Research on the Use of Machine Learning Methods for Forecasting Time Series when Making Management Decisions in IT Projects Under Martial Law

Viktor Morozov, Vladyslav Deineha and Andrii Khlevnyi

Taras Shevchenko National University of Kyiv, 24, Bohdan Gavrilishin Str., Kyiv, 04116, Ukraine

Abstract

This article will highlight the importance of time series forecasting in the modern world. The relevance of this topic is conditioned in the need for organizations and individuals to predict events based on the analysis of past data. Time series forecasting plays a critical role in planning, risk management, and strategic decision-making in IT projects, making it a key component of modern analytics and martial law management.

The topic of applying machine learning methods for time series forecasting will also be discussed in detail. The main methods will be reviewed, including exponential smoothing, ARIMA (Autoregressive Integrated Moving Average), and a hybrid method that combines different methods to improve forecast accuracy.

Exponential smoothing is a simple and effective method based on taking into account weighted averages of previous observations. ARIMA, in turn, is a classical statistical method that combines autoregression, integration, and moving average to model time series. A hybrid method in time series forecasting is a combination of two or more methods, and in this case, includes an improved approach with MAPE-dependent weights, allowing the weights of the methods to be adapted depending on their performance on time series data.

The literature review covers relevant scientific works in the field of time series forecasting using machine learning methods. Different approaches, their advantages and limitations will be discussed to provide a complete understanding of the current state of the field.

The paper will also present the results of practical implementation of time series forecasting methods, including exponential smoothing, ARIMA and hybrid methods. Using real data and the presented methods, forecasting will be performed.

Keywords [1]

Forecasting, time series, exponential smoothing, machine learning, decision making, IT projects, ARIMA, hybrid method

1. Introduction

Time series forecasting is an important area of research in data analysis, scientific modeling, and machine learning. This topic has become especially relevant in the context of the modern information society under martial law, where data accumulation is occurring at a tremendous speed, and the ability to identify patterns in the dynamics of time series for the needs of IT project management is becoming a key issue for many industries.

Time series forecasting is widely used in finance, economics, manufacturing, healthcare, climatology, management, and many other industries. In business, accurate forecasts allow you to manage IT project inventory more efficiently, plan production processes, and adapt to changes in market conditions. The same forecasting is essential in supply chain management, ensuring effective planning in the organization [1]. In the medical field, various time series forecasting methods are widely used to analyze and predict various health indicators, incidence, improve the accuracy of diagnosing various diseases and other important parameters [2].

Information Technology and Implementation (IT&I-2023), November 20 -21, 2023, Kyiv, Ukraine
EMAIL: viktor.morozov@knu.ua (A.1); thedynkan@knu.ua (A.2); andrii.khlevnyi@knu.ua (A.3)
ORCID: 0000-0001-7946-0832 (A.1); 0009-0008-3123-2302 (A.2); 0000-0001-8942-6670 (A.3)

With the advent of big data and the development of machine learning methods, time series forecasting has become more accurate and flexible. Machine learning [3] for time series forecasting provides powerful tools for analyzing and predicting data dynamics over time. Algorithms trained on historical time series are able to detect hidden patterns and provide accurate forecasts, making them indispensable in industries ranging from finance to healthcare and especially for managing the development of modern IT projects and startups. Applying the same deep learning to financial time series forecasting can achieve high forecast accuracy [4].

In a rapidly changing business environment and uncertainty associated with various factors, such as economic and political events, time series forecasting is becoming an indispensable tool for making strategic management decisions. Thus, research in this area not only enriches our theoretical knowledge, but also has practical value, contributing to the development of more accurate and adaptive forecasting methods.

2. Analysis of the practical application of time series forecasting in management decision-making

The practical use of time series forecasting in management allows organizations to make informed decisions based on an analysis of future trends. For example, in business, time series forecasting can be a key tool for production planning and inventory management. Based on the analysis of previous data, forecasting systems are able to predict product demand, which provides IT companies with the ability to optimize production processes and avoid unnecessary costs.

In finance, time series forecasting is an important tool for developing investment management strategies. Investors and financial analysts can use forecasts to determine the future value of assets, make portfolio allocation decisions, and manage risks.

In marketing, time-series forecasts help predict demand for goods and services, enabling companies to effectively plan advertising campaigns and resources to meet consumer demands. This is especially important in today's dynamic market environment, where rapid changes in consumer demands can have a significant impact on the success of IT business development projects.

Time series forecasting also has applications in healthcare, allowing for optimized allocation of medical resources based on the projected demand for services and medicines. This improves the efficiency of the healthcare system and helps prevent resource shortages in critical situations.

Thus, the practical use of time series forecasting in management provides organizations with not only more accurate forecasting of future events, but also the ability to make more informed, data-driven decisions, which contributes to increased business efficiency and sustainability.

3. Analysis of the use of time series for business forecasting

Time series forecasting in business is one of the elements of planning and effective management. Predicting future trends helps companies adapt to changes in demand, market changes, and other influencing factors. In the field of inventory management, time series forecasts allow you to optimize inventory levels, avoiding excessive costs or shortages of goods.

Time series forecasting is especially important in retail, where changes in seasonality, trends, and consumer behavior can have a significant impact on business success. Companies use these forecasts to develop effective marketing strategies, plan promotions, and optimize pricing policies.

In finance, time series forecasting is important for budgeting, liquidity management, and assessing a company's financial stability. Accurate demand forecasts also contribute to the development of investment strategies and decisions on the efficient allocation of resources.

The use of time series forecasting methods in business is becoming especially important in the context of modern technologies and the availability of large amounts of data. Machine learning methods and statistical models allow for more accurate analysis and prediction of time series dynamics, which contributes to more flexible and adaptive management of IT projects in the face of constant changes in the business environment. Thus, time series forecasting is becoming one of the key tools for businesses aimed at sustainability and successful development.

4. Analysis of recent research and publications

This article [5] discusses a method of statistical analysis of time series known as ARIMA, which is used to predict future values of time series based on their historical data. ARIMA models are used in various industries to predict future values of time series. They are especially widely used in demand forecasting, providing management with reliable recommendations for supply chain management. In addition, ARIMA models can be used to predict future stock prices based on previous price data.

However, ARIMA has some drawbacks that may limit its effectiveness in some cases. For example, the choice of optimal values of the parameters "p" and "q" may depend on experience and modeling skills, which can be a subjective process. Obviously, this implies manual model tuning when using this forecasting tool. In addition, ARIMA models can be complex and have low explanatory power compared to simpler models such as exponential smoothing.

The conclusions of the article indicate that ARIMA models are useful in short-term time series forecasting, similar to other models. However, to achieve optimal results, model builders should be careful when choosing parameters and if simpler models have higher explanatory power, it is recommended to give them preference. Thus, the paper emphasizes the importance of careful model selection for time series forecasting, as well as the need to take into account historical data and experience when developing models to achieve the best forecasting results.

In the article [6] the authors present the AR, MA, and ARIMA models, apply these models to forecast risks on the National SME Stock Trading (New Third Board). The results of the analysis reflect the real situation and provide an opportunity for effective forecasting of financial risks.

The authors use historical price information to estimate the probability of distribution of future stock price changes in order to prevent possible significant losses. In particular, the authors processed first-order differences for a series of stock prices to obtain a stable differential sequence, obtained information on the stationarity of the series, etc. Based on this, the article emphasizes the importance of applying time series models in the field of finance for risk forecasting and making informed decisions. The use of ARIMA for risk prediction at NTB is a concrete example of the application of time series models in real-world scenarios. The time series analysis method proposed in this article can be a useful tool for predicting extreme situations and, thus, reducing financial risks.

However, it should be borne in mind that the use of these models and methods may be limited by the fact that past trends in time series do not always accurately predict future changes. Therefore, it is important to regularly review and analyze the results of the analyses performed and to consider additional factors that may affect future trends. This article [7] describes a study aimed at predicting the closing prices of several cryptocurrencies, such as Bitcoin, Ripple, Dash, Litecoin, and Ethereum. The authors used price, market capital, and volume information from the previous days, and applied various statistical methods and machine learning algorithms, including simple and multiple linear regression, as well as the Long Short-Term Memory (LSTM) model, to predict prices based on historical data. In the study, the authors analyzed the accuracy of various forecasting models using different metrics, in particular RMSE. Based on the comparative analysis of time series forecasting methods including ARIMA and LSTM, it is found that ARIMA showed the best result while LSTM showed the worst result, which confirms the preference of using ARIMA in our study.

Overall, the study emphasizes that cryptocurrency price forecasting is achievable using appropriate methods, and that the use of historical data can be an effective tool for analysts and investors.

This article [8] discusses the exponential smoothing method as an effective time series forecasting tool often used in finance, economics, and marketing. Exponential smoothing is a statistical method that uses past observations to predict future values of a time series by assigning exponentially decreasing weights to past data. The article highlights various variants of the method, such as simple, double, and triple exponential smoothing, and emphasizes its advantages in terms of forecast accuracy and computational efficiency. However, it is worth noting that the method has limitations, especially in the case of complex patterns or outliers in the time series data, and emphasizes the need for judicious use of the method depending on the characteristics of the time series data and forecasting goals.

According to this article [9], the use of an exponential smoothing model in forecasting the oil transportation market proves to be an effective approach. The author refined the existing model and achieved more accurate forecasts of the average time charter equivalent of a tanker along different routes of oil transportation in the world oceans during the crisis period of 2015-2019. According to the study, the proposed method demonstrates better accuracy compared to naive methods, autoregressive methods and machine learning models, according to all applied error metrics. Thus, this method can be successfully applied for commercial purposes by tanker fleet operators and charterers with an accuracy rate of 71%. According to the article [10], the common statement that ARIMA models are more widely applicable than exponential smoothing models is inaccurate. In fact, both classes of models overlap and complement each other, and each has its own strengths and weaknesses. Linear exponential smoothing models are special cases of ARIMA models, but nonlinear exponential smoothing models do not have an equivalent analog in ARIMA. The article argues that all state space models with exponential smoothing are non-stationary, so if a stationary model is required, it is better to use ARIMA models. However, one of the advantages of exponential smoothing models is that they can be nonlinear and better model time series that exhibit nonlinear characteristics, including heteroscedasticity. Conclusions from the article: ARIMA and exponential smoothing models are not mutually exclusive, and each of them has its advantages. The optimal choice of model depends on the specific conditions of the task, accuracy requirements, and data features, which emphasizes the need to adapt the model to a specific context.

Article [11] describes the development of effective one-dimensional models for predicting the spread of tuberculosis in Taiwan. Using monthly tuberculosis case data from 2005 to 2020, various modeling methods were compared, including SARIMA, ETS, and hybrid SARIMA-ETS algorithms. After evaluating the accuracy metrics, the optimized SARIMA, ETS and the hybrid SARIMA-ETS models were selected as candidate models, respectively. The study revealed that the hybrid SARIMA-ETS outperforms ARIMA and ETS in short-term forecasts, while ETS demonstrates higher performance in long-term forecasting. It is worth mentioning that in short- or medium-term forecasting, all models performed better than in long-term forecasting. Overall, the article emphasizes that time series modeling is a useful tool for predicting the spread of tuberculosis and planning the efficient use of resources in the fight against the disease. The conclusions of the study indicate the superiority of the hybrid SARIMA-ETS model and the ETS model in predicting tuberculosis incidence in Taiwan, with the latter showing higher performance in the long run.

Papers [12,13] address the issues of state prediction in the development and management of IT projects and startups. It shows the use of neural networks for forecasting, which can also be attributed to the use of machine learning methods, but the use of time series is not specified.

The article [14] examines the effectiveness of different models for forecasting time series related to the COVID-19 pandemic. Three models - autoregressive integrated moving average (ARIMA), artificial neural networks (ANN), and error, trend, and seasonality (ETS) models - are analyzed by the authors. The effectiveness of each model is evaluated based on real data for the period from January 2020 to January 2021, covering two waves of the pandemic.

The results of the study show that ETS and ARIMA models demonstrate higher efficiency in forecasting COVID-19 time series compared to ANN models and hybrid variants. The authors also conclude that the best approach is to use only one of the ETS or ARIMA models for time series forecasting, instead of complex hybrid models, which can significantly save time and resources spent on data analysis and forecasting pandemic trends. Overall, the article demonstrates that ETS and ARIMA models can be effective for forecasting COVID-19 time series, and that using one of these models can be more efficient than complex hybrid models. This approach can help accelerate data analysis and forecasting of pandemic trends, which is important for decision-making in the health and economic sectors. However, it is important to note that the results of the study may be unique to the COVID-19 time series and may not necessarily be transferable to other types of time series.

5. Using the exponential smoothing method

Exponential Smoothing is one of the basic approaches to time series forecasting. This method is based on the weighted consideration of previous values of the time series using exponential coefficients. The basic idea is that each observation of the time series is assigned a certain weight, which decreases exponentially with the increase of time backward. Exponential smoothing has three main components: level, trend, and seasonality, if present.

- Level: This is the current estimate of the average value of the time series.

- Trend: A change in a level over time that can be positive, negative, or zero.
- Seasonality: If the time series is subject to repeated cyclical fluctuations, seasonality takes these patterns into account.

There are several variants of exponential smoothing, such as simple exponential smoothing, double exponential smoothing, and the Holt-Winters method. Let's look at their main features:

1. Simple exponential smoothing:
   - Simple exponential smoothing is used when the time series does not have a pronounced seasonality and trend.
   - Simple exponential smoothing formula [15]:

$$F(t+1) = \alpha * Y(t) + (1-\alpha) * F(t), \tag{1}$$

where F(t) is the forecast for the current period; Y(t) is the value of the series for the current period; α is the smoothing coefficient.

2. Double exponential smoothing:
   - Double exponential smoothing is used when there is a trend in the data but no seasonality.
   - It contains two levels of smoothing: one for observations and one for the trend.
   - Formula for level, trend and forecast [16]:

$$S(t) = \alpha * Y(t) + (1-\alpha) * (S(t-1) + b(t-1)), \tag{2}$$

$$b(t) = \beta * (S(t) - S(t-1)) + (1-\beta) * b(t-1), \tag{3}$$

$$F(t+m) = S(t) + m * b(t), \tag{4}$$

where S denotes the smoothed value; Y denotes the time series; α, β are smoothing coefficients for the smoothed value and the trend; F(t+m) is the forecast for the period t+m; m is the number of periods ahead.

3. Holt-Winters method:
   - Used when the data has both a trend and seasonality.
   - It contains three levels of smoothing: for observations, trend, and seasonality.
   - Basic equations of the Holt-Winters method [17]:

The formula for the level:

$$S(t) = \alpha * (Y(t) - I(t-L)) + (1-\alpha) * (S(t-1) + b(t-1)), \tag{5}$$

Formula for the trend:

$$b(t) = \beta * (S(t) - S(t-1)) + (1-\beta) * b(t-1), \tag{6}$$

Formula for seasonality:

$$I(t) = \gamma * (Y(t) - S(t)) + (1-\gamma) * I(t-L), \tag{7}$$

where L denotes the period; I denotes the estimate of the seasonal component; γ is the smoothing factor for seasonality.

The exponential smoothing method allows you to account for changes in the data by assigning weights to different observations. Exponential smoothing is a simple but effective method for modeling time series and can be adapted to account for various factors, such as trend and seasonality, for more accurate predictions.

6. Using the ARIMA method

ARIMA (Autoregressive Integrated Moving Average) is a time series forecasting method that combines three main components: autoregression (AR), integration (I), and moving average (MA). This method was developed to model complex time series structures, including trends and seasonality:

- Autoregression (AR): This component displays the dependence of the current value of the time series on its previous values. The parameter "p" determines the number of lags that are taken into account in the model.
- Integration (I): This component represents the degree of differentiation of the time series required to achieve stationarity. If the time series is non-stationary, differentiation is applied. The parameter "d" determines the number of times the series has to be differentiated.
- Moving Average (MA): This component models the dependence of the current value of the time series on the random errors of previous time points. The parameter "q" determines the number of error lags taken into account in the model.

The ARMA model can be represented by the following equation [18]:

$$Y(t) = \theta 1Y(t-1) + \theta 2Y(t-2) + ... + \theta pY(t-p) + \alpha 1e(t-1) + \alpha 2e(t-2) + ... + \alpha qe(t-q) + e(t), \quad (8)$$

where Y(t) is the current value of the time series; θ1, θ2, ..., θp - autoregression coefficients that reflect the influence of previous values on the current value; α1, α2, ..., αq - moving average coefficients that reflect the influence of previous errors on the current value; Y(t-1), Y(t-2), ..., Y(t-p) - previous values of the time series; e(t-1), e(t-2), ..., e(t-q) - previous error values; e(t) - random error or noise.

The ARIMA model extends the basic ARMA model by introducing an additional parameter "d", which is the order of integration. This parameter determines the number of times the differentiation should be performed to bring the series to steady state. The idea of differentiation is to calculate the difference between the current and previous values of the series. So, the ARIMA model includes three key parameters: "p" for the autoregression order, "d" for the integration order, and "q" for the moving average order. ARIMA allows you to model a wide range of time series, including both short-term and long-term trends, as well as seasonal fluctuations. The optimal values of the p, d, q parameters can be selected using methods such as autocorrelation and partial autocorrelation analysis or an automated approach.

7. Using the hybrid method

Before describing the hybrid method, it is important to mention the accuracy metric that will be used as an accuracy metric for forecasting methods.

MAPE (Mean Absolute Percentage Error) is a metric for assessing the accuracy of forecasts, especially in the context of time series. This metric measures the average absolute percentage error between actual and forecast values. A low MAPE value indicates a high forecast accuracy. However, the context of a particular task and the peculiarities of the time series should be taken into account when interpreting this metric.

MAPE is calculated using the following formula [19]:

$$MAPE = (1 / n) * \sum[(|Y(t) - \hat{Y}(t)|) / |Y(t)|] * 100, \quad (9)$$

where n - sample size; Y(t) - represents the actual value of the time series at time t; $\hat{Y}(t)$ - represents the forecast value at time t.

A hybrid time series forecasting method is a combination of two or more forecasting methods in order to improve the accuracy of forecasts. In this case, the combination of ARIMA and exponential smoothing methods is considered. The hybrid method combines these two approaches, allowing the different characteristics of each method to be taken into account. The process of creating a forecast in the hybrid method may include the following steps:

- ARIMA forecast: For example, ARIMA is the first to be used to forecast a time series for a specific time horizon.
- Exponential smoothing forecast: Exponential smoothing is then applied to create an additional forecast.
- Forecast combination: The resulting forecasts from ARIMA and exponential smoothing are combined. This can be accomplished by, for example, assigning different weights to each of the forecasts.

The main advantage of the hybrid method is that it allows the different features and strengths of each method to be taken into account, which can lead to more accurate time series forecasts.

In some cases, a simple approach to combining methods is used, which assumes equal weights for each method. For example, if two forecasting methods are used, each of them can be assigned a weight of 0.5, which means that each method has equal influence on the final forecast.

However, this paper plans to use a more complex and adaptive approach, namely weighted arithmetic mean, where the weights depend on the MAPE metrics of each method. This approach is more flexible as it allows to consider the performance of each method on the time series.

When the weights depend on MAPE, it allows us to consider how well each method performs on specific characteristics of the data. If one method performs better on certain parts of the time series, it can be assigned a higher weight in the final combination. This makes the approach more adaptive to changes in the structure of the data and can improve the overall accuracy of the forecasts.

This weighting method is based on empirical data and provides a more flexible and intelligent approach to combining forecasting methods, which can lead to more effective time series forecasts.

If the difference in MAPE values between two forecasting methods is significant, it may indicate that one of the methods is more efficient and accurate in a given context. In such a case, the use of a hybrid model may be unnecessary and it is more reasonable to favor the method that demonstrates a lower MAPE. Let describe the formula for building the hybrid model, given that there are two forecasting methods exponential smoothing and ARIMA. Let denote the forecasts obtained using exponential smoothing and ARIMA as $Y_{Exponential}$ and Y_{ARIMA}, respectively, and their MAPEs as $MAPE_{Exponential}$ and $MAPE_{ARIMA}$. The formula for the hybrid model can be represented as follows:

$$Y_{Hybrid} = \frac{\left(Y_{Exponential} * \frac{1}{MAPE_{Exponential}}\right) + \left(Y_{ARIMA} * \frac{1}{MAPE_{ARIMA}}\right)}{\frac{1}{MAPE_{Exponential}} + \frac{1}{MAPE_{ARIMA}}} \tag{10}$$

The weights for each method in this formula depend on inverse MAPE values, which means that the lower the MAPE for a particular method, the greater the contribution of that method to the hybrid forecast. As an option, different modifications of the formula are possible, such as using in it rounding of values, squaring and others. In the practical part, the formula as presented above will be used.

In this context, this formula will be used to create a hybrid time series forecast, taking into account the forecasts obtained using the exponential smoothing and ARIMA methods and considering their relative performance based on MAPE. In the practical part of the paper, two approaches to building a hybrid model for time series forecasting will be carried out.

1. Simple approach with equal weights: In the first approach, a simple method will be used where each of the two forecasting methods (ARIMA and exponential smoothing) will be assigned an equal weight, such as 0.5. This means that the influence of each method will be considered equal in the formula for the hybrid model.

2. Improved approach with MAPE-dependent weights: The second approach will use an improved formula that takes into account the weights that depend on the MAPE metrics of each forecasting method. This will allow to adapt the weights to the performance of each method on the time series.

Each approach will be applied to real time series data. In the end, after applying each of the two approaches to build the hybrid model, the MAPE value for each approach will be calculated. This will evaluate the accuracy of each method and compare the results of the two approaches.

Comparing the MAPE for the two approaches will determine which of the forecasting methods (equal weights or MAPE-dependent weights) performs better for a particular time series. This comparison is a key step in evaluating the performance of the hybrid model and deciding whether one approach is preferable to the other in a particular context.

8. Algorithm for building a time series forecast

The following algorithm for building a time series forecast will be used, which includes the use of three main methods: ARIMA, exponential smoothing, and a hybrid method, which is a combination of ARIMA and exponential smoothing. The main process is divided into several steps:

1. Data splitting: The original time series is split into training and test samples.

2. Model training: On the training sample, forecasts are made for each of the three methods: ARIMA, exponential smoothing, and a hybrid method. Each method forecasts time series values based on training data.

3. Comparison and selection of the main model: The forecasts obtained by each method are compared to the corresponding values of the test set. This allows to evaluate the accuracy of each method on data that they did not use in the training process. The MAPE metric, which measures the average absolute percentage error between the predictions and the actual values of the time series, is used to select the main model. The model with the “best” MAPE score on the training data is selected as the main model.

4. Forecast on all data: The selected main method forecasts time series values over the entire data range, which is the input time series.

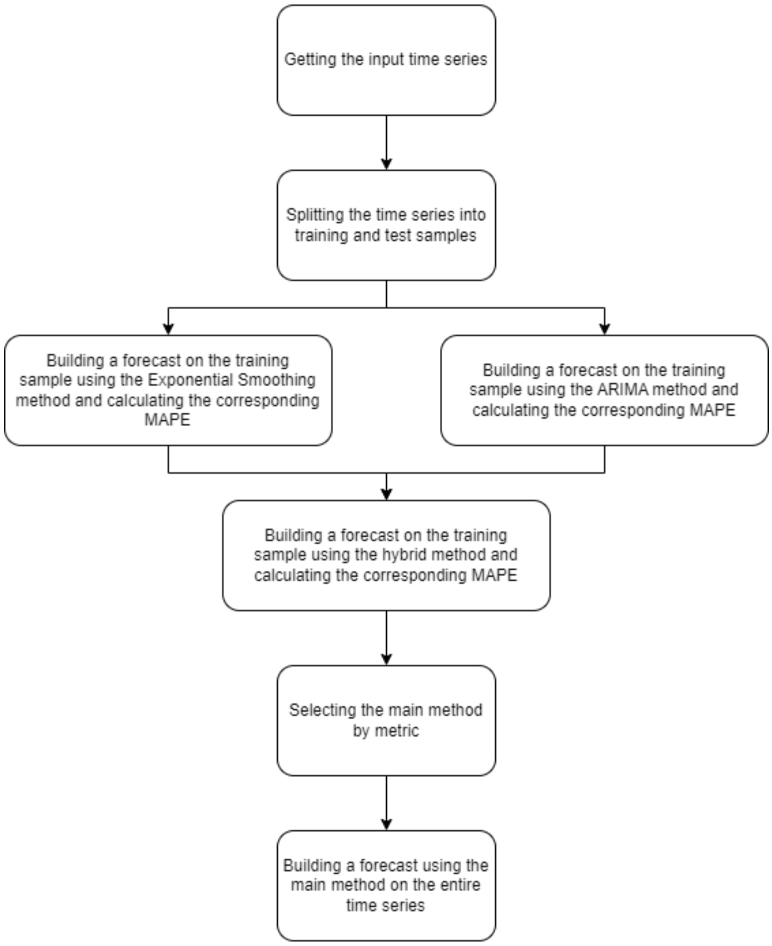

Figure 1: Scheme of the algorithm used to build a time series forecast

This approach allows you to choose the best forecasting model optimized for specific characteristics of the time series and get more accurate and stable forecasts for further use in analysis and planning.

9. Experimental research

For a practical demonstration of building a time series forecast, specific data from the Kaggle website [20] were selected, which are daily time series for the last 5 years for various indicators of website traffic with educational materials on statistical forecasting. In this case, Python, various libraries for working with data and statistical analysis, as well as the Jupiter Notebook development environment were used to analyze and build a time series forecast. Here are some key tools:

- Python: It is a high-level programming language that provides powerful tools for data analysis, statistical modeling, and machine learning.
- Pandas: A library for data processing and analysis that provides convenient data structures and tools for manipulating time series.
- Matplotlib: A data visualization library that allows you to analyze trends and patterns in time series.
- Statsmodels: A library for statistical modeling, including the implementation of ARIMA models and other statistical methods.
- Jupyter Notebook: An interactive development environment that provides a convenient way to combine code, textual explanations, and visualizations in one document. This allows you to analyze data and demonstrate results step by step.

After loading all the necessary libraries, proceed to loading the input time series. For forecasting purposes, the indicator with the daily number of downloaded pages - Page.Loads - is selected. From this time series, the first 60 values are selected, on the basis of which the forecast for the next 7 values will be built, thereby representing a 7-day forecast.

```
web = pd.read_csv('daily-website-visitors.csv')
ts = web["Page.Loads"].str.replace(',','')
ts = ts.astype(float)
ts = ts.head(60)
```

Print the input time series plot.

```
plt.figure(figsize=(24,10))
ts.plot()
```

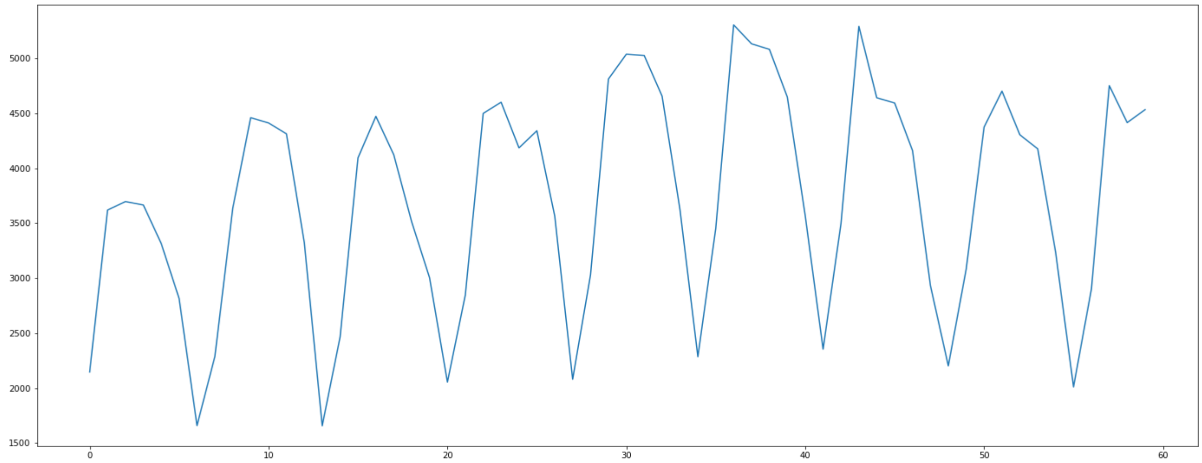

Figure 2: Input time series

Next, the forecast horizon is determined, i.e. the period for which the forecast will be made. In this case, the forecast horizon is 7. After that, the time series is divided into training and test samples.

```
horizon = 7
train_length = len(ts) - horizon
ts_train = ts.iloc[0:train_length]
ts_test = ts.iloc[train_length:len(ts)]
```

Next, the model is built using the Exponential Smoothing method on the training sample. According to the input data, it is seen that there is seasonality, with a period value of seven, which makes it possible to specify the corresponding values in the model. After building the model, the predicted values are obtained.

```
model = ExponentialSmoothing(ts_train, seasonal_periods=7, trend='additive', seasonal='additive').fit()
forecastexp = model.predict(0, len(ts_train) + horizon - 1).iloc[len(ts_train):]
```

Next, for clarity, show a plot showing the training, test, and obtained values using the exponential smoothing model.

```
plt.figure(figsize=(24,10))
plt.plot(ts_train, label="Train")
plt.plot(forecastexp, label="Prediction")
plt.plot(ts_test, label="Test")
plt.legend(loc="best")
plt.show()
```

Next, calculate the Mean Absolute Percentage Error (MAPE) for the built exponential smoothing model and display its value. MAPE is a metric that measures the percentage error of a prediction relative to actual values.

*MAPEexp = np.mean(np.abs((ts_test - forecastexp)/ts_test))*100*

print("MAPEexp is ", MAPEexp)

The resulting values:

MAPEexp is 2.965060577251055

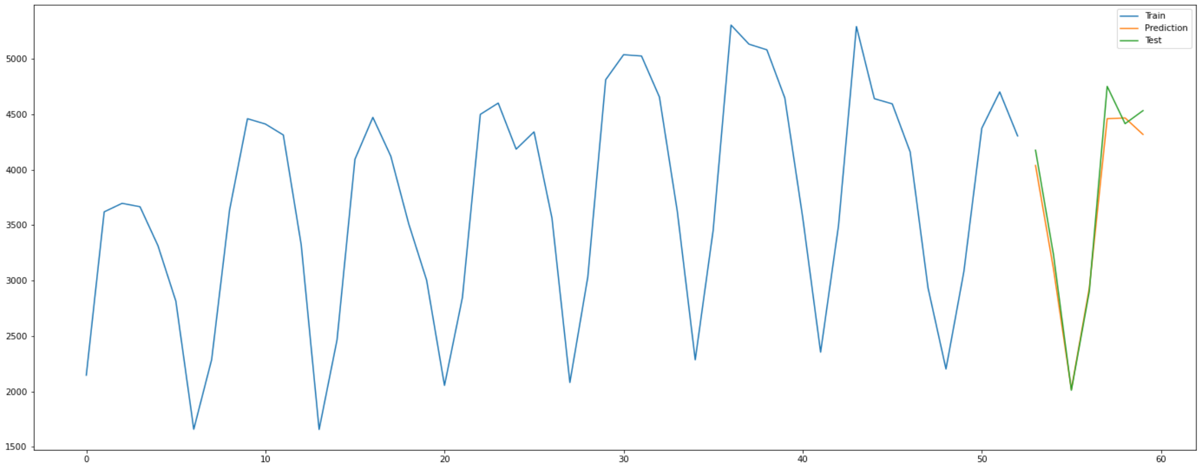

Figure 3: Exponential smoothing model on the training set

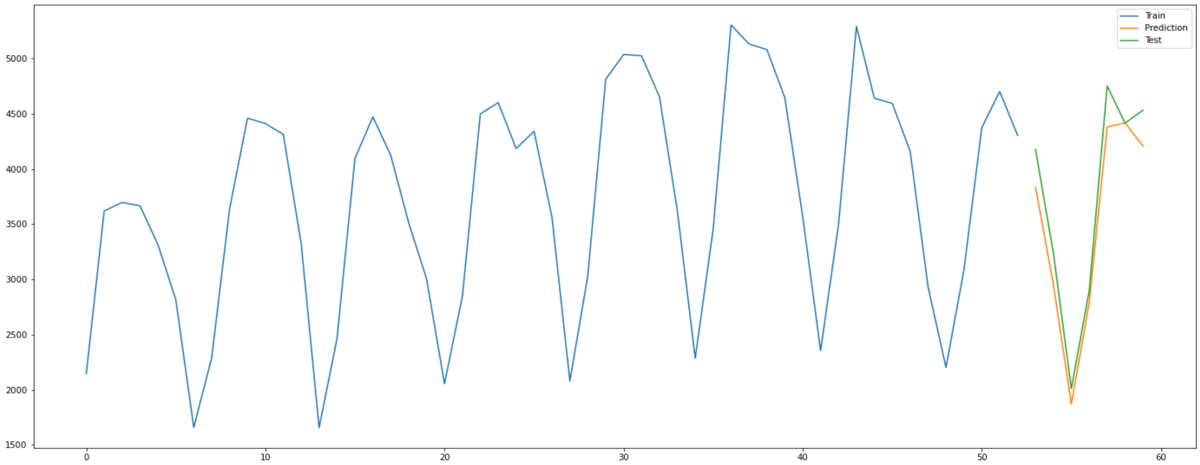

Figure 4: ARIMA model on the training sample

Next, the ARIMA automatic parameter selection model is built using the pmdarima library. In this case, the model will automatically select the optimal values of the parameters (p, d, q) for the ARIMA model based on minimizing the information criterion. Additional parameters are set by default. The resulting model is used to build a forecast, and a plot is drawn showing the training, test, and forecast values using the ARIMA model.

```
model = pm.auto_arima(ts_train, m=7, seasonal=True)
forecastarm = model.predict(n_periods=horizon)
plt.figure(figsize=(24,10))
plt.plot(ts_train, label="Train")
plt.plot(forecastarm, label="Prediction")
plt.plot(ts_test, label="Test")
plt.legend(loc="best")
plt.show()
```

The corresponding MAPE is calculated.

*MAPEarm = np.mean(np.abs((ts_test - forecastarm)/ts_test))*100* *print("MAPEarm is ", MAPEarm)*

The resulting values:

MAPEarm is 6.107361536607362

Move on to the building of a hybrid model, which is a combination of forecasts obtained from the exponential smoothing model and the ARIMA model. For example, let's build two types of hybrid models, one model using the simple approach with equal weights and one using the improved approach with weights depending on MAPE. Print their MAPE values on the screen.

*forecasthyb = 0.5 * forecastexp + 0.5 * forecastarm*

*forecasthybmape = ((forecastexp * (1/MAPEexp)) + (forecastarm * (1/MAPEarm))) / ((1/MAPEexp) + (1/MAPEarm))*

MAPEhyb = np.mean(np.abs((ts_test - forecasthyb)/ts_test))*100

print("MAPEhyb is ", MAPEhyb)

*MAPEhybmape = np.mean(np.abs((ts_test - forecasthybmape)/ts_test))*100*

print("MAPEhybmape is ", MAPEhybmape)

Corresponding MAPE values:

MAPEhyb is 4.329272723465154

MAPEhybmape is 3.7134184220187154

After calculating the MAPE values for the two approaches, it can be observed that the improved approach, which takes into account the MAPE-dependent weights, has a lower MAPE value. This indicates the higher accuracy of the improved hybrid model compared to the simple approach. Plotting the training, test and forecast values.

plt.figure(figsize=(24,10))

plt.plot(ts_train, label="Train")

plt.plot(forecasthybmape, label="Prediction")

plt.plot(ts_test, label="Test")

plt.legend(loc="best")

plt.show()

Next, the main forecasting method is selected based on the MAPE values for each method. In this case, since the exponential smoothing method showed the lowest MAPE value, it is selected as the main forecasting method. Then the forecast for the entire time series is built using the selected main method. A plot with the final forecast and input data is displayed.

model = ExponentialSmoothing(ts, seasonal_periods=7, trend='additive', seasonal='additive').fit()

forecast = model.predict(0, len(ts) + horizon - 1)

plt.figure(figsize=(24,10))

plt.plot(ts,label="Time Series")

plt.plot(forecast, label= "Forecast")

plt.legend(loc="best")

plt.show()

As a result, a plot was created that shows the initial data of the time series and the resulting forecast, built using the selected forecasting method.

10. Conclusions

The article discusses the relevance of time series forecasting and its application in various fields of activity. Time series forecasting is one of the key tools for decision-making in various industries, including finance, economics, healthcare, logistics, and many others. Accurate time series forecasts help to manage resources more efficiently, make informed management decisions, and plan future operations. The study considered specific time series forecasting methods, such as exponential smoothing and the ARIMA model. Also, the relevant literature was analyzed to justify the choice of time series forecasting methods, determine their applicability and evaluate their performance based on existing research and practical experience.

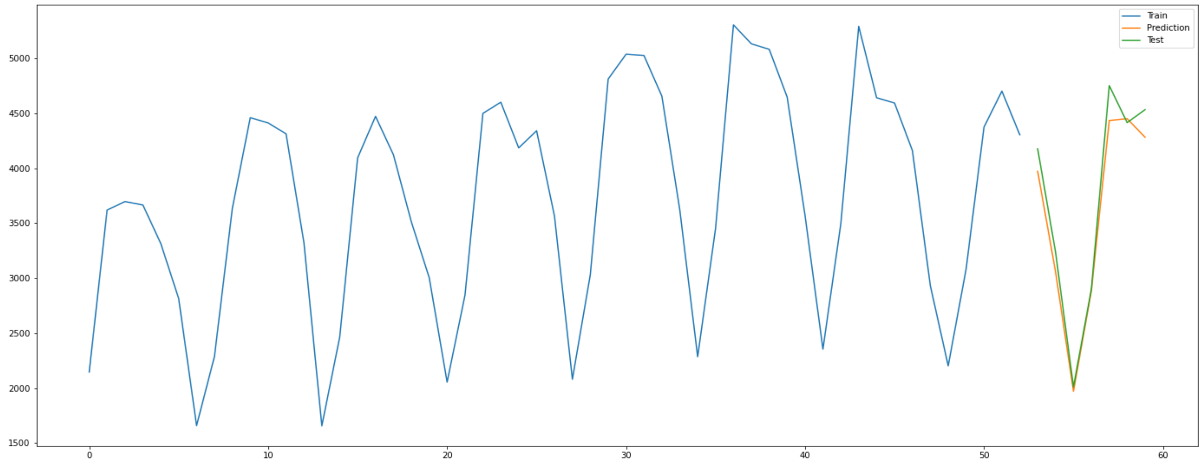

Figure 5: Hybrid model on the training set

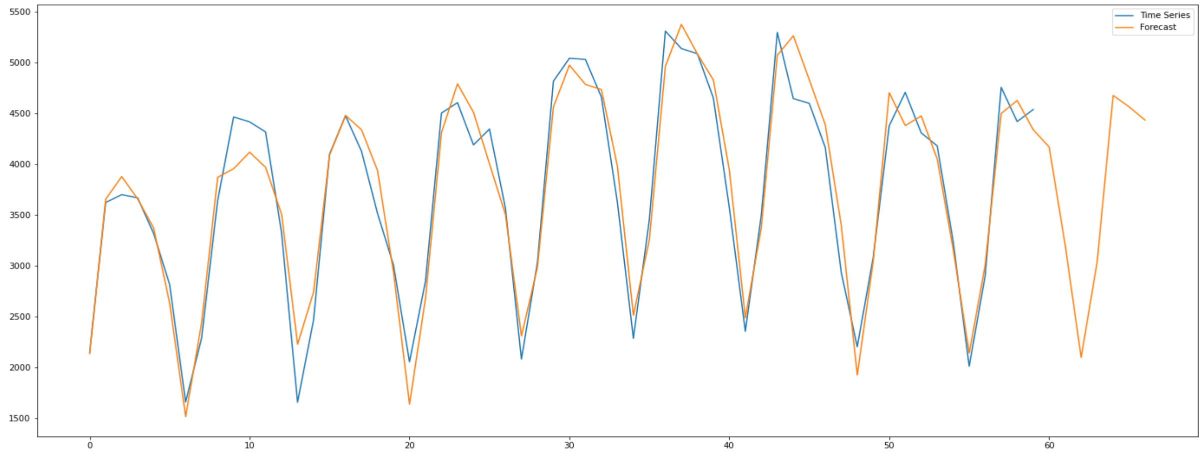

Figure 6: Initial series and forecast values

In conclusion, this study on time series forecasting emphasizes the importance of choosing an appropriate method to achieve accurate and reliable results. In this study, time series forecasting was practically implemented using various forecasting methods, such as exponential smoothing, ARIMA model with automatic parameter selection auto-arima, and their hybrid combination.

To build the hybrid model, an improved approach was presented that takes into account the weights depending on the MAPE metrics for each of the forecasting methods (ARIMA and exponential smoothing). In the practical part of the study, model building was conducted using both approaches (simple with equal weights and improved with MAPE-dependent weights) as a result, the improved approach showed a better result. This may indicate that the use of the improved approach with MAPE-dependent weights allows creating more adaptive and accurate hybrid time series forecasting models. Experiments on training and test data revealed that the exponential smoothing method demonstrated the best performance in terms of minimizing forecast errors measured by MAPE. This result allows to conclude that this method is applicable for a specific time series in this case. It is important to note that the study also considered the auto ARIMA model and the hybrid method. They provide alternative options, and the choice between them may depend on the specific characteristics of the time series and forecast requirements. In other cases, an analyst may choose other forecasting models based on other, subjective considerations.

The results of the study provide practical recommendations for applying time series forecasting methods in real-world scenarios. Determining the optimal method is important for management decision-making, planning, and effective resource management in various business sectors.

11. References

[1] Robert Fildes, Paul Goodwin, Michael Lawrence, The design features of forecasting support systems and their effectiveness, Decision Support Systems, Volume 42, Issue 1, 2006, Pages 351-361, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2005.01.003.

[2] Bui, C., Pham, N., Vo, A., Tran, A., Nguyen, A., Le, T. (2018). Time Series Forecasting for Healthcare Diagnosis and Prognostics with the Focus on Cardiovascular Diseases. In: Vo Van, T., Nguyen Le, T., Nguyen Duc, T. (eds) 6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6) . BME 2017. IFMBE Proceedings, vol 63. Springer, Singapore. https://doi.org/10.1007/978-981-10-4361-1_138

[3] Oleksandr Gerasymov, Machine learning for time series forecasting – URL: https://codeit.us/blog/machine-learning-time-series-forecasting

[4] He, K.; Yang, Q.; Ji, L.; Pan, J.; Zou, Y. Financial Time Series Forecasting with the Deep Learning Ensemble Model. Mathematics 2023, 11, 1054. https://doi.org/10.3390/math11041054

[5] Understanding ARIMA Models for Machine Learning – URL: https://www.capitalone.com/tech/machine-learning/understanding-arima-models/

[6] F. Wang, M. Li, Y. Mei and W. Li, "Time Series Data Mining: A Case Study With Big Data Analytics Approach," in IEEE Access, vol. 8, pp. 14322-14328, 2020, doi: 10.1109/ACCESS.2020.2966553.

[7] S. I. Ansari and V. H Y, “System for Analysis and Prediction of Trends in Cryptocurrency Market”, EAI Endorsed Trans IoT, vol. 7, no. 27, pp. 1–5, Mar. 2022.

[8] Nikhil Malkari, Exponential Smoothing: A Method for Time Series Forecasting – URL: https://medium.com/@nikhilmalkari18/exponential-smoothing-a-method-for-time-series-forecasting-7ea35ca2c781

[9] George Moiseev, Forecasting oil tanker shipping market in crisis periods: Exponential smoothing model application, The Asian Journal of Shipping and Logistics, Volume 37, Issue 3, 2021, Pages 239-244, ISSN 2092-5212, https://doi.org/10.1016/j.ajsl.2021.06.002

[10] Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03

[11] Kuan M-M. 2022. Applying SARIMA, ETS, and hybrid models for prediction of tuberculosis incidence rate in Taiwan. doi 10.7717/peerj.13117

[12] Morozov V., Kolomiiets A., Development of a model for evaluating the effectiveness of innovative startups based on information cycles and using neural networks. Indonesian Journal of Electrical Engineering and Computer Science, 2021, 23(1), pp. 396–404 doi:10.11591/ijeecs.v23.i1.pp396-404

[13] Morozov V., Elaboration of Communication Models in Adaptive Systems of E-learning with Using Neural Networks. SIST - 2021 IEEE International Conference on Smart Information Systems and Technologies, 2021. doi: 10.1109/SIST50301.2021.9465945

[14] Safi, Samir K. and Sanusi, Olajide Idris. ‘A Hybrid of Artificial Neural Network, Exponential Smoothing, and ARIMA Models for COVID-19 Time Series Forecasting’. 1 Jan. 2021 : 25 – 35

[15] Ostertagova, Eva & Ostertag, Oskar. Forecasting Using Simple Exponential Smoothing Method. Acta Electrotechnica et Informatica. 12. 62–66. (2012). 10.2478/v10198-012-0034-2.

[16] SAP HANA Predictive Analysis Library (PAL), Double Exponential Smoothing – URL: https://help.sap.com/docs/SAP_HANA_PLATFORM/2cfbc5cf2bc14f028cfbe2a2bba60a50/4665493227564e8bb11ab11a95c1d8a3.html

[17] Yibin Ng, Forecasting Stock Prices using Exponential Smoothing – URL: https://towardsdatascience.com/forecasting-stock-prices-using-exponential-smoothing-b37dfe54e8e9

[18] Rushad Faridi, Time Series Analysis – URL: https://bookdown.org/rushad_16/TSA_Lectures_book/general-arma-models.html

[19] What Is MAPE? A Guide to Mean Absolute Percentage Error – URL: https://www.indeed.com/career-advice/career-development/what-is-mape

[20] Daily website visitors (time series regression) – URL: https://www.kaggle.com/datasets/bobnau/daily-website-visitors/