DOMAIN: Applied Data Science.

PROJECT NAME: product demand prediction with machine learning.

PROBLEM DEFINITION:

- The problem is to create a machine learning model that forecasts product demand based on historical sales data and external factors.
- The goal is to help businesses optimize inventory management and production planning to efficiently meet customer needs.

DESIGN THINKING:

- This project involves 6 steps. They are data collection, data preprocessing, feature engineering, model selection, training, and evaluation.

1.DATA COLLECTION:

- In this part datas are collected to be first from various sources. Then the datas that are used for train the model.
- But in our case, we are already provided by datasets for product demand prediction from KAGGLE platform.

  Dataset link: https://www.kaggle.com/datasets/chakradharmattapalli/product-demandprediction-with-machine-learning

- This dataset contains ***5 columns and 150150 rows***.
- The datas in the columns are based on

  1. ID of the product
  2. ID of the store
  3. Total price of the product(i.e manufacturing price)
  4. Base price of the product (i.e retail price)
  5. No.of.units sold.

2. DATA PREPROCESSING:

In this part include major of 3 things;

1. Clean and preprocess the data:

Here , we are going to clear outliers that perform distinct significant from the other observation

2.Handle the missing values:

- we are working with large amount of data,therefore every is need to be consistent and every datas are need to be verified before the process.
- If we found any entity with missing values, we have to replace a ZN () function replaces a Null value for the data field that is placed inside the brackets with a zero.

3.Convert categorical features into numerical values:

- Here, we are converting the data values into numerical values for example, most of the price values are in INR and some of them are in $, so here we have to convert them ($) into INR.

3. FEATURE ENGINEERING:

In this step we are adding additional feature that capture seasonal patterns/ trends, and external influences on product demand.

1.seasonal patterns/trends: A certain time series with repetitive or predictable patterns of demand due to re-occurring seasonal events. These patterns can re-occur over days, weeks, months or quarters and can make it harder for businesses to forecast future demand trends. E.g., MEESHO Maha indian sale.

2. External influences on products demand: Such as,

- based on demand of the product.
- Competition between same product but on different manufacturers
- Suppliers of raw materials and goods
- economic conditions
- buyers
- Government

4.MODEL SELECTION:

PHASE_2 : INNOVATION

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE(ARIMA)

- In this part we are choosing ARIMA algorithm for demand forecasting.
- ARIMA, which stands for Autoregressive Integrated Moving Average, is a powerful time series forecasting algorithm widely used in product demand prediction.
- It combines the concepts of autoregression (AR) and moving averages (MA) with differencing to make time series data stationary, which is crucial for accurate forecasting.
- ARIMA models are particularly valuable when dealing with historical demand data, as they can capture seasonality, trends, and other patterns, allowing businesses to make informed decisions about inventory management and production planning.
- By understanding the fundamentals of ARIMA, you can harness its predictive capabilities to optimize your product demand forecasts.

KEY TAKEAWAYS OF ARIMA:

- 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.
- They are widely used in technical analysis to forecast future security prices.

Autoregressive models implicitly assume that the future will resemble the past.

- Therefore, they can prove inaccurate under certain market conditions, such as financial crises or periods of rapid technological change.
- 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.

COMPONENTS OF ARIMA:

An ARIMA model can be understood by outlining each of its components as follows:

- **Autoregression (AR):** refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
- **Integrated (I):** represents the differencing of raw observations to allow the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- **Moving average (MA):** incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

ARIMA PARAMETERS:

- 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.
- For example, a linear regression model includes the number and type of terms. A value of zero (0), which can be used as a parameter, would mean that particular component should not be used in the model. This way, the ARIMA model can be constructed to perform the function of an ARMA model, or even simple AR, I, or MA models.

ARIMA AND STATIONARY DATA:

- In an autoregressive integrated moving average model, the data are differenced in order to make it stationary.
- A model that shows stationarity is one that shows there is constancy to the data over time. Most economic and market data show trends, so the purpose of differencing is to remove any trends or seasonal structures.
- Seasonality, or when data show regular and predictable patterns that repeat over a calendar year, could negatively affect the regression model.
- If a trend appears and stationarity is not evident, many of the computations throughout the process cannot be made and produce the intended results.
- A one-time shock will affect subsequent values of an ARIMA model infinitely into the future. Therefore, the legacy of the financial crisis lives on in today’s autoregressive models.

PROS AND CONS OF ARIMA:

ARIMA models have strong points and are good at forecasting based on past circumstances, but there are more reasons to be cautious when using ARIMA. In stark contrast to investing disclaimers that state "past performance is not an indicator of future performance...," ARIMA models assume that past values have some residual effect on current or future values and use data from the past to forecast future events.

PROS:

- Good for short-term forecasting.
- Only needs historical data.
- Model's non-stationary data.

CONS:

- Not built for long-term forecasting.
- Poor at predicting turning points.
- Computationally expensive.
- Parameters are subjective.

DIFFERENCE BETWEEN AUTOREGRESSIVE AND MOVING AVERAGE MODELS:

- ARIMA combines autoregressive features with those of moving averages. An AR (1) autoregressive process, for instance, is one in which the current value is based on the immediately preceding value, while an AR (2) process is one in which the current value is based on the previous two values.
- A moving average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set to smooth out the influence of outliers. As a result of this combination of techniques, ARIMA models can take into account trends, cycles, seasonality, and other non-static types of data when making forecasts.

ARIMA FORECASTING WORK:

- ARIMA forecasting is achieved by plugging in time series data for the variable of interest. Statistical software will identify the appropriate number of lags or amount of differencing to be applied to the data and check for stationarity.
- It will then output the results, which are often interpreted similarly to that of a multiple linear regression.
- The Bottom Line
    - ✓ The ARIMA model is used as a forecasting tool to predict how something will act in the future based on past performance. It is used in technical analysis to predict an asset's future performance.

- ✓ ARIMA modeling is generally inadequate for long-term forecasting, such as more than six months ahead, because it uses past data and parameters that are influenced by human thinking. For this reason, it is best used with other technical analysis tools to get a clearer picture of an asset's performance.

5.MODEL TRAINING:

- To begin building an ARIMA model for an investment, you download as much of the price data as you can.
- Once you've identified the trends for the data, you identify the lowest order of differencing (d) by observing the autocorrelations.
- If the lag-1 autocorrelation is zero or negative, the series is already differenced. You may need to difference the series more if the lag-1 is higher than zero.
- Next, determine the order of regression (p) and order of moving average (q) by comparing autocorrelations and partial autocorrelations. Once you have the information you need, you can choose the model you'll use.

6.MODEL EVALUATION:

To forecast with ARIMA in Python, follow these steps using the statsmodels library:

- ✓ Import necessary packages.
- ✓ Prepare time series data.
- ✓ Determine the order of differencing (d), AR (p), and MA (q) terms using ACF and PACF plots.
- ✓ Fit ARIMA model using ARIMA().fit().
- ✓ Generate forecasts using forecast() method.
- ✓ Visualize predictions and evaluate model performance

CONCLUSION:

- These are all the steps that are involved in product demand prediction forecasting with machine learning.
- These are the things that we are understood in the project. By this way we are going to done our project.