CS 457 - Homework Assignment 12: Time Series Analysis

Due Date: Monday, April 10 at 11:59 pm

Purpose:
Demonstrate understating of Time Series Analysis for prediction and forecasting.

Points: 100

Deliverables: Submit ipynb code file along with your answer

- Use the dataset `bitcoin-price.csv`
- Perform time series analysis on the following questions**.** Make sure to include interpretation of each result including visualizations to support your answer

1. Read the dataset visualize the trend, seasonality, residual etc. Discuss your observations in the visualization.
2. Keep the data in days as provided in the dataset. Do not aggregate it for months as well are trying to predict everyday price.
3. Split the data into train and test sets. Use all the data for year 2020 for training and 2021 for testing.
4. Use ARIMA, visualize the results and report the RMSE.
5. Use SARIMA, visualize the results and report the RMSE.
6. Try to come up with optimal parameters for SARIMA. Pick the best one, visualize the results and then report the RMSE. Compare both ARIMA and SARIMA and conclude which one is better?
7. Now use any two Machine Learning regressor techniques and use the same training and testing set (used for ARIMA and SARIMA). Visualize the results and report the RMSE.
8. Compare all the RMSEs (ARIMA, SARIMA and two regression algorithms) and discuss which one is better.
9. Generate future dates from April 1st 2022 to June 30th 2022 and predict the price using SARIMA model (with optimal parameters).
10. Using the same future dates, predict the price using any one ML regression model
11. Visualize the predictions from 9 and 10 and discuss their differences.