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: autoregression, integration, and moving average. Autoregression 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:

1. Data Preparation

- 1.1 Data Collection

- 1.2 Data Cleaning

- 1.3 Data Exploration

- 1.4 Data Transformation

2. Model Development

- 2.1 Stationarity

- 2.2 Model Identification

- 2.3 Parameter Estimation

- 2.4 Model Checking

3. Model Evaluation

- 3.1 Train-Test Split

- 3.2 Model Training

- 3.3 Model Validation

- 3.4 Performance Metrics

4. Forecasting

- 4.1 Model Deployment

- 4.2 Inventory Optimization

- 4.3 Decision Making

**1. Data Preparation**

**1.1 Data Collection**

Gather historical sales data that includes relevant features such as date, product ID, store ID, and sales quantity. Ensure data consistency, accuracy, and completeness.

**1.2** **Data Cleaning**

Clean the data by handling missing values, duplicates, and outliers. Ensure that the data is in a suitable format for time series analysis.

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.

**1.3 Data Exploration**

Explore the data through visualization and statistical analysis to understand trends, seasonality, and any other patterns that may exist.

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.

**1.4 Data Transformation**

Transform the data into a time series format, making sure it is in chronological order. This may involve aggregating sales by day, week, or another relevant time unit.

**2. Model Development**

**2.1 Stationarity**

Check and ensure that the time series data is stationary, meaning it has a constant mean and variance over time. If not, apply differencing techniques to achieve stationarity.

**2.2 Model Identification**

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.

Identify the appropriate order (p, d, q) for the ARIMA model. This involves selecting the number of autoregressive (p) and moving average (q) terms, as well as the differencing parameter (d).

**2.3 Parameter Estimation**

Estimate the model parameters using methods like Maximum Likelihood Estimation (MLE). This step involves fitting the ARIMA model to the training data.

**2.4 Model Checking**

Check the model's goodness-of-fit using diagnostic plots (e.g., ACF and PACF plots) and statistical tests (e.g., Ljung-Box test) to ensure that the model adequately captures the underlying patterns in the data.

**3. Model Evaluation**

**3.1 Train-Test Split**

Split the data into training and testing sets, typically reserving a portion (e.g., 20%) for testing to evaluate the model's performance.

**3.2 Model Training**

Train the ARIMA model using the training data, utilizing the identified parameters (p, d, q).

**3.3 Model Validation**

Validate the model by forecasting sales on the testing data and comparing the predictions to the actual sales values.

**3.4 Performance Metrics**

Evaluate the model's performance using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Forecast Accuracy.

**4. Forecasting**

**4.1 Model Deployment**

Deploy the trained ARIMA model to forecast future sales. Continuously update the model as new data becomes available.

**4.2 Inventory Optimization**

Use the sales forecasts to optimize inventory management, ensuring that stock levels are sufficient to meet demand without excessive overstocking.

**4.3 Decision Making**

Make informed business decisions based on the data-driven sales predictions. This includes demand planning, pricing strategies, and supply chain optimization.

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.