INTERNSHIP PROJECT

Website Traffic Forecasting

By Saad Mohiuddin Ajaz

www.firstquadrantlabs.com

PROJECT OVERVIEW

OBJECTIVE:

Develop a time series machine learning forecasting model to accurately predict daily website traffic.

KEY FOCUS AREAS:

- Improve forecasting accuracy.
- Capture complex traffic patterns.
- Provide insights for resource optimization and decision-making.

- Tools Used: Python,Excel, Time Series Analysis.

PROJECT BENEFITS

Benefits of Website Traffic Forecasting

- **Improved Forecasting:** More accurate predictions of daily traffic.
- **Resource Optimization:** Enhanced server resource allocation and content planning.
- **Informed Decision-Making:** Scheduling content releases and marketing campaigns based on traffic insights.

DATA OVERVIEW

Data Description

- **Dataset:** Daily website traffic data for 5-6 years (~2100 records).
- **Key Variables:** Page loads, Unique visits.
- **Time Span:** 2014 - 2020.
- **Challenges:** Seasonality and spikes in traffic, missing data points.

DATA PREPROCESSING

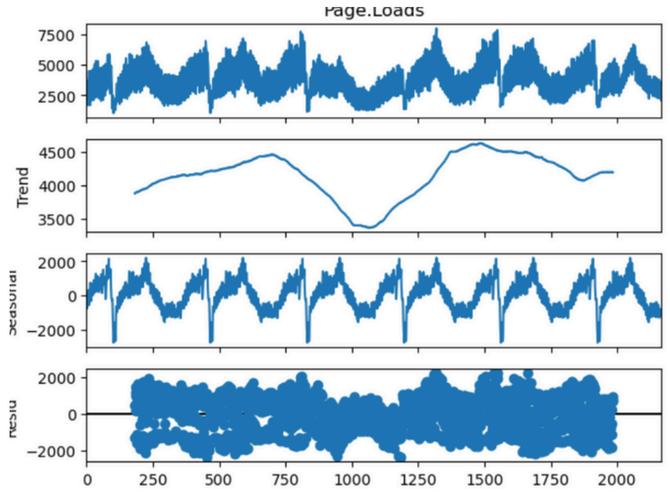
Data Preprocessing Steps

- Date Formatting: Ensured proper date-time format.
- Stationarity Check: Performed ADF test.
- Seasonality: Detected yearly seasonality based on traffic patterns.

```
### Testing For Stationarity

from statsmodels.tsa.stattools import adfuller

result= adfuller(y_train)

print('ADF Statistic: %f' % result[0])
print('p-value: %f' %result[1])
print('Critical Values:')

for key,value in result[4].items():
    print('\t%s: %.3f' %(key, value))
```

```
ADF Statistic: -3.630057
p-value: 0.005212
Critical Values:
        1%: -3.558
        5%: -2.917
        10%: -2.596
```

clearly it shows it is stationary as vale of p is less than 0.05



METHODOLOGY

MODEL DEVELOPMENT

- Time Series Models: Implemented SARIMA (seasonal ARIMA) for capturing seasonality and trend.

MODEL HYPERPARAMETER TUNING

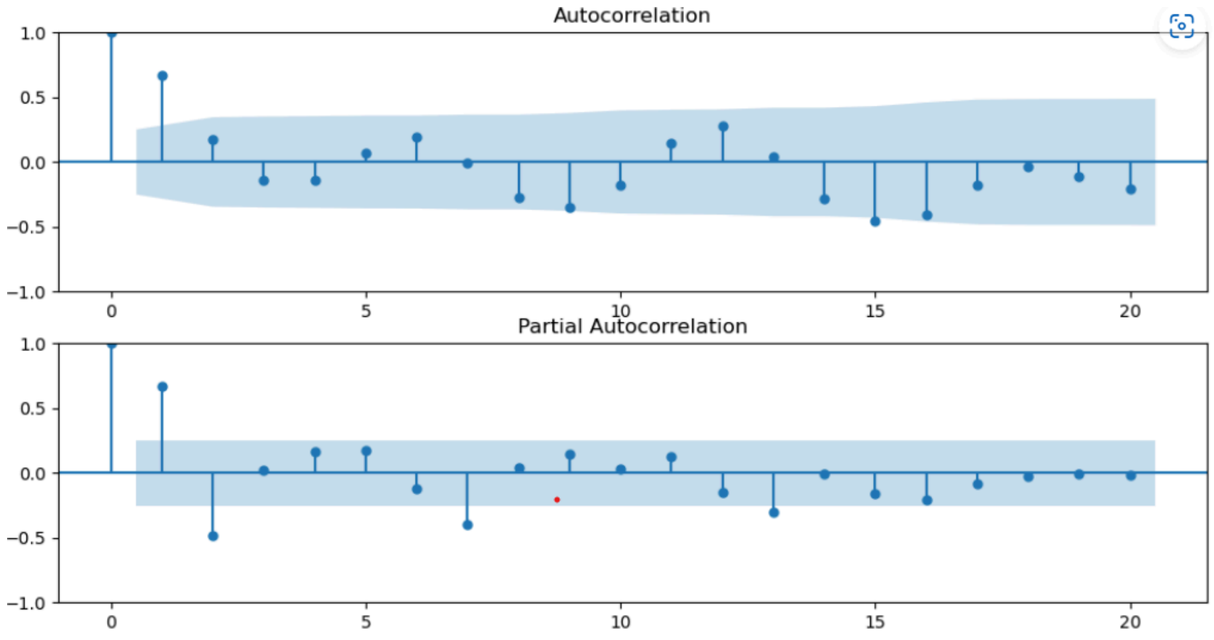
- SARIMA Parameters: Tuned seasonal (P, D, Q) and non-seasonal (p, d, q) components.
- Challenges: Ensuring convergence in SARIMA and avoiding overfitting in machine learning models.

MODEL EVALUATION

- Model Evaluation Metrics:
- Root Mean Squared Error (RMSE): Measures prediction error.
- ACF and PACF plots to evaluate or understand p,d,q values effectively.

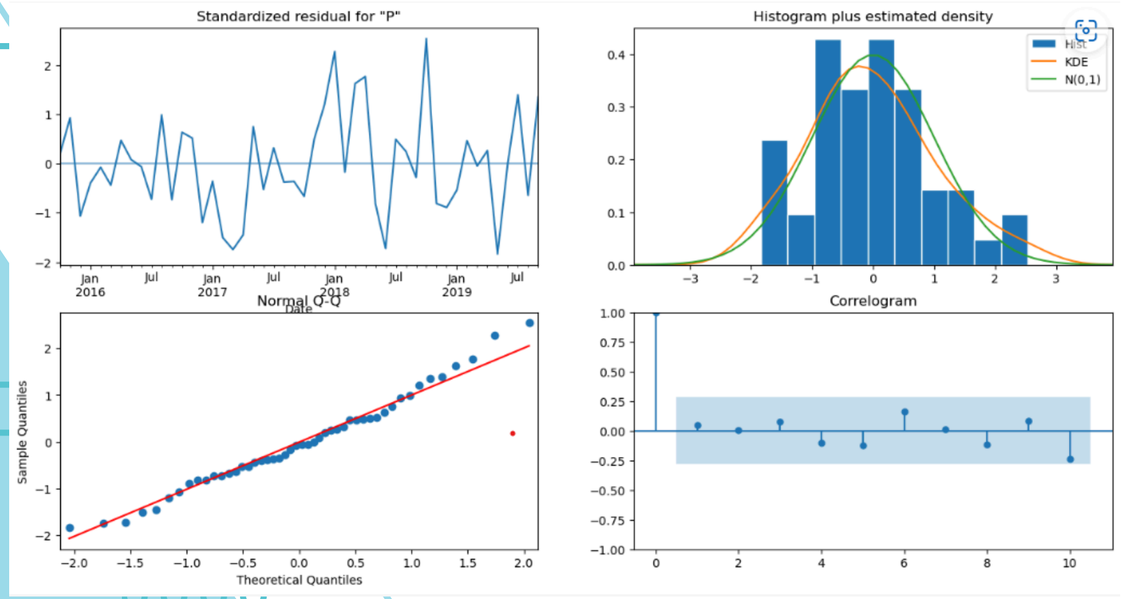
```
forecast=model_aic.get_prediction(start=pd.to_datetime('2020-01-01'), dynamic=False)
predictions=forecast.predicted_mean

actual = y_test['2020-01-01':]

rmse= np.sqrt((predictions - actual)**2).mean()
print('The Root mean squared error of our forecasts is {}'.format(round(rmse,2)))
```

```
The Root mean squared error of our forecasts is 1240.78
```

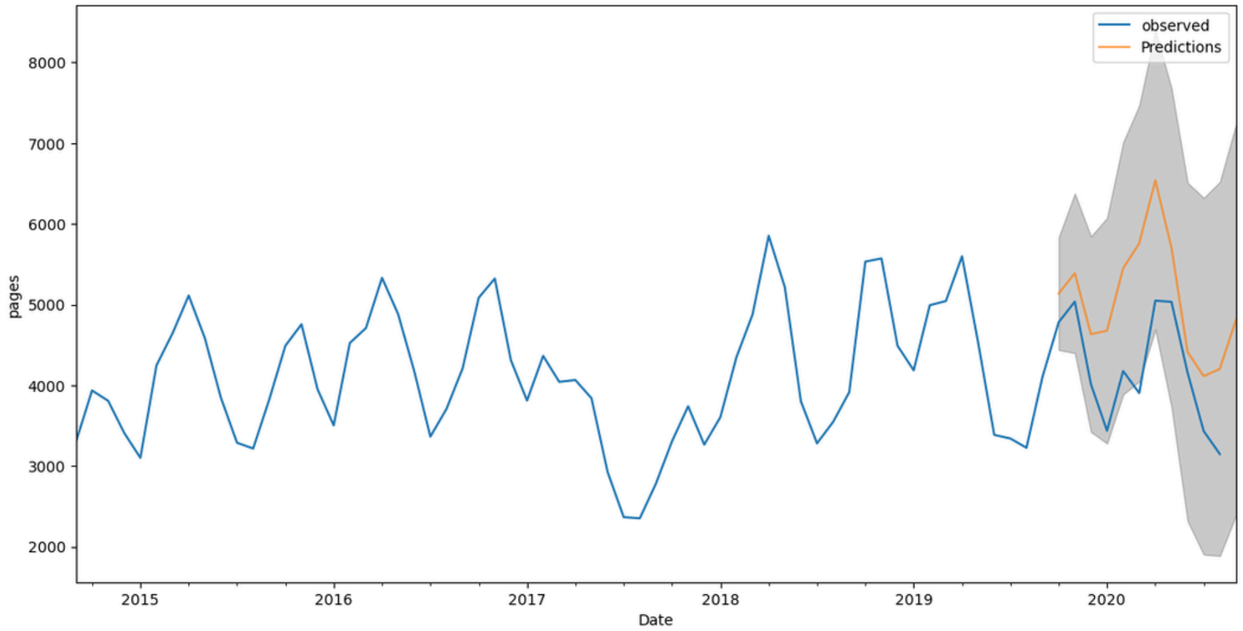


RESULTS AND FORECASTS

- **Insights from Forecast: Highlight key insights such as peak traffic seasons, trend shifts, and anomaly detection.**



Visualizing Forecast: Line plots of actual vs predicted values



NEXT STEPS

- Refinements: Further improve model accuracy with more granular data.
- Implementation: Integrate model into live traffic monitoring system.
- Future Work: Extend forecasting to include user behavior patterns, bounce rates, and conversions.

THANK YOU

By Saad Mohiuddin Ajaz

Questions?

Email:saadajaz20@gmail.com