DS250

Data Analytics and Visualization

Assignment 3(a)

Animesh Raj : 11940120

Puja Bansal : 11940910

Colab Link : → Link

Introduction

In this assignment, we were given the gold prices dataset, with gold prices in many different currencies. For which we have Imputed the missing values from the data-set, to visualize and predict the gold prices for the next 20 days via different models (ARIMA, SARIMA, Linear Regressor, Facebook-prophet).

Part A

Pre-processing :

- Forward-fill Backward-fill → Gaps in data can be filled by propagating the non-NaN values forward or backward along a Series of attributes dof data sets.
- Simple Imputer → The SimpleImputer class provides basic strategies for imputing missing values. Missing values can be imputed with a provided constant value, or using the statistics (mean, median or most frequent) of each column in which the missing values are located.
- Extension of Simple Imputer → Make new columns indicating what will be imputed in the data-set by using SimpleImputer class.

Part B

Visualization :

Interactive visualization using plotly. Where we implemented dropdown (select from country's currency list) and range slider (to zoom-in and zoom-out the particular range of the graph).

Also tried to visualize via Altair (due to some restrictions of Altair to process data of 5000 rows only, we didn't proceed further with Altair).

Part C

Model :

- **ARIMA :** An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is autoregressive if it predicts future values based on past values. Here we created a model with order (p=1,d=1,q=1).
- **SARIMA :** The difference between ARIMA and SARIMA is about the seasonality of the dataset. If data is seasonal, like it happens after a certain period of time. Then we will use SARIMA. Here we created a model of seasonal_window 7 and (p=1,d=1,q=0).
- **Linear Regressor :** Created a lag-features and tested the desired model on that feature for 100 most recent days validation.
- **Facebook - prophet :** This model accounts for 'change points' or specific shifts in trend in the time series. We created 4 forecast's for the given model by changing the parameters to get changepoints and visualizing that particular forecast.

Result : Report the MAE, MAPE and RMSE accuracy for all the models, and where we got the minimum error value, we used that model for 20 days prediction.

Part D

Predictions for next 20 days :

After reporting all the error values (MAE, MAPE, RMSE), one for which we got the minimum error_accuracy value, we implemented that model for prediction.

In our case the model we used for prediction was the ARIMA model, since the value of MAE, MAPE, RMSE is minimum among all models.

THANK YOU!!