Website Traffic Forecasting Using SARIMA

DEVI AHILYA VISHWAVIDHYALYA INDORE

In Partial Fulfillment of the Requirements
for the Degree of

MASTER OF TECHNOLOGY
DATA SCIENCE

<u>Forecasting Methods</u>

Submitted By:

Afifa Jafari(DS7A-2202)

Under the Supervision of

Dr. Vandit Hedau
Assistant Professor

SCHOOL OF DATA SCIENCE AND FORECASTING

2022-23

Abstract

Accurate website traffic forecasting is crucial for businesses and organisations to plan resource allocation, marketing tactics, and infrastructure requirements. In this paper, we offer a forecasting method for predicting website traffic patterns based on the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMA) model. To capture the underlying seasonality, trends, and other significant elements, SARIMA employs both time series analysis and external factors.

The proposed SARIMA model has various advantages for anticipating website traffic. It can detect seasonality and trends in website visitor patterns, allowing firms to plan promotions and optimise resource allocation accordingly. Exogenous variables allow the model to consider external factors that may affect website traffic, resulting in a more thorough and accurate forecast.

In conclusion, by combining time series analysis with external variables, the SARIMA model provides an excellent strategy for website traffic predictions. Because the model can capture seasonality, trends, and the impact of external events, it is a great tool for businesses and organisations looking to optimise their online presence and make data-driven decisions.

Introduction

Websites are critical to the success of businesses and organisations in today's digital world. Understanding and properly forecasting website traffic patterns is critical for resource allocation, marketing tactics, and infrastructure optimisation. Forecasting website traffic enables organisations to deploy resources more efficiently, plan marketing initiatives, and deliver a consistent user experience.

Because of the changing nature of internet platforms, projecting website traffic is a difficult undertaking. Seasonality, trends, marketing initiatives, social media activities, holidays, and other external elements all have an impact on website traffic. Traditional forecasting approaches frequently fail to capture the complex interactions between these variables, resulting in erroneous predictions.

In this research project, we look at a forecasting method that uses the capability of the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables(SARIMA) model to forecast website traffic. SARIMA combines time series analytic techniques with the incorporation of external variables to provide a comprehensive framework for properly capturing and forecasting website traffic patterns.

The SARIMA model is well-suited for projecting website traffic since it takes into account both temporal dependencies and the influence of external factors. The model can uncover underlying seasonality, patterns, and abnormalities by analysing past website traffic data, allowing businesses to forecast peak periods and arrange their operations appropriately. , including exogenous variables such as marketing campaigns or vacations helps the model to capture the impact of these factors on website traffic, hence improving the accuracy of the forecasts

This report aims to provide a detailed understanding of the SARIMA model and its application in website traffic forecasting. We will discuss the methodology used to implement the model, including data collection, preprocessing, and model training. Additionally, we will explore the evaluation metrics used to assess the accuracy of the forecasts and compare the results with other forecasting techniques.

The insights gained from accurate website traffic forecasting using SARIMA can significantly benefit businesses and organizations. By having a clear understanding of future website traffic patterns, companies can optimize their resource allocation, plan marketing campaigns effectively, and ensure their websites can handle anticipated surges in traffic. This report aims to empower businesses with the knowledge and tools necessary to make data-driven decisions and enhance their online presence.

LITERATURE REVIEW

To increase accuracy, Khashei and Bijari [10] suggested an ensemble forecasting model. Their model consists of an Auto-Regressive Integrated Moving Average (ARIMA) statistical model and an Artificial Neural Network (ANN) model. They recognised the ANN model's restriction in processing linear data, which inspired them to use a hybrid model based on Multi-Layer Perceptron (MLP) to analyse the time-series data's non-linear component. The linear component of time series data is handled by the ARIMA model. To improve overall forecasting accuracy, a hybrid forecasting model [10] integrating the ARIMA and ANN has been created. Their model was evaluated on three well-known data sets, and it performed better on all of them. The ARIMA model was used to create the necessary data for their methodology, which was then put into the ANN model to forecast the future. U. Kumar and V. Jain [11] studied the ability of the Auto-Regressive Moving Average (ARMA) and ARIMA models to forecast the future.

They experiment with several information criteria to fine-tune the model parameters p, q, and d, such as AIC (Akaike Information Criterion), HIC (Hannon-Quinn Information Criterion), BIC (Bayesian Information Criterion), and FPE (Final Prediction Error). They also take AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) plot data into account when determining the best performing model.

In their work, they investigated various model performance evaluation metrics such as MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), and RMSE (Root Mean Squared Error). M. Dastorani et al. [12] conducted a comparison of various statistical models such as AR (Auto-Regressive), MA (Moving Average), ARMA, ARIMA, and SARIMA.

Wikipedia Page Web Traffic Prediction -IEEE International 2018 (Big Data Conference (Big Data) [1]) by Navyasree Petluri, Eyhab Al-Masri

During the prediction model construction process, the system effectively rebuilt the old model and incorporated new features to notice increases in model efficiency.

New features were used in various combinations.

1) The median of each time series' set window length as an independent feature for measuring weekly, monthly, quarterly, and yearly page popularity.

2) Golden ratio-based median of medians of varied time frame windows. The study examined the obtained results and compared the accuracies in various scenarios to determine the significance of each attribute. As a next step, we will try to figure out how to tweak parameters in the present model in order to achieve better outcomes.

Wikipedia's pageview API is the data used for this project. That data contains daily page visits as a time se-ries to any post. Latest data is obtained through this API. The data is returned in JSON format. The fields extracted from this data are the recorded Dates and Visits on that date. It converts this data into a data frame and fits into the predictive model. [7,8]

2021 6th International Conference on Inventive Computation Technologies (ICICT)

- Mohammmad Asifur Rahman Shuvo, Muhtadi Zubair Afsara Tahsin Purnota , Sarowar Hossain, Muhammad Iqbal Hossain

Study intended to discover the most acceptable forecasting model based on time-series which helps us to anticipate future traffic data when there is enough dataset is available. Having this goal in mind, study began to search for models based on prediction, which would enable us to predict the value data. However, after further investigation, we discovered that it was not a prediction but rather forecasting, so we concentrated on that. Study discovered so many time-series forecasting models that it made our work both laborious and enjoyable.

Numerous case studies have demonstrated the effectiveness of the SARIMA model in forecasting website traffic. These studies have been conducted across various industries, including e-commerce, media, and online services. The SARIMA model has been used to predict website traffic during peak periods, optimize server capacity, plan marketing campaigns, and improve overall user experience. The case studies highlight the practical applications and benefits of utilizing the SARIMA model in real-world scenarios.

The selection of appropriate exogenous variables is crucial for the effectiveness of the SARIMA model in website traffic forecasting. Researchers have explored various external factors and their impact on website traffic. These include marketing activities, search engine trends, social media interactions, weather conditions, holidays, and economic indicators. The inclusion of these variables in the SARIMA model allows businesses to capture the complexity and interdependencies that affect website traffic patterns.

To assess the accuracy of website traffic forecasts, researchers commonly employ evaluation metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). These metrics provide insights into the performance of the forecasting models and allow for comparisons between different approaches. The literature provides examples of studies that use these metrics to validate the effectiveness of the SARIMA model in website traffic forecasting.

Problem Statement

Accurate website traffic forecasting is critical for businesses and organisations to efficiently plan their resource allocation, marketing tactics, and infrastructure requirements. Traditional forecasting methods, on the other hand, frequently fail to capture the complex dynamics of website traffic patterns, resulting in inaccurate predictions. Furthermore, these methodologies fail to effectively account for the impact of external factors such as marketing campaigns, social media activities, vacations, and other important variables that might have a substantial impact on website traffic.

As a result, the issue addressed in this research is the lack of a comprehensive forecasting approach that uses both time series analysis and external variables to effectively estimate website traffic. Existing methodologies do not adequately capture underlying seasonality, patterns, and the influence of external factors, resulting in inefficient resource allocation and wasted marketing opportunities.

We propose using the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMA) model for website traffic predictions to address this issue. By combining time series analysis techniques with external variables, the SARIMA model has the capacity to capture the complex dynamics of website traffic patterns. The SARIMA model can produce more accurate and trustworthy forecasts by adding elements such as marketing efforts, social media activities, holidays, and other important data, allowing businesses to make informed decisions and optimise their online strategies.

Methodology

1. Data Collection:

The first step in the methodology is to collect historical website traffic data. This data should include relevant information such as the timestamp (date and time) and the corresponding number of website visits or page views. Additionally, gather any available data on external variables that may influence website traffic, such as marketing campaign data, social media metrics, holidays, or other factors deemed relevant. Ensure that the data covers a sufficient time span to capture seasonal patterns and trends. The dataset I'm utilising for Website Traffic Forecasting is derived from *thecleverprogrammer.com's* daily traffic data. It includes daily traffic data from June 2021 to June 2022. The dataset is available for download here. Let's get started with the process of forecasting website traffic by importing the relevant Python modules and dataset:

2. Data Preprocessing:

Before applying the SARIMA model, it is essential to preprocess the collected data. This involves handling missing values, outliers, and any inconsistencies in the data. Additionally, perform exploratory data analysis (EDA) to identify any seasonality, trends, or other patterns in the website traffic data. If necessary, apply techniques such as differencing or seasonal differencing to remove any detected patterns.

3. Exogenous Variable Selection:

Identify the relevant exogenous variables that may impact website traffic. This can include marketing campaigns, social media metrics, holidays, or any other factors known to influence website traffic. Conduct a thorough analysis to determine which variables to include in the SARIMA model. Consider factors such as their availability, reliability, and significance in relation to website traffic patterns.

4. Model Training and Parameter Estimation:

Split the preprocessed data into training and validation sets. The training set is used to train the SARIMA model. Apply the SARIMA algorithm to the training data, specifying the appropriate seasonal and non-seasonal orders (p, d, q, P, D, Q, s) based on the identified patterns and the best fit for the data. Estimate the parameters of the SARIMA model using statistical techniques such as maximum likelihood estimation.

5. Model Validation:

Validate the SARIMA model using the validation dataset. Generate forecasts based on the trained model and compare them with the actual website traffic values in the validation set.

Evaluate the accuracy of the forecasts using evaluation metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), or any other appropriate metric. Assess the performance of the SARIMA model and make any necessary adjustments or refinements if the forecasts are not satisfactory.

6. Forecasting:

   Once the SARIMA model is validated, use it to forecast future website traffic. Incorporate the exogenous variables for which data is available in the forecast period. Generate forecasts for the desired time horizon, considering the temporal dependencies, seasonal patterns, trends, and the impact of the selected exogenous variables.

7. Performance Evaluation:

   Evaluate the performance of the SARIMA model by comparing the forecasted website traffic with the actual values. Measure the accuracy of the forecasts using appropriate evaluation metrics. Assess the model's ability to capture seasonality, trends, and the influence of exogenous variables. Analyze any discrepancies and identify areas for improvement or further refinement of the forecasting methodology.

8. Reporting and Interpretation:

   Present the findings of the SARIMA model in a clear and concise manner. Provide visual representations of the forecasted website traffic alongside the actual values. Discuss the accuracy and reliability of the forecasts, highlighting any significant observations or patterns identified. Interpret the results and their implications for resource allocation, marketing strategies, and overall website optimization.

Python Libraries

Plotly

Plotly is a powerful and versatile data visualization library that allows you to create interactive and visually appealing plots. It supports a wide range of chart types, including line plots, scatter plots, bar charts, pie charts, and more. Plotly provides APIs for several programming languages, including Python, R, and JavaScript.

Statsmodels

Statsmodels is a powerful Python library that provides a comprehensive suite of statistical models and tools for data analysis. It offers a wide range of statistical techniques, including regression analysis, time series analysis, hypothesis testing, and more. Statsmodels is built on top of NumPy, SciPy, and pandas, making it an integral part of the scientific Python ecosystem.

Key features and functionalities of Statsmodels:

- Regression Analysis: Statsmodels supports various regression models, such as linear regression, generalized linear models (GLM), robust regression, mixed-effects models, and more. It provides extensive tools for model estimation, hypothesis testing, and result summary.
- Time Series Analysis: Statsmodels offers a comprehensive set of tools for time series analysis, including autoregressive integrated moving average (ARIMA) models, vector autoregressive (VAR) models, state space models, and seasonal decomposition of time series (STL). It allows for data preprocessing, model estimation, forecasting, and diagnostic checks.
- Hypothesis Testing: Statsmodels includes a wide range of statistical tests, such as t-tests, ANOVA, chi-square tests, and more. These tests enable hypothesis testing and help in drawing conclusions about data based on statistical significance.
- Descriptive Statistics: Statsmodels provides descriptive statistical measures, including mean, median, variance, standard deviation, and percentiles. It also offers tools for exploring and summarizing datasets, such as data visualization and exploratory data analysis (EDA).

Data Preprocessing

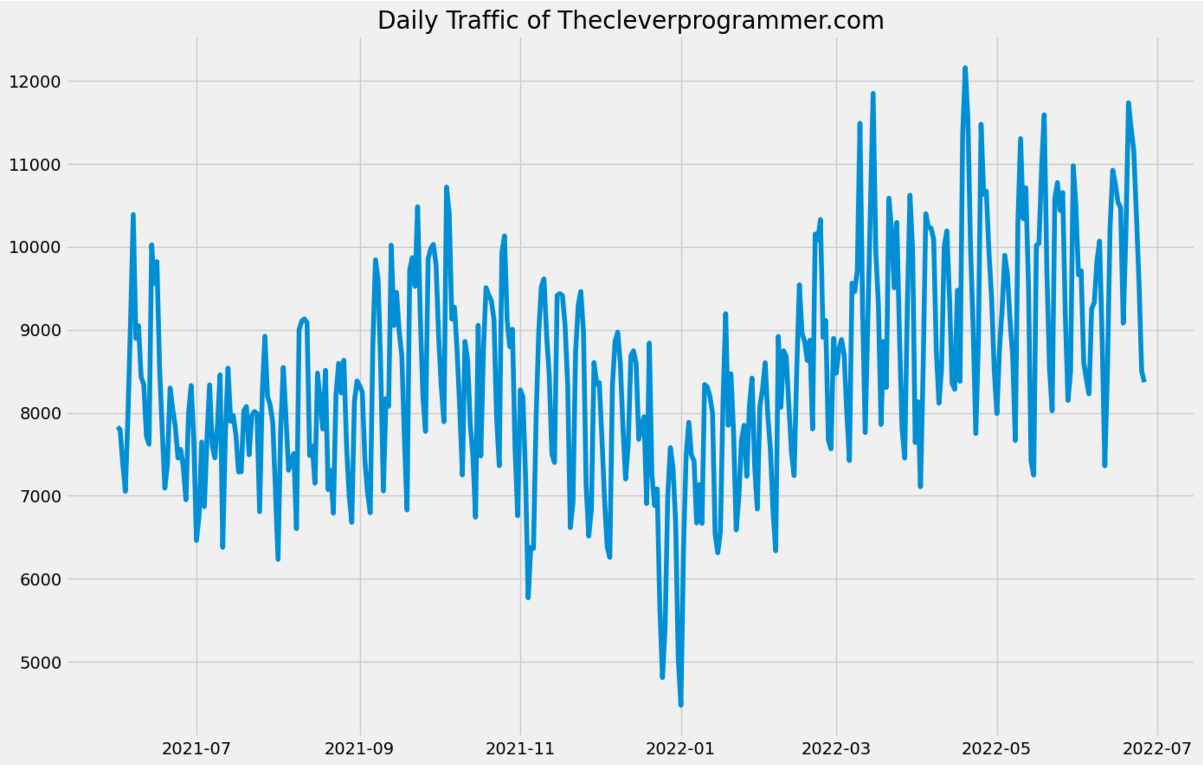
The dataset contains two columns, date and traffic. Before moving forward, I will convert the Date column into Datetime data type:

```
data["Date"] = pd.to_datetime(data["Date"],

                              format="%d/%m/%Y")

print(data.info())

data["Date"] = pd.to_datetime(data["Date"],

                              format="%d/%m/%Y")

print(data.info())
```

```
         Date  Views
0  01/06/2021   7831
1  02/06/2021   7798
2  03/06/2021   7401
3  04/06/2021   7054
4  05/06/2021   7973
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 391 entries, 0 to 390
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Date    391 non-null    datetime64[ns]
 1   Views   391 non-null    int64
dtypes: datetime64[ns](1), int64(1)
memory usage: 6.2 KB
None
```

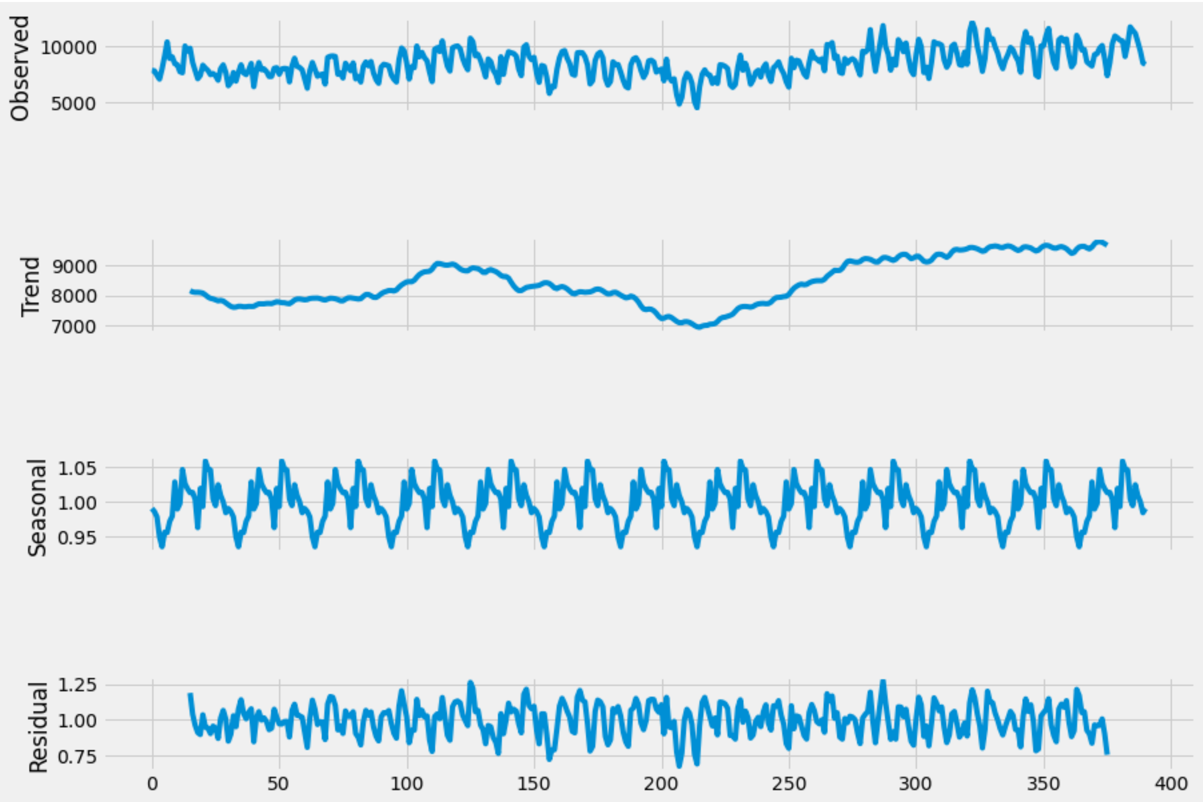
Daily traffic of the website




```
plt.style.use('fivethirtyeight')
plt.figure(figsize=(15, 10))
plt.plot(data["Date"], data["Views"])
plt.title("Daily Traffic of Thecleverprogrammer.com")
plt.show()
```

The code is using the matplotlib library to plot a line chart with a specific style (fivethirtyeight). It sets the figure size, plots the "Views" data against the "Date" data, and adds a title to the plot.Make sure to replace data["Date"] and data["Views"] with the appropriate variables or data sources containing the date and views data for "Thecleverprogrammer.com".

Our website traffic data is seasonal, as traffic peaks during the week and falls over the weekends. When working on the Time Series Forecasting problem, it is useful to know whether the dataset is seasonal or not.



ARIMA

ARIMA is an acronym that stands for Autoregressive Integrated Moving Average. It is a Time Series Data Forecasting Algorithm. ARIMA models, such ARIMA(p, d, q), have three parameters. Here, p, d, and q are defined as follows:

- p:The number p represents the number of lagged values that must be added or subtracted from the values (label column). It represents the autoregressive component of ARIMA.
- d is the number of times the data must differentiate in order to produce a stationary signal. If the data is steady, the value of d should be 0, and if the data is seasonal, the value of d should be 1. ARIMA's integrated portion is captured by d.
- q: The amount of lagged values for the error term added or subtracted from the values (label column) is given by q. It represents the moving average component of ARIMA

SARIMA

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is an extension of the ARIMA model that incorporates seasonality into the time series analysis. SARIMA is particularly useful for forecasting time series data with significant seasonal patterns.

The parameters of the SARIMA model are denoted by (p, d, q) × (P, D, Q, s), where:

- (p, d, q) represents the non-seasonal orders for the AR, I, and MA components, respectively.
- (P, D, Q, s) represents the seasonal orders for the seasonal AR, seasonal I, seasonal MA components, and the seasonality period (s), respectively.

To use the SARIMA model, you can follow these steps:

- Visualize the time series data and identify any obvious trends, seasonality, or other patterns.
- Determine the appropriate values for the SARIMA model parameters (p, d, q) × (P, D, Q, s). This can be done through methods such as autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, as well as trial and error.
- Split the data into training and testing sets. Typically, the earlier portion of the data is used for training, and the later portion is used for testing and evaluating the model's performance.
- Fit the SARIMA model to the training data using the determined parameter values. This involves estimating the model coefficients and optimizing the model's performance.
- Validate the model by comparing the model's forecasts with the actual values from the testing set. Use evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or others to assess the accuracy of the forecasts.
- Optionally, refine the model by iteratively adjusting the model parameters and re-fitting it to the data until a satisfactory level of accuracy is achieved.
- Once the model is validated and deemed satisfactory, use it to forecast future values of the time series.

SARIMAX

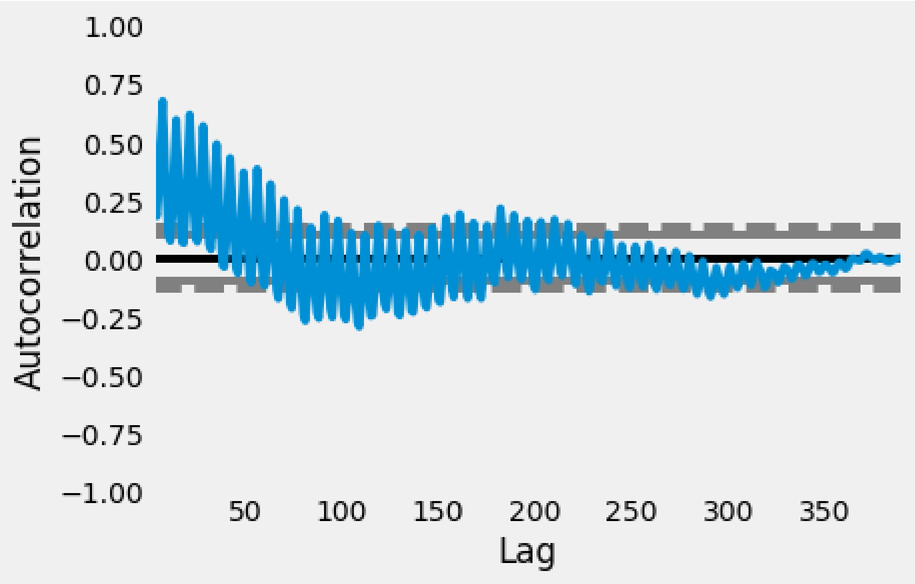
SARIMAX model capable of processing the exogenous features of the time series. The exogenous attribute calculated at time t impacts the not auto-regressive time seriesvalue at time t. We can change the SARIMA equation 3 above to make it an equivalent SARIMAX equation as follows

$$\Phi(L)^p \Phi(L^S)^{\bar{P}} \Delta^d y_t \Delta_S^D y_t = \phi(L)^q \phi(L^S)^{\bar{Q}} \Delta^d \epsilon_t \Delta_S^D \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i \tag{5}$$

is the exogenous attribute at time t and n is the

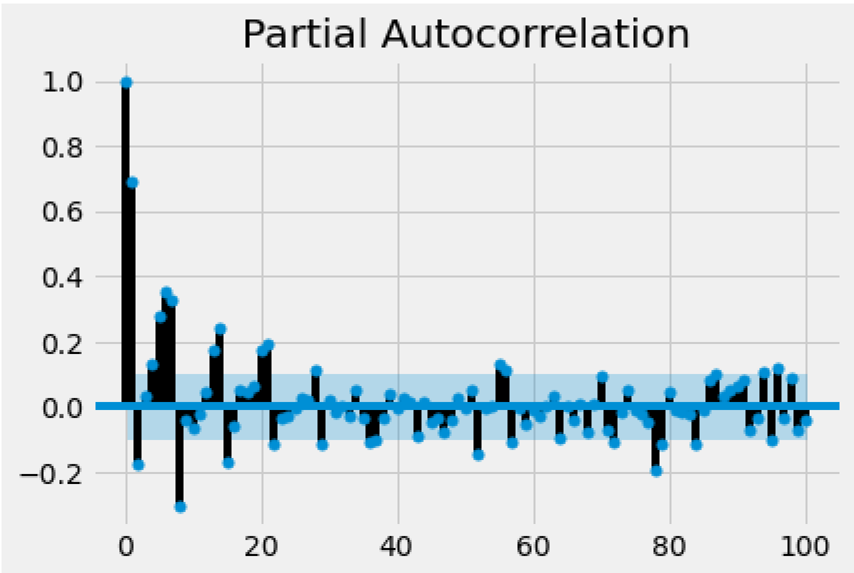
Statsmodels is a Python library that provides an implementation of the SARIMA model. It offers the SARIMAX class, which allows you to fit and forecast time series data using the SARIMA model.

As the data is not stationary, the value of d is 1. To find the values of p and q, we can use the autocorrelation and partial autocorrelation plots:



Ploting the view

```
plot_pacf(data["Views"], lags = 100)
```



Train the SARIMA model

```
p, d, q = 5, 1, 2
```

```
model=sm.tsa.statespace.SARIMAX(data['Views'],
                                order=(p, d, q),
                                seasonal_order=(p, d, q, 12))
model=model.fit()
print(model.summary())
```

The code provided fits a SARIMAX model to the "Views" data with the specified order and seasonal order. It then prints the summary of the fitted mode

The *sm.tsa.statespace.SARIMAX* class is a part of the statsmodels library in Python. It represents the SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) model and provides methods for model fitting, forecasting, and analysis.

- (p, d, q) represents the non-seasonal orders for the AR, I, and MA components, respectively.
- (P, D, Q, s) represents the seasonal orders for the seasonal AR, seasonal I, seasonal MA components, and the seasonality period (s), respectively.

data['Views'] represents the endogenous variable, which is the time series data for the "Views" variable.

(p, d, q, 12) represents the seasonal orders for the seasonal AR, seasonal I, seasonal MA components, and the seasonality period (12 months in this case).

The *fit()* method is called on the *SARIMAX* model to estimate the model parameters and fit the model to the data. The *model_fit* object contains the fitted model.

Finally, the *summary()* method is called on the *model_fit* object to print a summary of the fitted model, including information such as coefficients, standard errors, p-values, and various statistical measures.

Make sure to replace *data['Views']* with the appropriate variable or data source containing the time series data you want to model and forecast.

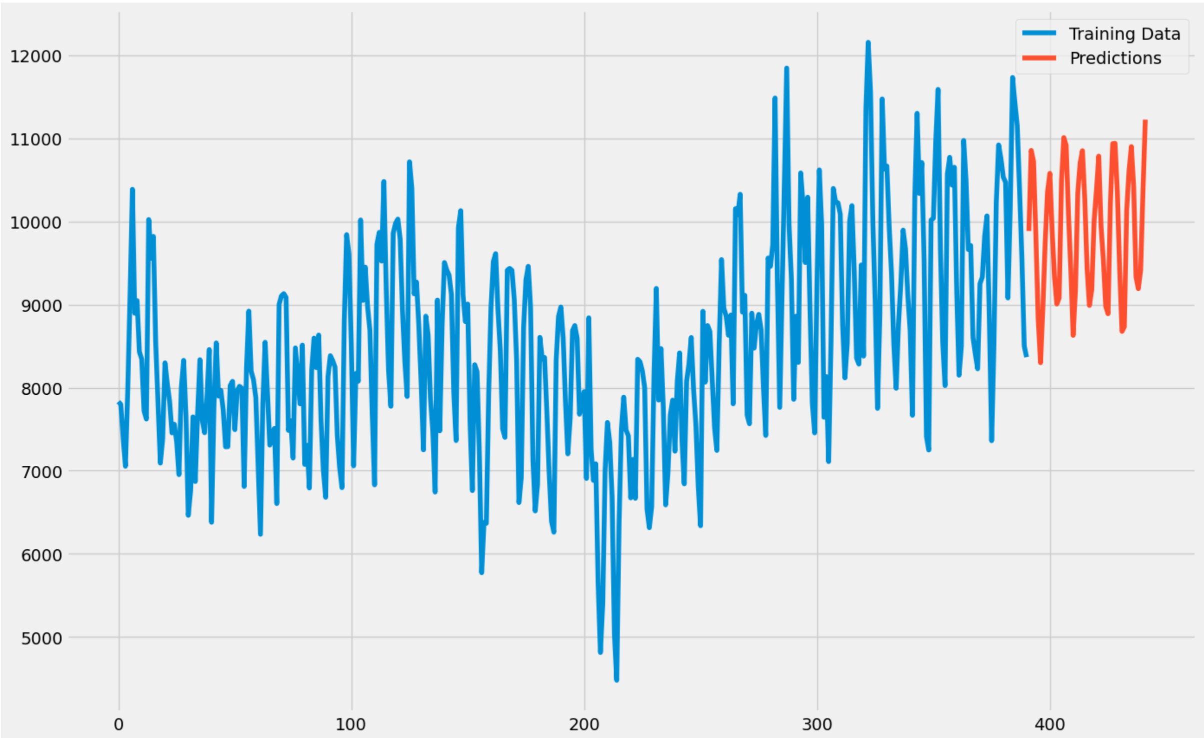
```
                                     Statespace Model Results
==========================================================================================
Dep. Variable:                                Views   No. Observations:                391
Model:             SARIMAX(5, 1, 2)x(5, 1, 2, 12)   Log Likelihood               -3099.402
Date:                              Tue, 28 Jun 2022   AIC                           6228.803
Time:                                      07:01:10   BIC                           6287.827
Sample:                                           0   HQIC                          6252.229
                                              - 391
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.7808      0.134      5.836      0.000       0.519       1.043
ar.L2         -0.7973      0.135     -5.920      0.000      -1.061      -0.533
ar.L3         -0.1442      0.170     -0.850      0.395      -0.477       0.188
ar.L4         -0.1833      0.151     -1.210      0.226      -0.480       0.114
ar.L5         -0.1548      0.139     -1.117      0.264      -0.426       0.117
ma.L1         -1.1826      0.094    -12.515      0.000      -1.368      -0.997
ma.L2          0.8856      0.078     11.304      0.000       0.732       1.039
ar.S.L12      -0.2606      4.608     -0.057      0.955      -9.293       8.772
ar.S.L24       0.0428      0.781      0.055      0.956      -1.488       1.573
ar.S.L36      -0.1880      0.246     -0.764      0.445      -0.670       0.294
ar.S.L48      -0.2151      0.959     -0.224      0.823      -2.095       1.664
ar.S.L60       0.0127      0.986      0.013      0.990      -1.920       1.946
ma.S.L12      -0.6902      4.611     -0.150      0.881      -9.728       8.348
ma.S.L24      -0.0994      3.637     -0.027      0.978      -7.228       7.029
sigma2      1.257e+06   1.59e+05      7.914      0.000    9.46e+05    1.57e+06
===================================================================================
Ljung-Box (Q):                      102.98   Jarque-Bera (JB):                 1.32
Prob(Q):                              0.00   Prob(JB):                         0.52
Heteroskedasticity (H):               1.03   Skew:                             0.14
Prob(H) (two-sided):                  0.85   Kurtosis:                         3.01
===================================================================================
```

Forecast traffic on the website for the next 50 days:

```
391     9885.647737
392    10855.361597
393    10725.051365
394     9828.059279
395     8824.289414
396     8300.183783
397     8949.513213
398     9745.277758
399    10353.338143
400    10578.741349
401     9883.800360
402     9329.632515
403     9005.638396
404     9078.941894
405    10487.891086
406    11009.439207
407    10916.618529
408    10089.760403
409     9422.263831
410     8629.263400
411     9164.667004
412    10344.709609
413    10703.173078
414    10851.358088
415    10275.869403
416     9425.038450
417     8988.807212
418     9178.557035
```

```
419     10003.114691
420     10369.257029
421     10786.438347
422      9937.805230
423      9529.679920
424      8978.573003
425      8889.000917
426     10212.106718
427     10936.393657
428     10940.448619
429     10350.456919
430      9403.312126
431      8676.332354
432      8732.720054
433     10130.564929
434     10605.927773
435     10901.099383
436     10423.594907
437      9332.921721
438      9190.436362
439      9413.960861
440     10364.213567
441     11225.501413
Name: predicted_mean, dtype: float64
```

Plot the predictions



Application and Use cases

1. Site owners and administrators can use SARIMAX-based traffic forecasting to estimate future demand and arrange server capacity accordingly. They can effectively deploy resources to ensure optimal website performance and user experience by understanding expected traffic patterns.
2. Ad Campaign Optimisation: Forecasting website traffic can help optimise ad campaigns for websites that rely on advertising revenue. Website owners can deliberately arrange and target their ad campaigns to maximise reach and engagement with the audience by predicting high traffic periods.
3. Content Strategy: Traffic predictions using SARIMAX can help website owners plan their content strategy. They can anticipate popular subjects or themes by identifying trends and patterns in website traffic and tailoring their content creation efforts accordingly. This guarantees that users have access to relevant and entertaining content during peak traffic periods.
4. Understanding future website traffic facilitates optimal resource allocation across many departments. Customer support teams, for example, can plan for increased traffic and modify their personnel levels and response techniques accordingly. It promotes seamless operations and increases client satisfaction.
5. E-commerce Inventory Management: E-commerce websites can utilize SARIMAX-based traffic forecasting to optimize their inventory management processes. By forecasting future traffic and correlating it with product sales, they can adjust their stock levels, monitor supply chains, and avoid stockouts or excess inventory situations.
6. Business Planning and Decision Making: Accurate traffic forecasting using SARIMAX empowers website owners and stakeholders to make informed business decisions. It provides insights into future performance, revenue projections, and resource requirements. This information supports strategic planning, budgeting, and overall business growth strategies.
7. Performance Monitoring and Anomaly Detection: SARIMAX-based forecasting models can help monitor website performance by comparing actual traffic with predicted values. Significant deviations from the forecasted traffic patterns can indicate anomalies, such as sudden spikes or drops in traffic, which may require further investigation or immediate action.
8. SEO Optimization: SARIMAX-based traffic forecasting can aid in Search Engine Optimization (SEO) efforts. By analyzing historical traffic patterns and predicting future trends, website owners can identify opportunities for optimizing keywords, improving content visibility, and enhancing organic search rankings.

Conclusion

In conclusion, Website Traffic Forecasting Using SARIMAX to estimate and plan for future website traffic proved to be a valuable tool for website owners, administrators, and businesses. Organisations may obtain insights into traffic patterns, forecast peak periods, and make informed decisions to optimise their website operations and improve user experience by using the power of SARIMAX modelling.

The use of SARIMAX in website traffic predictions has various advantages. It provides capacity planning, ensuring that server resources and infrastructure are deployed correctly to accommodate expected traffic loads. This prevents performance difficulties and ensures that website visitors have a pleasant surfing experience. Furthermore, SARIMAX aids in the optimisation of advertising campaigns by recognising high-traffic periods, allowing website owners to carefully schedule and target ads to maximise engagement and income.

Furthermore, SARIMAX forecasting makes content planning easier, allowing website owners to connect their content strategy with expected visitor patterns. This guarantees that relevant and engaging information is available during peak moments, increasing user satisfaction and participation. By anticipating traffic flows and permitting proper staffing levels for optimal performance, the model also facilitates resource allocation across multiple departments, such as customer assistance.

SARIMAX-based website traffic forecasting has implications beyond operations. It aids in the management of e-commerce inventory, allowing businesses to optimise stock levels, monitor supply chains, and minimise stockouts or surplus inventory issues. Furthermore, SARIMAX-based predictions help with corporate planning and decision-making by offering insights into future performance, revenue projections, and resource needs.

In conclusion, Website Traffic Forecasting Using SARIMAX empowers organizations to make data-driven decisions, optimize resource allocation, enhance user experience, and drive business growth. By harnessing the power of this forecasting technique, website owners can anticipate future traffic patterns, adapt their strategies, and position themselves for success in the dynamic online landscape.

Future Scope

Time Series Forecasting is one of the least researched fields, and numerous methods are being studied to increase forecast accuracy. The proposal's major goal is to forecast.Future online traffic will be used to make judgements for improved congestion control. Past values are used to forecast future values. We will also investigate multivariate time series and make recommendations for simplifying real-time decision-making. Time Series Forecasting is one of the least researched fields, and numerous methods are being studied to increase forecast accuracy. The proposal's major goal is to forecast future web traffic so that better congestion control decisions may be made. Past values are used to forecast future values. We shall also investigate multivariate time series.

Model Enhancement and Refinement: Enhancement and refinement of SARIMAX models for website traffic predictions is one area of future study and development. This includes experimenting with other SARIMAX models, such as SARIMA with exogenous variables, SARIMA with dynamic regression, and SARIMA with additional seasonal components. These modifications have the potential to increase the accuracy and resilience of traffic forecasts. Hybrid Models and Ensemble approaches: The creation of hybrid models that integrate SARIMAX with other forecasting approaches is another path for future research. To achieve more precise and dependable forecasts, hybrid models can combine the strengths of various models. Furthermore, ensemble approaches can be used to improve forecasting performance and minimise uncertainty by aggregating forecasts from multiple models.

Integrating SARIMAX-based traffic forecasting with analytics and visualization tools can enhance the usability and interpretability of the forecasts. Developing user-friendly interfaces and dashboards that allow website owners and administrators to vis Conducting comprehensive benchmarking and evaluation studies comparing SARIMAX with other forecasting methods can contribute to a better understanding of its strengths, limitations, and performance in different scenarios. This can help identify situations where SARIMAX excels and where alternative approaches may be more appropriate. visualize and explore the forecasted traffic patterns can facilitate better decision-making and planning.

SARIMAX-based traffic forecasting techniques developed for websites can potentially be applied to other domains with similar time series characteristics. Exploring the applicability of SARIMAX in fields such as retail, transportation, social media, or energy demand forecasting can expand its scope and impact.

Finally, the future scope of Website Traffic Forecasting Using SARIMAX includes model enhancement, external data integration, hybrid models, real-time forecasting, integration with analytics tools, long-term forecasting, benchmarking, and application in other domains. These developments will contribute to more accurate, robust, and adaptable website traffic forecasting approaches, allowing website owners and administrators to make more informed decisions and plan more effectively.

References

[1] "Predicting Computer Network Traffic: A Time Series Forecasting Approach using DWT, ARIMA and RNN" by Rishabh Madan, 2018.

[2] “Fast ES-RNN: A GPU Implementation of the ES-RNN algorithm ” by Andrew Redd and Kaung Khin, 2019.

[3] “Time Series Forecasting Based on Complex Network Analysis“ by SHENGZHONG MAO AND FUYUAN XIAO, 2019.

[4] “Web Traffic Prediction of Wikipedia Pages” by Navyasree Petluri, Eyhab Al-Masri, 2019.

[5] “Time series forecasting using improved ARIMA” by Soheila Mehrmolaei,2016.

[6] “Efficient Prediction ofy Network Traffic for Real-Time Applications” by Muhammad Faisal Iqbal , Muhammad Zahid, Durdana Habib, and Lizy Kurian John, 2019.

[7] https://wikitech.wikimedia.org/wiki/Analytics/AQS/Pageviews

[8] https://towardsdatascience.com/3-facts-about-timeseries-forecasting-that-surprise-experienced-machinelearning-practitioners-69c18ee89387.

[9] "Temporal Pattern Attention for Multivariate Time Series Forecasting" by Shun-Yao Shih Fan-Keng Sun Hung-yi Lee, 2018.

[10]“Time series forecasting using improved ARIMA” by Soheila Mehrmolaei,2016.

[11] “Efficient Prediction of Network Traffic for Real-Time Applications” by Muhammad Faisal Iqbal , Muhammad Zahid, Durdana Habib, and Lizy Kurian John, 2019.

[12] “Modelling Approaches for Time Series Forecasting and Anomaly Detection” by Shuyang Du , Madhulima Pandey, and Cuiqun Xing, 2018.

[13] “Neural Decomposition of Time-Series Data for Effective Generalization” by Luke B. Godfrey and Michael S. Gashler, 2017

[14] Tensorflow, https://www.tensorflow.org.

[15] Paxson V, Floyd S. Wide area traffic: the failure of Poisson modeling. IEEE/ACM

Transactions on Networking 1995;3(3):226–244