allowing for a deeper understanding of patterns, tendencies and irregularities [21]. By employing this method, one can predict and forecast future trends with accuracy, leveraging historical patterns as a basis for projection. Such prediction ability is especially important for subjects such as economics, finance, and environmental sciences [22]. In addition, time series analysis allows the study of difficult relationships in complicated systems, such as financial markets or weather patterns [23]. Furthermore, it allows to separate deterministic regular fluctuations from stochastic ones, which is an essential task in areas such as marine science and meteorology [24].

However, Time Series Analysis is not without its limitations and disadvantages. Many of the approaches that have been developed in the past have been developed for the analysis of single-dimensional data, but their suitability in multidimensional scenarios is restricted. The emerging approaches based on network science aim to alleviate these constraints, even though they are still under development [25]. In addition, when faced with indeterminate or inaccurate monitored data, ordinary time series models may not produce the best results, leading to the necessity of applying customized techniques [26]. While the prediction of systematic components of time series such as trends, cycles and seasonal fluctuations is feasible, the prediction of random variables appears to be difficult, which limits predictability in specific settings [22].

Autoregressive Integrated Moving Average (ARIMA)

As it was mentioned above, the ARIMA model is a well-known tool in studying time series data, particularly for predicting changes in financial markets, such as stock indices like the AIX (Astana International Exchange Index). Therefore, this section will detail how the ARIMA model is used to predict the movement of AIX, drawing on academic and real-world research.

ARIMA consists of three main parts: Autoregression (AR), Differencing (I), and Moving Average (MA) [31]. The AR part looks at how a variable in the present is related to its own past values. The autoregressive element simulates the relationship for an observation and a number of lagged measurements from the same time series. Mathematically, it can be expressed as:

$$\mathrm{Y_t = c + \varphi 1 Y_{\{t-1\}} + \varphi 2 Y_{\{t-2\}} + \ldots + \varphi p * Y_{\{t-p\}} + \varepsilon_t}$$

Where

$\mathrm{Y_t}$ –is the current value of the time series at time t,

c is a constant,

φ 1, φ 2, ..., φ p are the autoregressive variables

ε_t is the error term

The autoregression component represents the time dependence in the data, enabling the model to consider the effect of past measurements on future values.

Differencing helps make the data stable over time [32]. The concept of stationarity is underlying in many time series models, including ARIMA. Differentiation involves the computation of variance between successive measurements, which can be expressed as:

$$\Delta Y_t = Y_t - Y_{\{t-1\}}$$

A differentiated series (ΔY_t) should be ideally stationary, that is, its mean, variance and autocorrelation structure remain constant over time.

While the MA part shows the relationship between an observation and the errors from a moving average model applied to past observations, indicating short-term fluctuations around the average [33].

From a mathematical perspective, it can be formulated as:

$$Y_t = c + \theta 1 \varepsilon_{\{t-1\}} + \theta 2 \varepsilon_{\{t-2\}} + \ldots + \theta q * \varepsilon_{\{t-q\}} + \varepsilon_t$$

Where

θ1, θ2, ..., θq are the moving average components

$\varepsilon_{\{t-1\}}$, $\varepsilon_{\{t-2\}}$..., $\varepsilon_{\{t-q\}}$ are the error terms from earlier time steps.

The moving average parameter represents short-term fluctuations and occasional noise in the data, by providing a mechanism for smoothening out irregularities.

In applying ARIMA to the movement of the AIX index, we use a systematic process based on empirical research and methodological foundations. First, previous market data about the AIX index were obtained, usually covering daily or hourly values for a particular period of time. Afterwards, the data are cleaned by correcting any missing data or strange outliers to make certain that the data is valid [34]. This step is important because it helps ensure the data is reasonable and is stable over time.

After that, selecting the right metrics for the predicting model can be started. Figuring out the best values for these parameters is key to ensuring the prediction model perform well. A significant set of parameters is provided by (p, d, q), where:

- p stands for the index of the autoregressive component,
- d stands for the index of differencing,
- q stands for the index of the moving average component.

This is accomplished by analysis, frequently using statistical instruments such as autocorrelation and partial autocorrelation functions to identify patterns that are underlying the data [35]. Subsequently, estimation techniques, such as maximum likelihood estimation, are then applied to train an ARIMA model using a carefully trained dataset [36]. Both validation and verification have a vital role in assessing the effectiveness and suitability of the ARIMA model. For this purpose, approaches such as k-fold cross-validation or splitting the dataset into learning and testing subsets are applied [37]. The evaluation rubrics such as mean absolute error (MAE), mean squared error (MSE) and root mean square error (RMSE) determine the variance between the predicted and real values, providing an understanding of the precision and stability of the mode [32].

Forecasting and analysis involves the application of the verified ARIMA model to generate projections of upcoming flows of the AIX index, giving valuable assistance to investors and analysts in taking evidence-based decisions concerning portfolio management and risk reduction scenarios [33].

In reviewing the strong and weak points of the ARIMA model, it is essential to recognize its benefits and constraints. The ARIMA model has a number of advantages for forecasting the progress of the AIX index, in particular, it is able to catch the time dependence and can incorporate historical data into forecasts [2]. In addition, its interpreting power provides an easier understanding of the key patterns affecting market trends. Nevertheless, ARIMA also has some restrictions, such as its dependence on assumptions of linearity and stationarity. Therefore, these assumptions can create problems in the efficient modeling of fast-moving financial markets that have fluctuations in both volatility and nonlinear dynamics [36].

Support Vector Machines (SVM)

One more precious instrument in the field of machine learning is Support Vector Machines (SVM), which are often used for predicting and analyzing financial indices due to their reliability and ability to deal with confusing data. SVM algorithms allow analysts to predict potential stock market movements by thoroughly examining historical market data and discovering patterns that may suggest potential future price trends.

Within the framework of forecasting stock market behavior, support vectors (SVMs) work in a number of different modes:

• Selecting Features: SVM models utilize historical market data incorporating different market financial indicators such as price, volume, volatility and technical metrics. These selected characteristics aim to effectively capture market dynamics and trends.

• Training: In the training phase, the SVM model learns to associate input features with certain target labels that present the preferred output categories. By optimization, the SVM customizes its settings to both minimize the forecast errors and maximize the accuracy.

• Categorization: After learning, the SVM model can classify new data items into different groups based on the identified attributes and arguments. In the context of the stock market, the SVM classifies upcoming price movements, providing an understanding of prospective market trends.

• Forecasting: By leveraging its ability to classify, the SVM model produces forecasts of future stock price movements on the basis of present market circumstances and input characteristics. These forecasts help investors in making well informed choices about trading strategies and in managing their portfolios.

• Evaluation and Improvement: The prognostic efficiency of the SVM model is assessed with validation data to identify its precision and reliability. When required, the model is improved by correcting hyperparameters, choosing attributes, or updating the learning data to improve the forecasting performance.

Support vector machines (SVMs) function by building a hyperplane that efficiently divides data points which belongs to various categories in a space determined by a set of features. The goal is to find the hyperplane with the largest indentation, which is the maximum distance between the nearest data points, known as support vectors, from each category. With this algorithm, the approach is developed to obtain the best possible prediction performance and to provide stability when faced with new, unknown data.

From a mathematical perspective, SVM is designed to solve the following optimization problem:

$$min_{w,b} \frac{1}{2} ||w||^2$$

subject to the constraints:

$$y_i(w \times x_i + b) \geq 1 \quad for\ i = 1, \dots, N$$

where

- w is the weight vector
- b is the bias term
- x_i represents the feature vector of the i-th data point
- y_i is the class label of the i-th data point (either -1 or 1)

- N is the number of data points
- $||w||$ denotes the Euclidean norm of the weight vector
-

The decision function of SVM can be expressed as:

$$f(x) = sign(w \times x + b)$$

$$where\ x\ is\ a\ new\ data\ point\ to\ be\ classified$$

Support Vector Machines (SVM) show significant promise in financial forecasting due to their effectiveness in modeling complex, nonlinear relationships and capturing nuanced market dynamics, which is particularly advantageous in financial markets characterized by nonlinear and ever-changing behavior. Indeed, this ability is based on SVM's strong theoretical foundation, which is based on structural risk minimization principles, which allows it to summarize new data well and reduce the risk of overfitting [38].

In addition, SVM exhibits its adaptiveness in finding different types of finite data, involving high dimensional datasets. The ability to utilize different kernel functions such as linear, polynomial or radical kernel basis function allows it to efficiently catch different types of data and adapt to different market conditions [39].

In spite of its strengths, the SVM experiences issues, in particular computational complexity, especially when dealing with large data sets. Such complexity can overload computational resources and may demand a considerable amount of time to work through. In addition, SVM runtime performance is strongly affected by parameter selection, such as the choice of kernel function and regularization parameter, which requires careful parameter tuning to achieve optimal prediction performance [40].

In conclusion, Support Vector Machines (SVM) are emerging as powerful tools in financial forecasting and analysis, offering the ability to model nonlinear relationships and manage the diverse data patterns found in financial markets. However, it is critical to recognize the computational requirements of the SVM and the importance of parameter optimization for its effective use in financial applications.

Data Processing

In this investigation there were chosen historical market data on four issuers: Bank CenterCredit JSC (CCBN), KazMunayGas JSC (KMG), KazAtomProm JSC and Halyk Savings Bank of Kazakhstan Joint Stock Company (HSBK), which are traded on Astana International Exchange.

Joint Stock Company "Bank CenterCredit" issued 188,029,035 ordinary shares priced at 228.45 KZT each. This information is listed in the AIX data under the symbol "CCBN". The fluctuation in the prices of ordinary shares throughout the month of March is depicted in Figure 1.

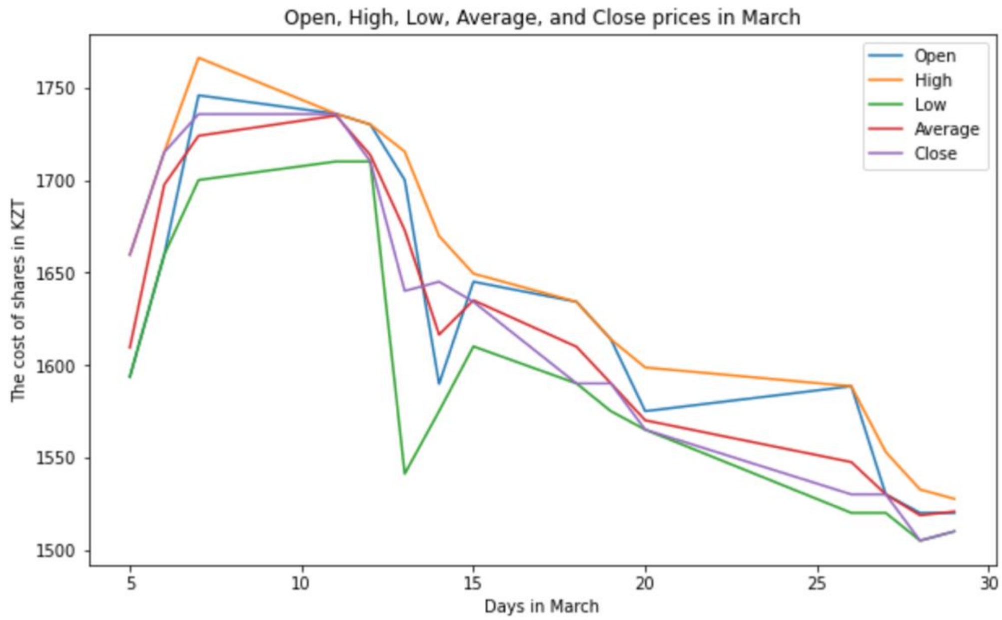

Figure 1. Ordinary share prices for the March in 2024 [27].

The subsequent Figure 2 illustrates the ordinary share prices of "KazMunayGas" JSC in March 2024. The overall trend of the cost exhibits a certain pattern, albeit with variations in open, close, high, low, and average prices. It's important to note that the AIX exchange remained closed on Saturdays and Sundays. Additionally, in March, the exchange was closed for holidays on March 20th and resumed operations on March 26th.

The ordinary share prices of Joint Stock Company "KazAtomProm" (KAP) in March exhibit a different pattern compared to the previously mentioned companies (refer to Figure 3). The fluctuations in the share price of this company are not as pronounced as those observed in other companies, except for one instance. On March 12th, 2024, the company experienced its lowest share price, with the stock prices closing at a minimum value.

The following Figure 4 illustrates the prices of common shares of Joint Stock Company "Halyk Savings Bank of Kazakhstan" (symbolized as "HSBK" on the AIX site).

As depicted in the figure, prices reached their peak on March 14th under various conditions. Subsequently, prices gradually declined, followed by a period of relative stability until the end of March.

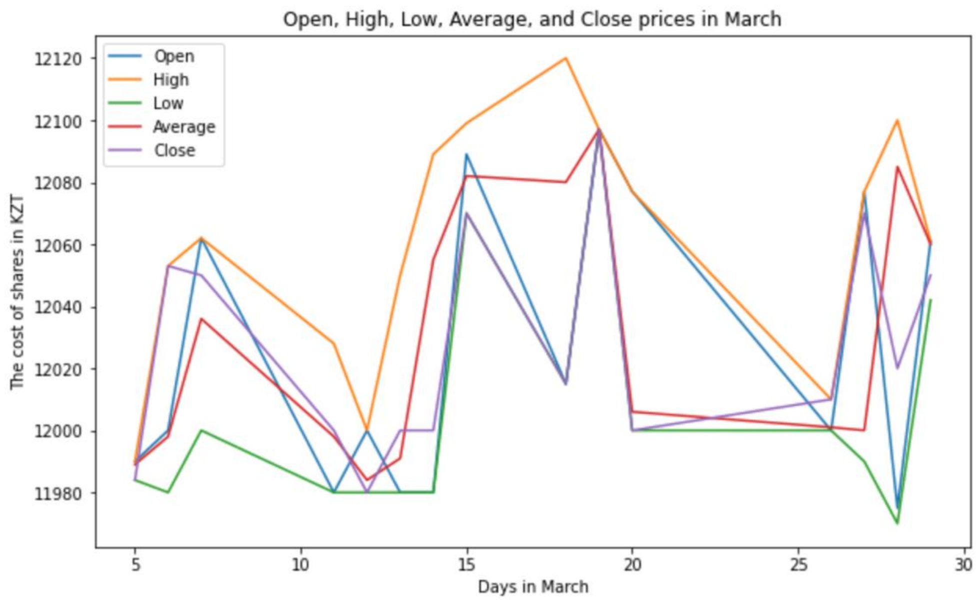

Figure 2. Ordinary share prices of “KazMunayGas” JSC in March in 2024 [28].

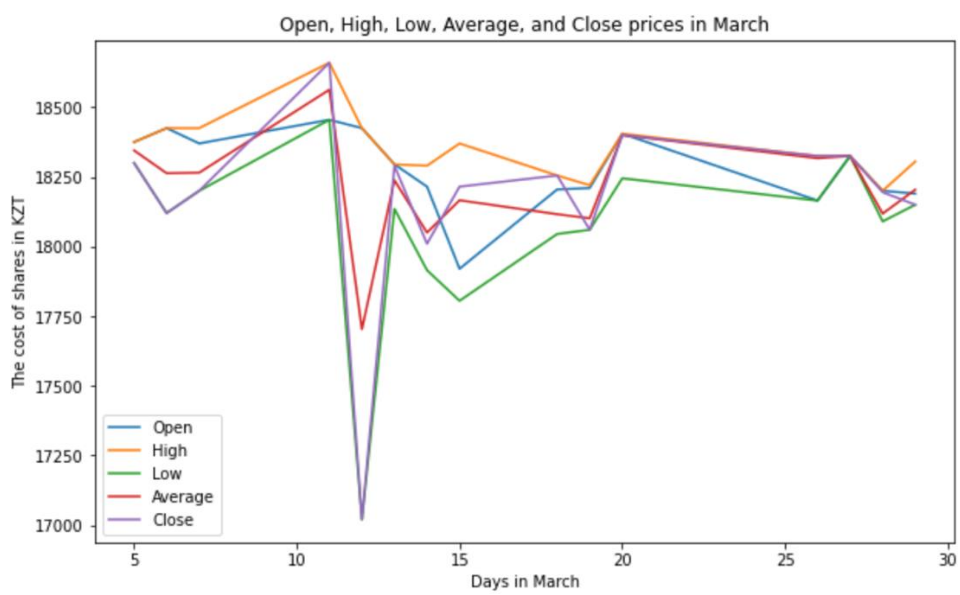

Figure 3. Ordinary share prices of JSC “KazAtomProm" in March in 2024 [29].

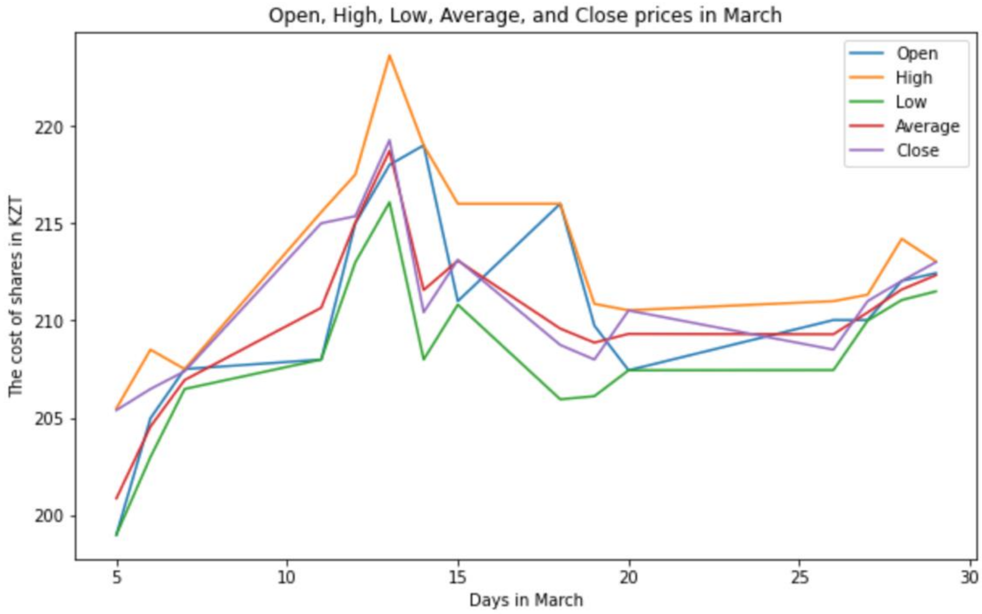

Figure 4. Common share prices of JSC “Halyk Savings Bank of Kazakhstan [30].

Linear Regression Method Results

As mentioned earlier, future trends can be predicted using the Linear Regression Method, which is implemented in the Python programming language for this research. Several libraries are utilized to obtain the final results. The first library, pandas, facilitates efficient manipulation of structured data, including tables and Excel datasets. Additionally, the matplotlib library is employed for rich data visualization, allowing for the representation of data through line graphs, pie charts, and bar charts. For numerical computing and various essential operations, the numpy library plays a crucial role. Lastly, the scikit-learn library is essential, particularly its "sklearn.linear_model" module, which incorporates linear regression functionality. After importing all necessary libraries, the next line of code will help to get the historical monthly data from Excel file. Additionally, all historical data can be downloaded from the main page of Astana International Exchange as an Excel file. To read the data from Excel file there has been used a library “pandas”, which reads so stores the data into a DataFrame “df”. The next lines of code under “#Assuming the first column is ‘Data’ as integer and the second is ‘Cost’ as float” read the date in March from the first column by indexing them and the cost of shares from the second column. The former is integer and the latter is float.

```python
import pandas as pd
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import numpy as np

# Path to the Excel file
file_path = '/Users/erzhigit/Desktop/Dissertation/Data/March(CCBN).xlsx'

# Read the Excel file
df = pd.read_excel(file_path)

# Assuming the first column is 'Date' as integer and the second is 'Cost' as float
X = df.iloc[:, 0].values.reshape(-1, 1)  # Reshaping for sklearn compatibility
Y = df.iloc[:, 1].values

# Create and fit the model
model = LinearRegression()
model.fit(X, Y)

# Predict using the model for the extended date range
X_predict = np.arange(X.min(), 36).reshape(-1, 1)  # Extending to date value of 35
Y_predict = model.predict(X_predict)

# Plotting the original data and the prediction line
plt.scatter(X, Y, color='blue', label='Original Data')
plt.plot(X_predict, Y_predict, color='red', label='Prediction Line')
plt.xlabel('Dates')
plt.ylabel('Cost of Share in KZT')
plt.title('Linear Regression Prediction for CCBN')
plt.legend()
plt.show()
```

Figure 5. The main code of linear regression (Result will be the prediction graph for CCBN).

To create the model, there are written two lines of python code. Linear regression model can be got by typing “model = LinearRegression()”. There is also shown model fitting. “Fit” method is important part of all machine learning methods. It helps to find the optimal parameters for Linear Regression Model. Below fitting method there have to be written the lines of codes for predicting values X and Y. The “X” is the date value, likewise “Y” is the share prices. The last part of code is written especially for printing the monthly data and on the basis of them predicted values of the share prices. It is very important to mention that after writing the python code there has to be used compiler to get the results. In this work, there has been used “Jupyter” Compiler. Finally, the finished code is depicted in Figure 5, tailored for the "Bank CenterCredit" data. Executing this Python code generates the main graph, as depicted in Figure 6. Other figures can be found by changing the importing the monthly historical data into the file path. In Figure 6, both the original data and prediction line results are displayed. The future trends of share prices are observed using the average prices of the shares. From this chart, it appears that shares of Bank CenterCredit JSC are expected to decline over the next 5 days. The historical data used for this analysis encompasses the month of March 2024. Notably, the chart displays the number 35, indicating that stock price data is available only until the 29th of March. The subsequent numbers from 30 to 35 represent the corresponding days of April for convenience.

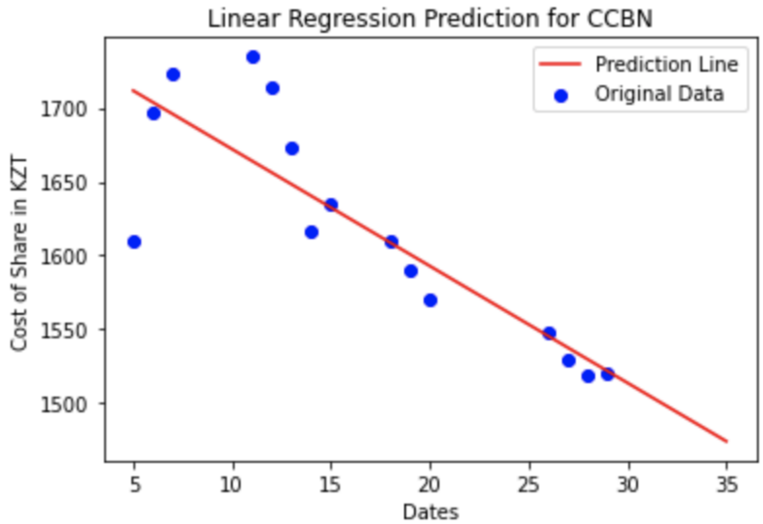

Figure 6. Linear Regression results for CCBN.

In Figure 7, the results of the Linear Regression Method suggest that KazMunaiGas shares are projected to increase over the next 5 days, up until April 5th.

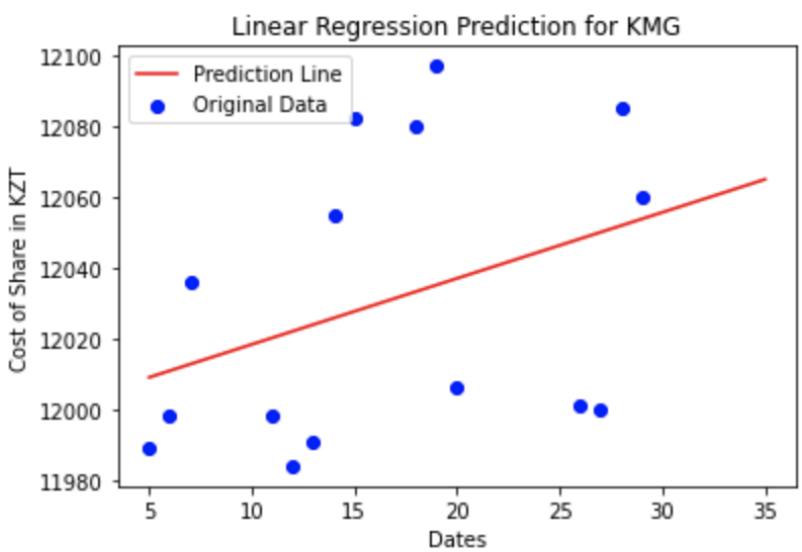

Figure 7. Linear Regression results for KMG.

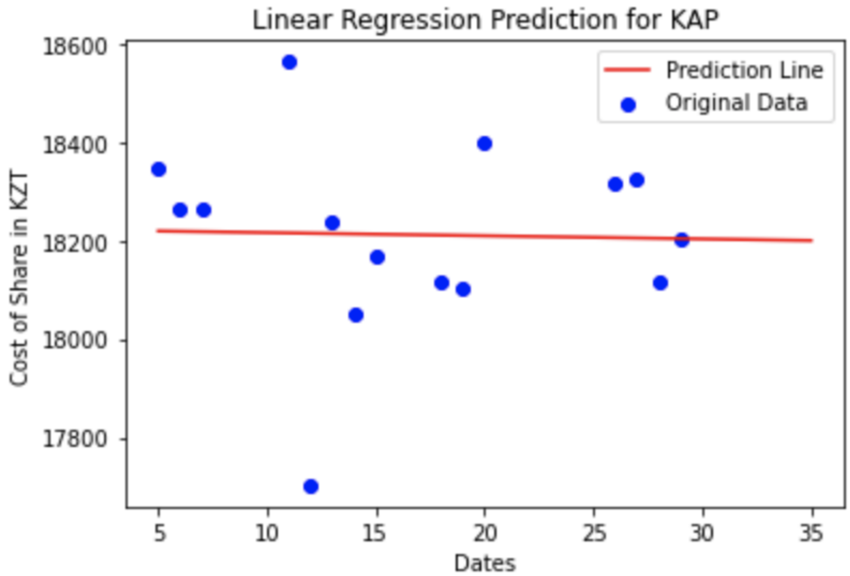

Figure 8. Linear Regression results for KAP.

A different scenario emerges from the results of the data analysis conducted on "KazAtomProm" JSC. The trend indicates that the average share price is expected to remain relatively stable over the next 5 days.

The results of linear regression of the last company “Halyk Savings Bank of Kazakhstan” JSC is demonstrated in Figure 9. The overall trend is similar to results of KMG. The average price of shares of the HSBK goes up during the next 5 days.

According to the above-mentioned results and demonstrated figures, the overall trend can be predicted, but the exact value of predicted prices of the shares will be somehow different.

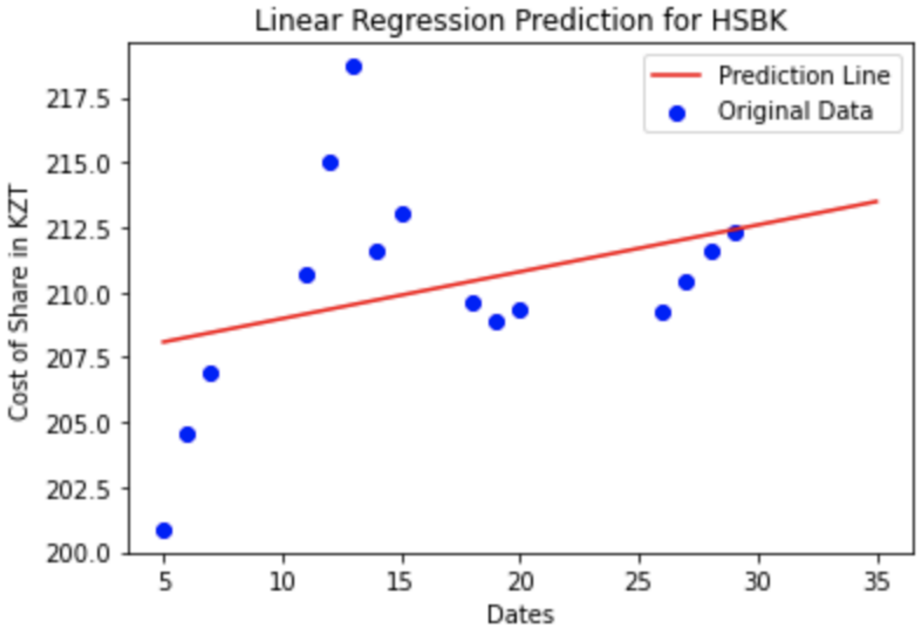

Figure 9. Linear Regression results for HSBK.

Multiple Regression Method Results

Multiple Regression is an advanced method derived from the linear regression technique. The results obtained from both linear and multiple regression methods exhibit similarities. Despite the close resemblance in the codes of these methods, there exist notable differences. The code for the multiple regression method is illustrated in Figure 10.

The initial results of code compilation for CCBN are depicted in Figure 11. Prediction results for three additional companies can be found in Figures 12, 13, and 14, respectively. These figures exhibit similar trends to Figures 7, 8, and 9. Subsequent sections of this research will delve into further analyses and comparisons of linear and multiple regression methods.

```python
import pandas as pd
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import numpy as np

# Path to the Excel file
file_path = '/Users/erzhigit/Desktop/Dissertation/Data/March(CCBN).xlsx'

# Read the Excel file
df = pd.read_excel(file_path)

# Assuming the first column is 'Date' as integer and the second is 'Cost' as float
X = df.iloc[:, [0]].values  # Reshaping for sklearn compatibility (and readiness for more variables)
Y = df.iloc[:, 1].values

# Create and fit the model
model = LinearRegression()
model.fit(X, Y)

# Predict using the model for the extended date range
X_predict = np.arange(X.min(), 36).reshape(-1, 1)  # Extending to date value of 35
Y_predict = model.predict(X_predict)

# Plotting the original data and the prediction line
plt.scatter(X, Y, color='blue', label='Original Data')
plt.plot(X_predict, Y_predict, color='red', label='Prediction Line')
plt.xlabel('Date')
plt.ylabel('Cost of Share in KZT')
plt.title('Multiple Regression Prediction for CCBN')
plt.legend()
plt.show()
```

Figure 10. Multiple regression code for CCBN.

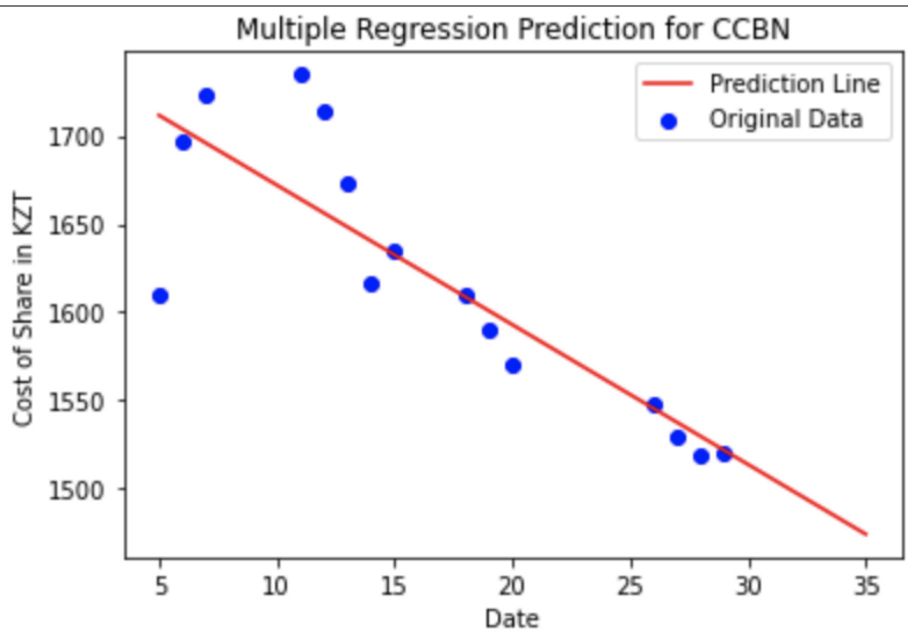

Figure 11. Multiple regression method results for CCBN.

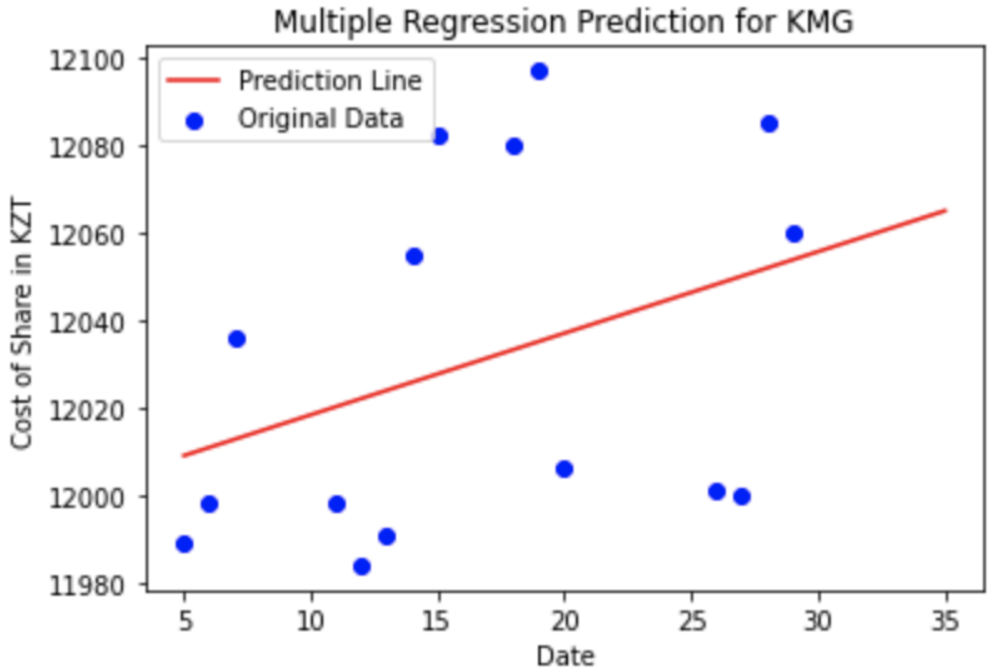

Figure 12. Multiple regression method results for KMG.

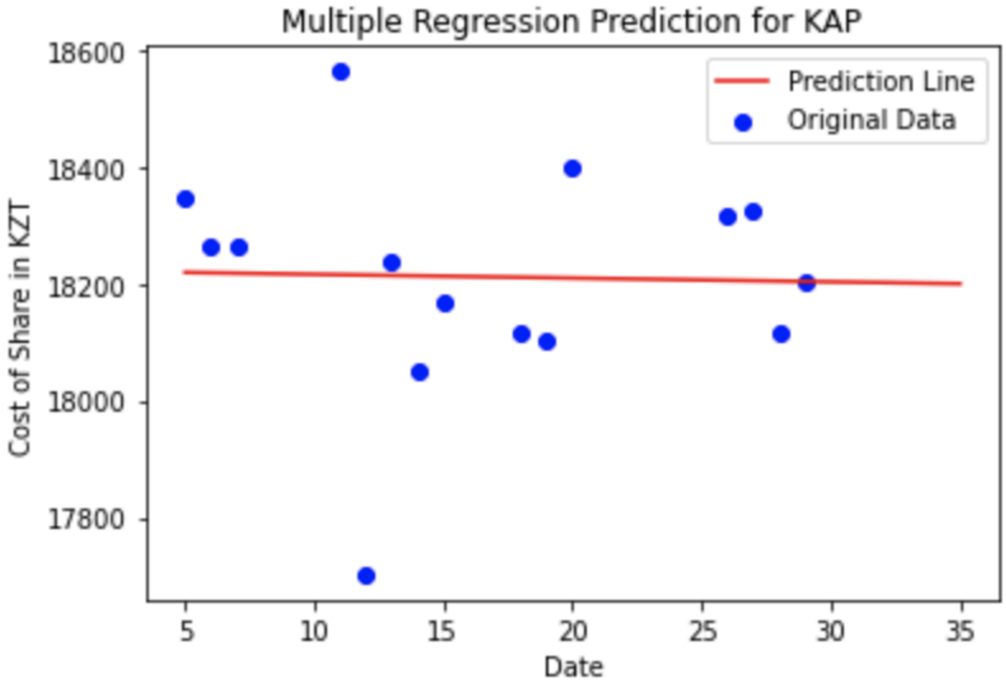

Figure 13. Multiple regression method results for KAP.

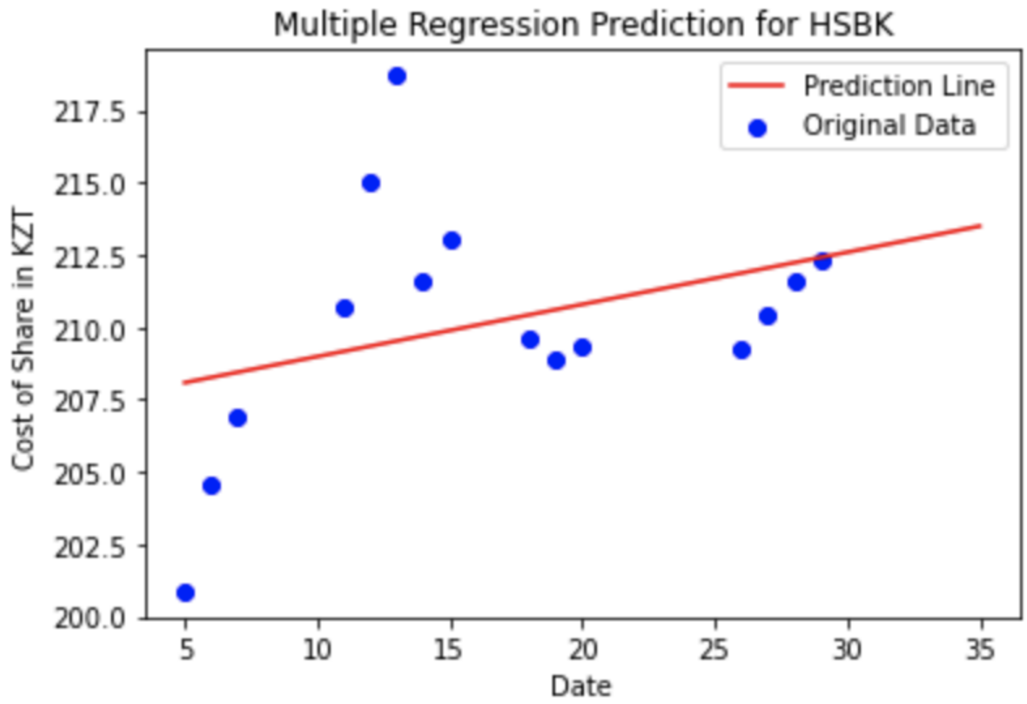

Figure 14. Multiple regression method results for HSBK.

ARIMA Results

One of the well-known methods for predicting future financial trends is the Autoregressive Integrated Moving Average (ARIMA) model, which is a statistical method. The advantages of this method have already been discussed. To demonstrate the predicted values, the first step is to install the necessary statistical models in Python. This can be achieved by simply typing "pip install statsmodels" in the Jupyter Compiler within the Anaconda software environment. Once installed, the Python code can be written. The primary code structure can be observed in Figure 15. The original data can be imported from an Excel file obtained from the Astana International Exchange platform. The main part of ARIMA method code is similar to the above-mentioned codes of Linear Regression and Multiple Regression Methods. However, there are some differences. The line of code “model =ARIMA (series, order= (5, 1, 0))” contains the information about the model ARIMA: the dataset ‘series’ to which ARIMA model is applied, ‘order = (5,1,0)’ which indicates that ‘p=5’ considers the last 5 observations in the series, ‘d=1’ indicates the series will be differenced once to make it stationary, and the last component ‘q=0’ gives information that there is not included moving average in this ARIMA model. The next step is fitting the model ARIMA. After fitting the model, there is a line “forecast = module_fit.forecast(steps = 10)” to forecast the next 10 values of share costs. After this line of code there are some lines of python code, which is vital for printing the obtained results. In Figure 15, the line where the ARIMA method is imported is also illustrated. Upon compilation of the code, the predicted share prices for the company "Bank of CenterCredit" JSC can be obtained (refer to Figure 16).

```python
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA

# Path to the Excel file
file_path = '/Users/erzhigit/Desktop/Dissertation/Data/March(CCBN).xlsx'

# Read the Excel file
df = pd.read_excel(file_path)

# Assuming the first column is integer-based date and the second column is the cost of shares
series = pd.Series(df.iloc[:, 1].values, index=df.iloc[:, 0])

# Fit an ARIMA model
# These are placeholder parameters for p, d, q
model = ARIMA(series, order=(5,1,0))  # Adjust order based on your data
model_fit = model.fit()

# Forecasting
forecast = model_fit.forecast(steps=10)  # Forecasting the next 10 points

# Plotting the results
plt.figure(figsize=(10, 6))
plt.plot(series, label='Original')
plt.plot(range(series.index[-1] + 1, series.index[-1] + 11), forecast, label='Forecast')
plt.xlabel('Date')
plt.ylabel('Cost of Share')
plt.title('ARIMA Forecast for CCBN')
plt.legend()
plt.show()
```

Figure 15. Code in python of ARIMA method.

As observed in Figure 16, the ARIMA method demonstrates the capability to predict precise share price values. For example, based on the ARIMA results for CCBN, it indicates an initial increase followed by a subsequent decrease in share prices. It's worth noting that the results presented above utilize average share prices from the AIX stock. However, it's also possible to forecast high and low share prices using historical data. This approach aids buyers and financial professionals in mitigating the risks associated with excessive spending of financial resources. By strategically purchasing shares at lower prices during stock operation, financial specialists can optimize their investment strategies.

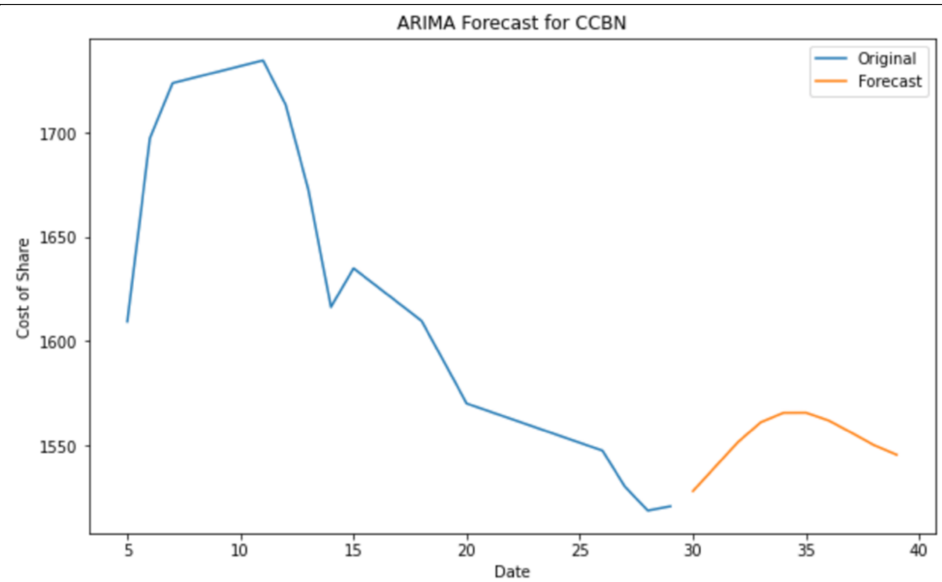

Figure 16. ARIMA method results for CCBN.

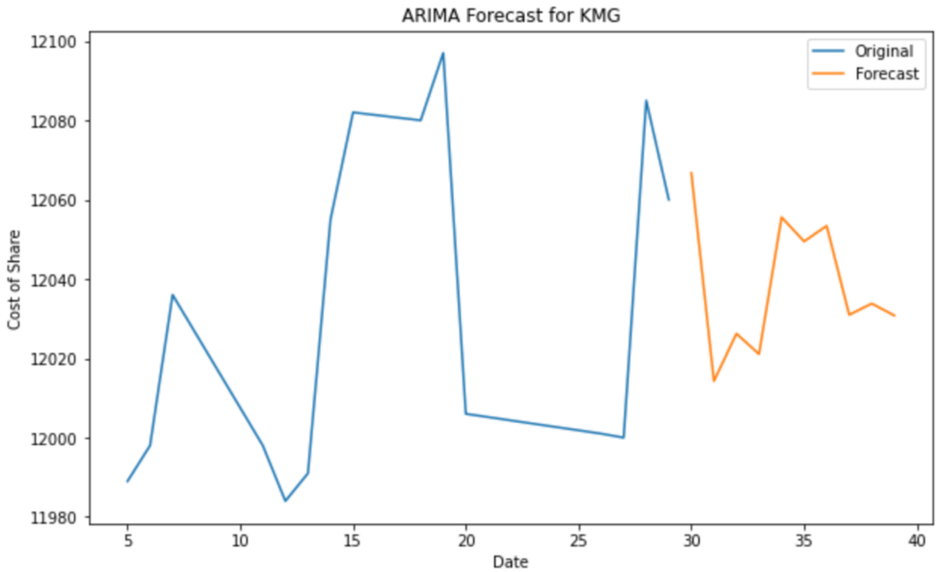

Figure 17. ARIMA method results for KMG.

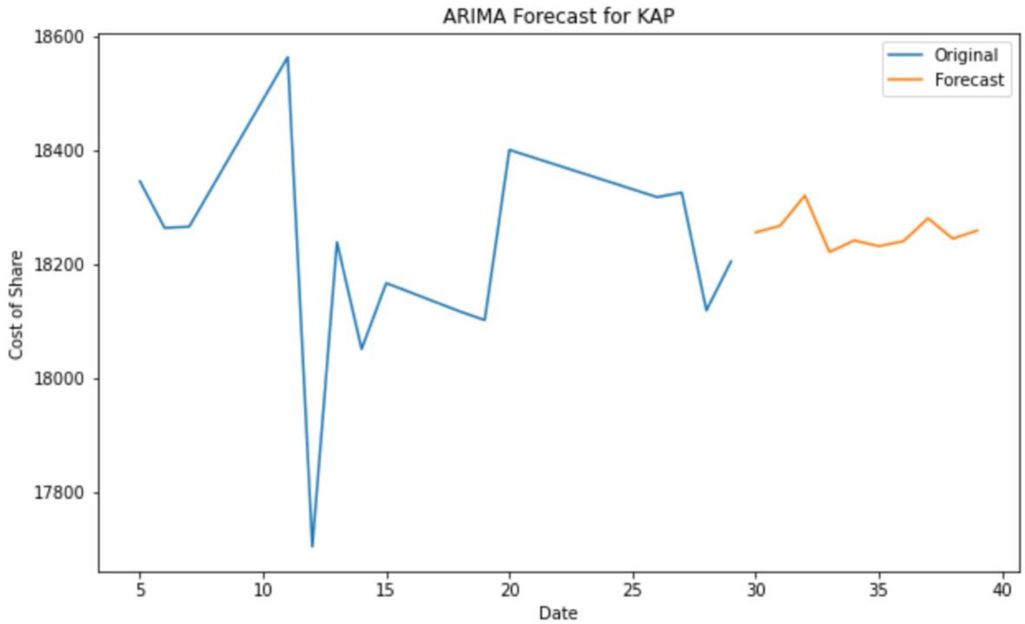

Figure 18. ARIMA method results for KAP.

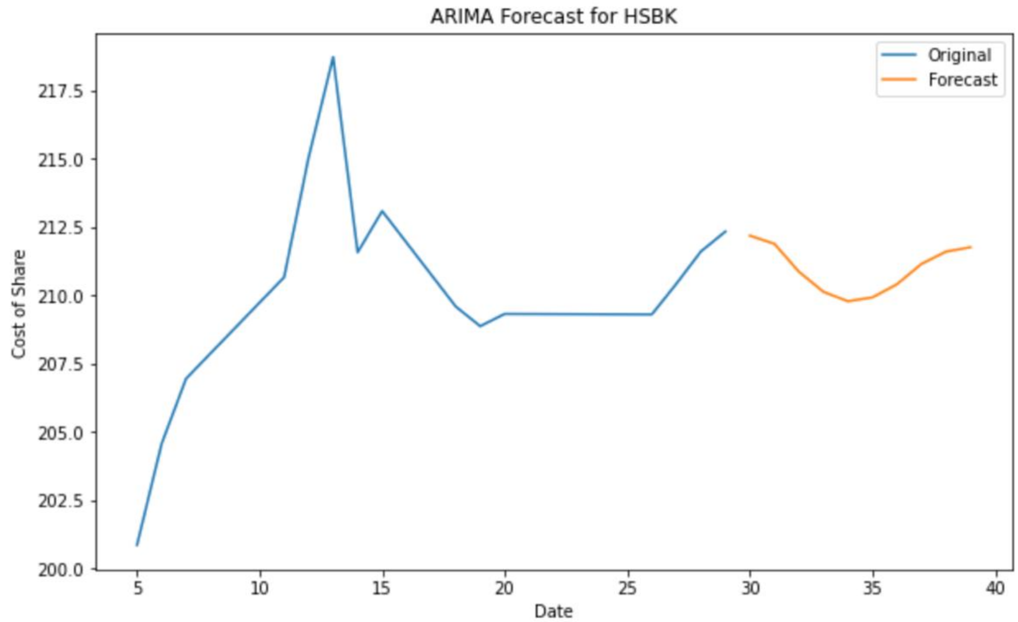

Figure 19. ARIMA method results for HSBK.

Similar results have been obtained for other companies, including "KazMunayGas" JSC, "KazAtomProm" JSC, and "Halyk Savings Bank of Kazakhstan" JSC. The predicted values for these companies are depicted in Figures 16, 17, 18, and 19, respectively. To assess the accuracy of the ARIMA method's predictions, a comparison will be made between the obtained results and the original data available at AIX. It is essential to verify the accuracy of the predictions by comparing them with the actual data provided on the AIX platform.

AR Results

Another method employed for predicting future data is the Autoregressive (AR) method. The primary Python code structure resembles the examples illustrated above, albeit with some distinctions. The autoregressive package can be imported using the following code snippet: "from statsmodels.tsa.ar_model import AutoReg". Figure 20 illustrates the implementation of the autoregressive code in Python. The resulting predictions are visualized in Figures 21, 22, 23, and 24. These figures provide insights into the trends and share prices.

In Figure 21, it is projected that the prices of "Bank of CenterCredit" JSC shares will decrease over the next 5 days. Similar trends are evident in Figures 22 and 24. Conversely, share prices of KazAtomProm and National Bank are expected to remain stable over the same period. However, for KazMunaiGas company, a decrease in share prices is anticipated (refer to Figure 23).

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.ar_model import AutoReg

# Path to the Excel file
file_path = '/Users/erzhigit/Desktop/Dissertation/Data/March(CCBN).xlsx'

# Read the Excel file
df = pd.read_excel(file_path)

# Assuming the first column is 'Date' as integer and the second is 'Cost' as float
# In AR models, only the dependent variable (cost) is used
costs = df.iloc[:, 1]

# Fit an AR model
# The lag value (lags) can be adjusted based on your data
model = AutoReg(costs, lags=1)
model_fit = model.fit()

# Forecasting
forecast_steps = 5  # number of steps to forecast
forecast = model_fit.predict(start=len(costs), end=len(costs) + forecast_steps - 1)

# Plotting the results
plt.figure(figsize=(10, 6))
plt.plot(df.iloc[:, 0], costs, label='Original')
forecast_dates = range(df.iloc[-1, 0] + 1, df.iloc[-1, 0] + 1 + forecast_steps)
plt.plot(forecast_dates, forecast, label='Forecast')
plt.xlabel('Date')
plt.ylabel('Cost of Share in KZT')
plt.title('AR Model Forecast for CCBN')
plt.legend()
plt.show()
```

Figure 20. The Autoregressive method code in python.

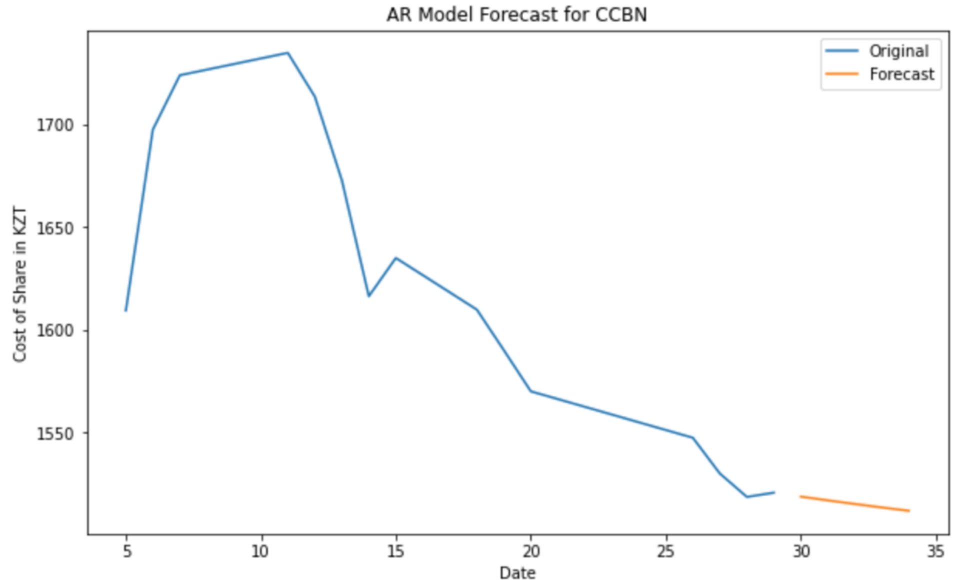

Figure 21. The AR results for CCBN.

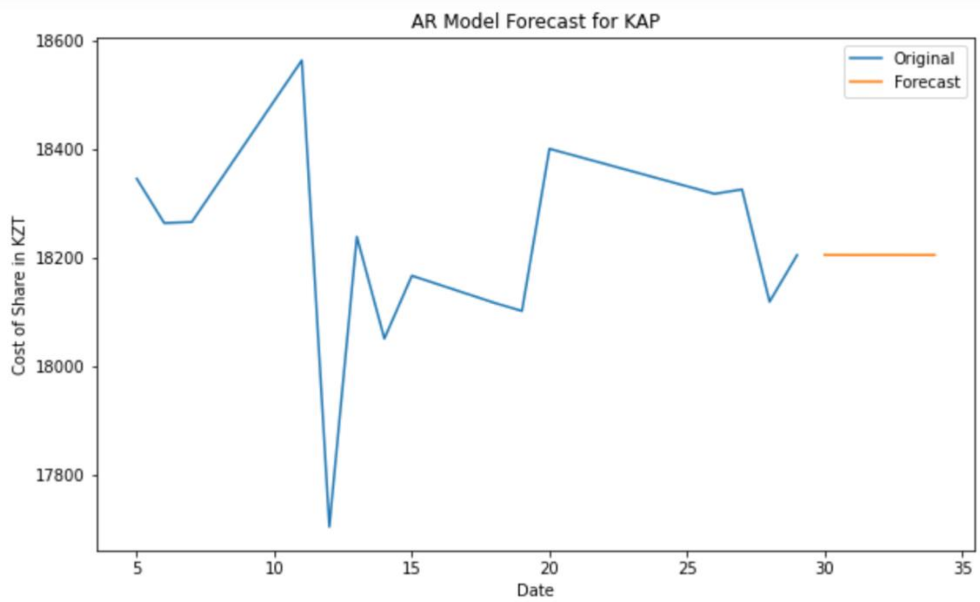

Figure 22. The AR results for KAP

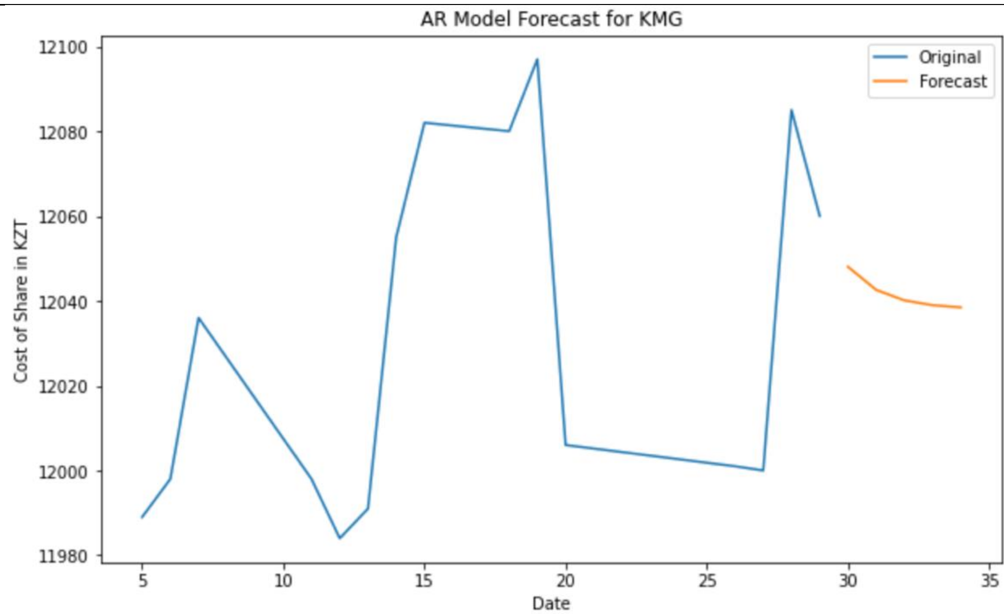

Figure 23. The AR results for KMG.

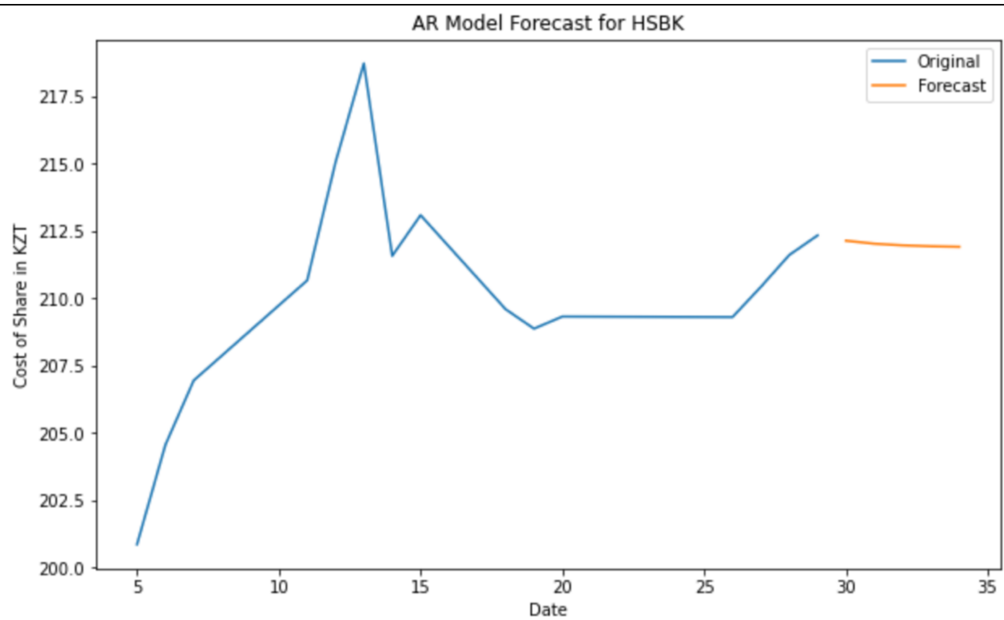

Figure 24. The AR results for HSBK.

Support Vector Regression

There are shown results of different methods for chosen different companies in above chapters of the research. The same procedure will be done for the method which is called Support Vector Regression (SVR). There have already talked the pros and cons of this method compared to the other. The main code of this method in programming language python is shown in Figure 25. The main code is similar to above-mentioned codes in python. To use the SVR to predict the future values there has to be imported the from the library “sklearn.svm” the method “SVR”. The line from sklearn.preprocessing import StandardScaler in Python imports the StandardScaler class from the preprocessing module of the sklearn library. StandardScaler is a tool used to normalize or scale features in your data. Other differences are shown in the code and the every operation’s name is written above each operation’s line of code.

```python
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.svm import SVR
from sklearn.preprocessing import StandardScaler

# Path to the Excel file
file_path = '/Users/erzhigit/Desktop/Dissertation/Data/March(CCBN).xlsx'

# Read the Excel file
df = pd.read_excel(file_path)

# Assuming the first column is integer-based date and the second column is the cost of shares
dates = df.iloc[:, 0].values.reshape(-1, 1)  # Reshape for sklearn compatibility
prices = df.iloc[:, 1].values

# Standardizing the data
scaler_x = StandardScaler()
scaler_y = StandardScaler()
dates_scaled = scaler_x.fit_transform(dates)
prices_scaled = scaler_y.fit_transform(prices.reshape(-1, 1))

# Support Vector Regression Model
svr_rbf = SVR(kernel='rbf', C=100, gamma=0.1, epsilon=.1)

# Fit the model
svr_rbf.fit(dates_scaled, prices_scaled.ravel())

# Predicting future values, let's say next 10 points
future_dates = pd.Series(range(dates.max() + 1, dates.max() + 11)).values.reshape(-1, 1)
future_dates_scaled = scaler_x.transform(future_dates)
future_prices_scaled = svr_rbf.predict(future_dates_scaled)
future_prices = scaler_y.inverse_transform(future_prices_scaled.reshape(-1, 1))

# Plotting the results
plt.figure(figsize=(10, 6))
plt.scatter(dates, prices, color='black', label='Data')
plt.plot(dates, scaler_y.inverse_transform(prices_scaled), color='blue', label='SVR Fit')
plt.plot(future_dates, future_prices, color='red', label='Future Prediction')
plt.xlabel('Date')
plt.ylabel('Cost of Share in KZT')
plt.title('SVR Forecasting for CCBN')
plt.legend()
plt.show()
```

Figure 25. The code in python of the Support Vector Machines (SVR).

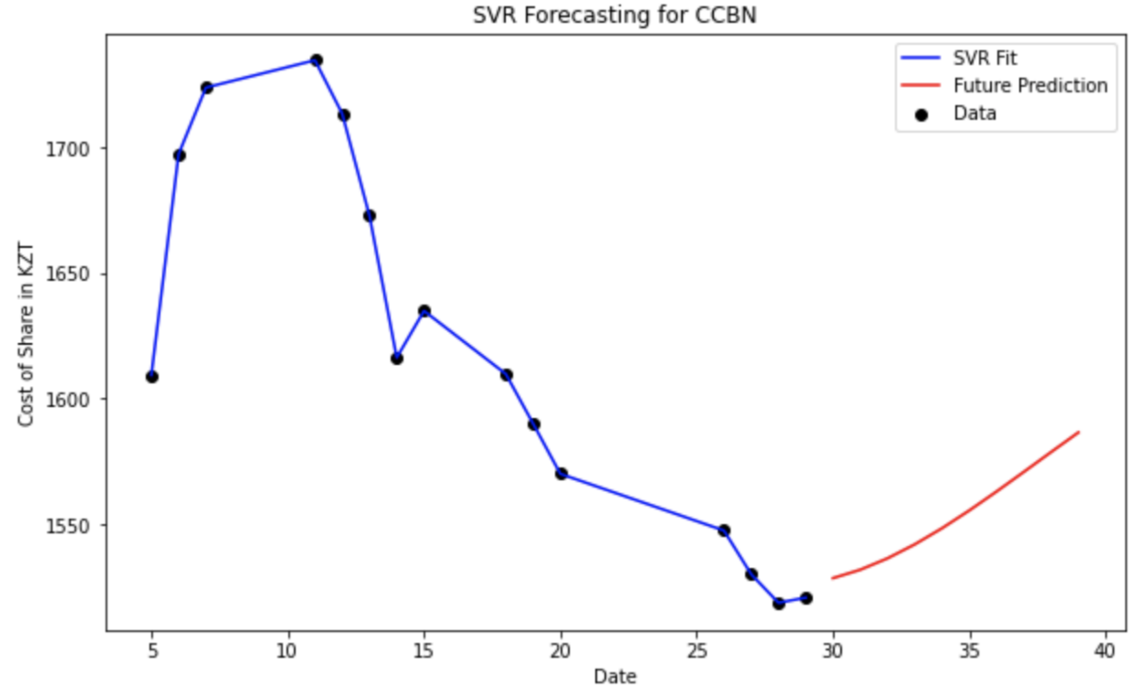

Figure 26. The SVR results for the company CCBN.

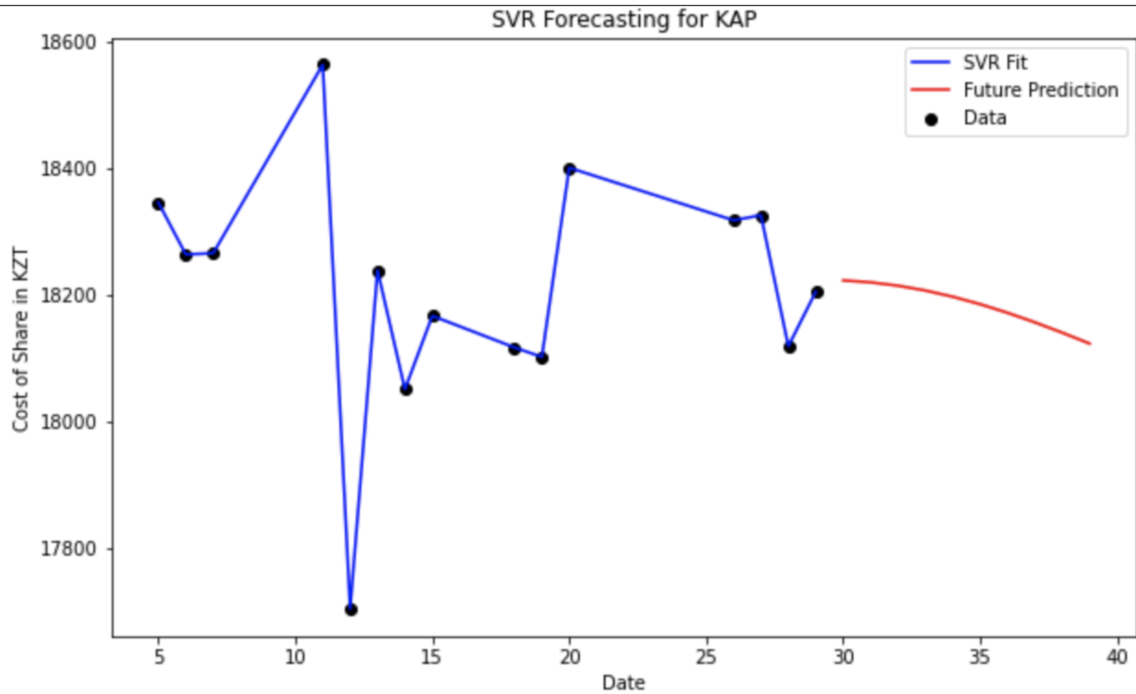

Figure 27. The SVR results for the company KAP.

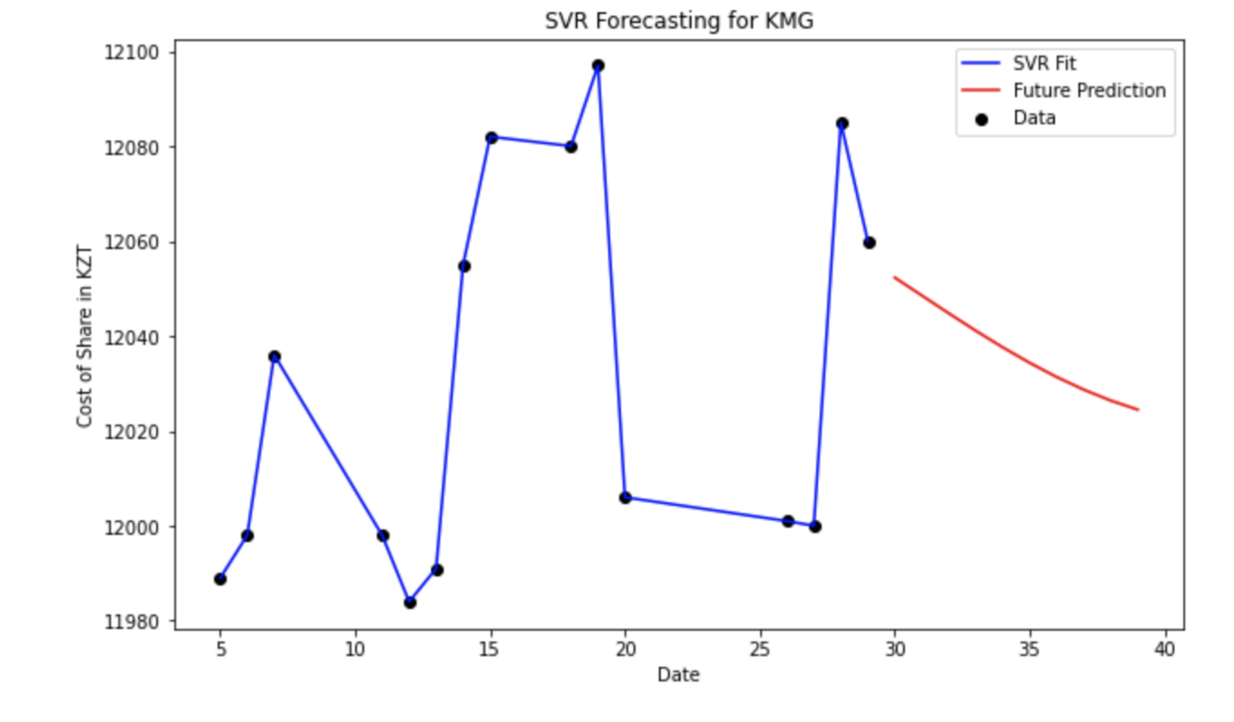

Figure 28. The SVR results for the company KMG.

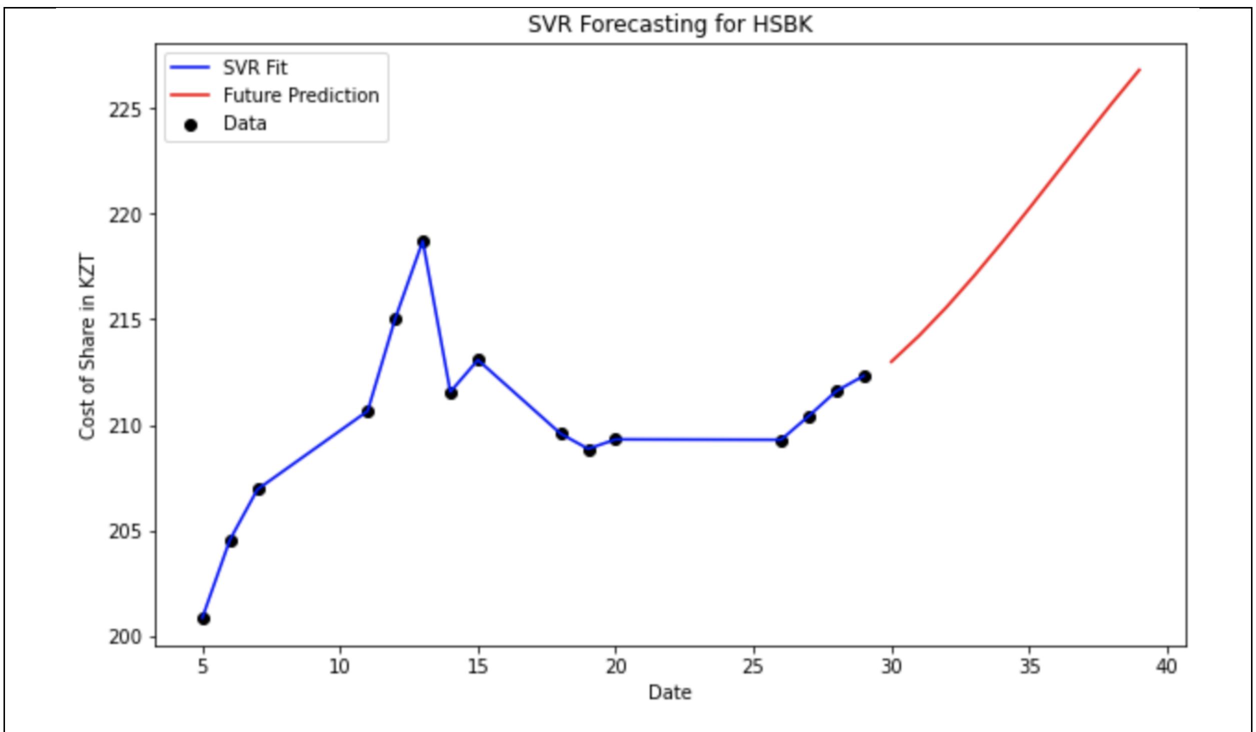

Figure 29. The SVR results for the company HSBK.

Comparisons between different methods' results

The results obtained from various methods have already been the subject of some discussions in the past. In the following paragraph, the results of these different methods will be compared amongst themselves, as well as with the current data.

In the preceding chapters of the research, historical data spanning the entirety of March was utilized, with forecasted values extending up to the 5th of April. Therefore, the forecasted values obtained from different methods (refer to Tables 1, 2, 3, 4) should be compared.

Table 1. Comparisons of the average prices of shares of CCBN.

Date	Linear Regression (in KZT)	AR (in KZT)	ARIMA (in KZT)	SVR (in KZT)	Actual (in KZT)
1st of April	1513.77	1518.81	1528.00	1528.49	1575.14
2nd of April	1505.85	1516.97	1539.86	1531.84	1660.94
3rd of April	1497.93	1515.22	1551.66	1536.38	1625.64
4th of April	1490.02	1513.53	1561.05	1541.96	1610.02
5th of April	1482.09	1511.92	1565.59	1548.38	1580.08

As it can be seen from Table 1, forecasts, obtained by ARIMA method, exactly coincide with actual prices of Bank CenterCredit shares. Actual data on the company's shares showed a growth up to KZT1,660.94 from 1 to 2 April, followed by a decline. The similar tendency was observed in forecasts obtained by ARIMA method. As it is shown in figure 16, the prices were growing till April 5, and then a decline followed. Also, the closest results were obtained using the SVR method. Like the actual data, the data predicted by the SVR method shows a slight increase on the 2nd of April, but from the 2nd of April to the 5th of April there is a gradual decline in values, whereas the figures predicted by the

ARIMA and SVR methods increase thereafter. Looking at the average price values generated by the AR method, it can be noted that the results generated by the AR method show significant comparability with the actual data. Except for the fact that the AR method does not show a price increase on 2 April, but only a gradual price decrease over the whole period. Although the simplest linear regression model gives favorable results, they are not superior to the results of the above-mentioned methods.

Table 2. Comparisons of the average prices of shares of KAP.

Date	Linear Regression (in KZT)	AR (in KZT)	ARIMA (in KZT)	SVR (in KZT)	Actual (in KZT)
1st of April	18203.06	18204.32	18255.03	18222.00	18290.
2nd of April	18202.41	18204.23	18266.82	18218.99	18690
3rd of April	18201.78	18204.26	18319.82	18213.66	18875
4th of April	18201.14	18204.24	18220.35	18206.04	18991
5th of April	18200.49	18204.25	18240.95	18196.25	18881

The table above shows the results of four methods in comparison with actual data on shares of KazAtomProm JSC. Looking at the results in more detail, it can be seen that the data generated by LR method is gradually decreasing, while the actual data shows progressive growth for this period. The same conclusions can be drawn regarding the SVR method, which shows a general decline in prices. Despite the fact that AR generated a more accurate forecast of BCC share prices, in this case, the data diverges significantly from the present figures, as in general the forecasted prices remain unchanged over the period of 5 days. The closest figures were predicted by ARIMA method, however there are slight divergences with actual data.

Table 3. Comparisons of the average prices of shares of KMG.

Date	Linear Regression (in KZT)	AR (in KZT)	ARIMA (in KZT)	SVR (in KZT)	Actual (in KZT)
1st of April	12055.66	12048.04	12066.73	12052.30	12072
2nd of April	12057.53	12042.59	12014.24	12048.53	12123
3rd of April	12059.39	12040.11	12026.22	12044.77	12071
4th of April	12061.26	12038.98	12021.06	12041.11	12128
5th of April	12063.12	12038.46	12055.57	12037.60	12057

The following table 3 presents the results of the study obtained using four methods: Linear regression, AR, ARIMA, and SVR in comparison with the current data. All these methods demonstrate high efficiency in forecasting share prices, as evidenced by comparison with actual data available on the main site of Astana International Exchange (AIX). However, it can be noted that linear regression is less accurate in this case, as this method showed an increase in average prices on a daily basis, while the real share prices of KMG fluctuated during the whole period, but generally fell. The most similar average prices to the real prices were predicted by the ARIMA method.

Table 4. Comparisons of the average prices of shares of HSBK.

Date	Linear Regression (in KZT)	AR (in KZT)	ARIMA (in KZT)	SVR (in KZT)	Actual (in KZT)
1st of April	212.60	212.13	212.18	212.99	212.58
2nd of April	212.78	212.02	211.88	214.22	213.07
3rd of April	212.96	211.96	210.86	215.58	212.48
4th of April	213.14	211.93	210.13	217.05	213.34
5th of April	213.32	211.91	209.78	218.62	211.02

According to Table 4, the best results are observed in the case of Halyk Savings Bank of Kazakhstan (HSBK), where all four methods show strong correlation with actual data. However, it should be said that the most accurate data was obtained using the AR method, on 5 April the real share price was 211.02tg and the AR predicted price was 211.91tg. While the most distant results were obtained using the SVR method, which showed a significant increase in share prices for the period from 1st to 5th April. It is noteworthy that the simplest linear regression forecasting model shows a good approximation.

Conclusion

The purpose of this study was the implement and analyze the effectiveness of four machine learning models, in particular, the model of Linear Regression, AutoRegressive (AR), AutoRegressive Integrated Moving Average (ARIMA) and support vector machine (SVM), in predicting index movements on the Astana International Exchange (AIX) using technical analysis methods. The methodology was developed using literature from previous studies on the implementation of machine learning algorithms for predicting stock market trends. However, this is the first approach to apply such algorithms, especially in the context of AIX. For learning and testing the machine learning models historical data of 2023 was used, including the values of the AIX index, along with a number of technical indicators.

The paper focuses on four core companies that are quoted on the Astana International Exchange (AIX): "Bank of CenterCredit" JSC, "KazAtomProm" JSC, "KazMunayGaz" JSC, and "Halyk Savings of Bank of Kazakhstan" JSC. The research methodology includes the realization and demonstration of the Python code for each of the machine learning models. These models are used to predict the stock prices of the chosen companies. The study thoroughly compares the predicted prices obtained by these models with the actual market data available at the Astana International Exchange.

The results of the study show that all four methods—Linear Regression, AR, SVM and ARIMA—can produce satisfactory outcomes in predicting the movement of stock prices with a reasonable degree of accuracy. However, in a few cases ARIMA demonstrates noticeable effectiveness. Through a comprehensive comparison of predicted and actual prices, the study provides valuable insight into the performance and suitability of each model for stock price forecasting in the context of AIX. In consequence, the above mentioned models could become precious instruments for both traders and investors, aspiring to make reasonable choices regarding their investments strategies on AIX. But it is also should be noted that, as well as technical analysis as such these simulations, while being utilized in actual trading or investing must be applied in tandem with other existing instruments.

Although the results of this study illustrate the prospective performance of Linear Regression, AutoRegressive (AR), AutoRegressive Integrated Moving Average (ARIMA) and support vector machine (SVM) models in projecting the changes of AIX Index, however, several restrictions exist that should to be addressed. One of the primary constraints is the comparatively short range dataset used in this investigation, which only includes the time frame from January till December of 2023 year. The research has demonstrated that widening of time limits can increase the efficiency of the models. But on the case of the AIX Index one would need some time for more new information to collect. That is why prospective investigations may as well look into the possibility of applying these models to particular shares quoted on AIX, which have a more extended time period.

In addition, the investigation used a narrow range of technical indicators to learn and perform model testing, which may have limited the precision of the models. For this reason, the inclusion of supplementary technical indicators may be helpful. More importantly, it may be intriguing to witness in upcoming studies the use of not only complementary technical indicators, as well as economic variables and market trends as functions to learn and verify models and examine their impact on forecast precision.

By considering the constantly evolving breakthrough character of artificial intelligence and machine learning, prospective studies may explore the effectiveness of other machine learning algorithms for example BERT, WaveNet and Prophet in forecasting the performance of the AIX index or quoted stocks on it.

As a conclusion, the study suggests avenues for further research. It suggests that alternative methods other than those used in this study should be examined to improve forecasting precision. It also suggests that combining tools that can account for market contingencies, black swan events, changing economic indicators, and unexpected conditions such as the COVID-19 pandemic could significantly improve forecasting performance. In the light of these observations, future research could improve stock price forecasting models and may provide more reliable information to investors and financial analysts navigating dynamic market conditions. The obtained results, as it is hoped, will bring an input to the study of AIX as a forming and growing stock market and will fuel the enthusiasm for this subject among researchers of this country.

References:

1. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.
2. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
4. Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press.
5. Cortes, C., & Vapnik, V. (1995). "Support-vector networks." Machine Learning, 20(3), 273-297.
6. Cherkassky, V., & Ma, Y. (2004). "Practical selection of SVM parameters and noise estimation for SVM regression." Neural Networks, 17(1), 113-126.
7. E. Zhantileuov, A. Smaiyl, A. Aibatbek, and S. Kassymkhanov, "A Case Study of Machine Learning Comparisons for Predicting Apartment Prices in Astana," in 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), 2023, pp. 305-309. DOI: 10.1109/SIST58284.2023.10223463. Available: IEEE Xplore.
8. J. Patel, S. R. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," Expert Syst. Appl., vol. 42, pp. 259-268, 2015. DOI: 10.1016/j.eswa.2014.07.040. Available: ScienceDirect.

9. J. Patel, S. R. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," Expert Syst. Appl., vol. 42, pp. 2162-2172, 2015. DOI: 10.1016/j.eswa.2014.10.031. Available: ScienceDirect.
10. N. Ayyıldız, "Prediction of Stock Market Index Movements with Machine Learning," 2023. DOI: 10.58830/ozgur.pub354. Available: Ozgur Publications.
11. M. Hanias, P. Curtis, and E. Thalassinos, "Time Series Prediction with Neural Networks for the Athens Stock Exchange Indicator," European Research Studies Journal, vol. 15, pp. 23-32, 2012. DOI: 10.35808/ERSJ/351. Available: European Research Studies Journal.
12. E. Güresen, G. Kayakutlu, and T. Daim, "Using artificial neural network models in stock market index prediction," Expert Syst. Appl., vol. 38, pp. 10389-10397, 2011. DOI: 10.1016/j.eswa.2011.02.068. Available: ScienceDirect.
13. N. Sagadatova, A. Nugumanova, N. Zhakiyev, and B. Talas, "FORECASTING ELECTRICITY DEMAND WITH FORECASTERAUTOREG: CASE STUDY IN ASTANA," Scientific Journal of Astana IT University, 2023. DOI: 10.37943/14trmf1662. Available: Astana IT University.
14. Overview - AIX," [Online]. Available: https://aix.kz/about-aix/overview/.
15. R. Husain and R. Vohra, "Applying Machine Learning in the Financial Sector," in International Education and Research Journal, vol. 3, 2017. [Online].
16. H. Ghoddusi, G. G. Creamer, and N. Rafizadeh, "Machine Learning in Energy Economics and Finance: A Review," in Energy eJournal, 2018. [Online]. Available: Machine Learning in Energy Economics and Finance: A Review - Consensus

17. S. Emerson, R. Kennedy, L. O'Shea, and J. R. O'Brien, "Trends and Applications of Machine Learning in Quantitative Finance," in Machine Learning eJournal, 2019. [Online]. Available: Trends and Applications of Machine Learning in Quantitative Finance - Consensus
18. D. Snow, "Financial Event Prediction using Machine Learning," in S&P Global Market Intelligence Research Paper Series, 2019. [Online]. Available: Financial Event Prediction using Machine Learning - Consensus
19. T. Hastie, J. Friedman, and R. Tibshirani, "Linear Methods for Regression," pp. 41-78, 2001. [Online].
20. K. Kumari and S. Yadav, "Linear regression analysis study," Journal of the Practice of Cardiovascular Sciences, vol. 4, pp. 33-36, 2018. [Online].
21. V. Silva, M. E. Silva, P. Ribeiro, and F. M. A. Silva, "Time series analysis via network science: Concepts and algorithms," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 11, 2021. [Online].
22. D. Özdemir, "The Analysis of Time Series," 2016. [Online].
23. J. Zhang, R. Shan, and W. Su, "Applying Time Series Analysis Builds Stock Price Forecast Model," Mathematical Models and Methods in Applied Sciences, vol. 3, pp. 152, 2009. [Online].
24. R. Thomson and W. Emery, "Time-series Analysis Methods," 2014. [Online].
25. V. Silva, M. E. Silva, P. Ribeiro, and F. M. A. Silva, "Time series analysis via network science: Concepts and algorithms," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 11, 2021. [Online].

26. X. Yang and B. Liu, "Uncertain time series analysis with imprecise observations," Fuzzy Optimization and Decision Making, 2018. [Online].
27. “Market watch,” AIX, https://market.aix.kz/details/CCBN/profile (accessed Apr. 8, 2024).
28. “Market watch,” AIX, https://market.aix.kz/details/KMG/profile (accessed Apr. 8, 2024).
29. “Market watch,” AIX, https://market.aix.kz/details/KAP/profile (accessed Apr. 8, 2024).
30. “Market watch,” AIX, https://market.aix.kz/details/HSBK/profile (accessed Apr. 8, 2024).
31. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). Time Series Analysis: Forecasting and Control. Wiley.
32. Brockwell, P. J., & Davis, R. A. (2016). Introduction to Time Series and Forecasting. Springer.
33. Chatfield, C. (2019). The Analysis of Time Series: An Introduction (7th ed.). Chapman & Hall/CRC.
34. Shumway, R. H., & Stoffer, D. S. (2017). Time Series Analysis and Its Applications: With R Examples (4th ed.). Springer.
35. Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods (2nd ed.). Pearson.
36. Enders, W. (2014). Applied Econometric Time Series (4th ed.). Wiley.
37. Tsay, R. S. (2010). Analysis of Financial Time Series (3rd ed.). Wiley.
38. Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer.

39. Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.
40. Hsu, C., Chang, C., & Lin, C. (2003). A Practical Guide to Support Vector Classification. National Taiwan University.
41. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). Introduction to linear regression analysis. John Wiley & Sons.
42. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian data analysis. CRC press.
43. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer Science & Business Media.
44. Fox, J. (2016). Applied regression analysis and generalized linear models. Sage Publications.
45. Husain, R., & Vohra, R. (2017). Application of machine learning in financial sector: A survey. International Journal of Computer Applications, 169(5), 10-16.
46. V. Nasteski, "An overview of the supervised machine learning methods," HORIZONS.B, vol. 4, pp. 51–62, Dec. 2017. doi:10.20544/horizons.b.04.1.17.p05