Project title: Future sales prediction

ARIMA PREDICTION MODEL

Autoregressive integrated moving average (ARIMA) models predict future values based on past values. ARIMA makes use of lagged moving averages to smooth time series data it captures the patterns , trends and seasonality of the data using a combination of past values, differences and errors.

An ARIMA model is a mathematical representation of a time series that describes how the current value depends on the past values and the random errors. The model has three main components: auto regression, integration, and moving average. Auto regression means that the current value is a linear combination of the previous values, with some coefficients. Integration means that the model is applied to the differences between consecutive values, rather than the original values, to make the series stationary. Stationary means that the mean, variance, and autocorrelation of the series do not change over time. Moving average means that the current value is also a linear combination of the past errors, with some coefficients. The errors are assumed to be independent and normally distributed.

TIME SERIES ANALYSIS:

Suppose you have monthly sales data for a retail store, and you want to forecast the next 12 months. We can plot the data and see that it has an upward trend and some seasonality. You can also perform a stationarity test, such as Augmented Dickey-Fuller, and find that the series is not stationary. We can then apply a first-order difference to make it stationary, and plot the autocorrelation and partial autocorrelation functions to identify the potential values of p and q. we can then fit several ARIMA models with different combinations of p, d, and q, and compare them using AIC, BIC, and RMSE. You can also plot the forecasts and the residuals for each model, and check their assumptions. we can then select the model that has the lowest AIC, BIC, and RMSE, and the best forecast and residual plots. In this case, it may be an ARIMA(1,1,1) model with a seasonal component.

An autoregressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

PARAMETERS:

The parameters can be defined as:

* *p*: the number of lag observations in the model, also known as the lag order.
* *d*: the number of times the raw observations are differenced; also known as the degree of differencing.
* q: the size of the moving average window, also known as the order of the moving average.

Process Flow:

DATA COLLECTION

Sales data details of various branches of the company are gathered and consolidated into a single dataset, using the SQL Server with the help of link tables. Which contains 2,78,559 instances and is saved as an Excel file. Data is imported into a data frame for further processing in Python.

DATA PRE-PROCESSING

Steps involved in data pre-processing, (i) Dropping unwanted columns. (ii) Checking for null values and dropping them. (iii) Converting the data type and sorting the data based on date. (iv) Grouping and summarizing the values based on the date. As a result, the model can be built with clean data.

EXPLORATORY DATA ANALYSIS

The data is visualized in order to identify the major points and elements influencing the company's growth. The company’s top and the least sold products and monthly sales are visualized.

ARIMA MODEL

Arima model Predicted 𝑌𝑡 = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

RESULT AND ANALYSIS:

Implementing ARIMA model and model that will predict the sales. 70% of the data will be used for training and 30% will be used for testing in this project. The model will be trained, and it will be used to forecast. Based on the forecast made, a graph will be plotted for future sales.

CONCLUSION:

ARIMA is a popular technique for analysing stationary univariate time series data. Model identification, model estimation, and model checking are the three main stages in building an ARIMA model, with model classification being the most important stage. As a result, the survey provides insight into the various time series prediction and forecasting models using ARIMA. A large number of real-world applications conducted by various individuals were also studied, and it was discovered that ARIMA is a real-world toll for time series prediction, forecasting, and analysis with accuracy.