ARIMA ALGORITHM

1. Data collection: The first step was to extract the stock price dataset for analysis.We are using the API of Yahoo finance(yfinance) for collection of data.
2. Data preprocessing: Once the dataset is collected, the next step is to preprocess the data. This includes , checking the null values, statistical description (number of sample data, mean, standard deviation, calculations of  various quantiles), and transforming the data into a format that is suitable for modeling. The value to be predicted is the ‘Close’ value of the stock price, we are removing it from the dataset to make a 1-d data frame of close values.
3. Standard Decomposition:- To explain the characteristics of the time series data better, we can break it down into three components:  trend - T(t): a long-term upward /downward change in the average value based on selected period. seasonality - S(t): a periodic change to the value that follows an    identifiable pattern.  residual - R(t): random fluctuations in the time series data that do not follow any patterns.It’s a preliminary step used to check the Stationarity of data visually.
4. Stationarity/Seasonality check: Time series data needs to be a stationary forearm model to work. Thus, we will check for the stationarity of the data using various tests like ADF(Augmented Dickey–Fuller test) and KPSS(Kwiatkowski–Phillips–Schmidt–Shin).
5. Differencing-If the series is not stationary and is seasonal in nature(p-value>0.05) then there is differencing to be done for making the series stationary/non-seasonal.We would continue differencing until the series has become stationary.
6. Test/Train:- we are splitting the data into two parts i.e test and train. On train data, we would be applying the ARMA model to obtain the best-fit line and on test data, we are trying to predict the remaining values of the dataset.
7. Order of the model:-. The order of the model is represented by the parameters p,q and d. The parameter p represents the order of the autoregressive component.The parameter q represents the order of the moving average component and the integration/difference factor is also included in the computation of the final variable. We are using ACF(Auto-Correlation Function) and PACF (Partial Auto-Correlation Function) to find the respective p and q values.The model order with minimum aic (Akaike information criterion ) is selected.
8. Fitting: Once the appropriate model is selected.. We will use the "fit" method to fit the model to the data.
9. Testing the residuals- Residuals are tested for their normality to see if the residual values are close to a normal distribution  and the autocorrelation between the residual values.
10. Model evaluationWe will use various metrics like Mean Square Error (MSE), R2 score to evaluate the performance of the model. We will also use visualization techniques like line plots, autocorrelation plots, partial autocorrelation plots, etc.