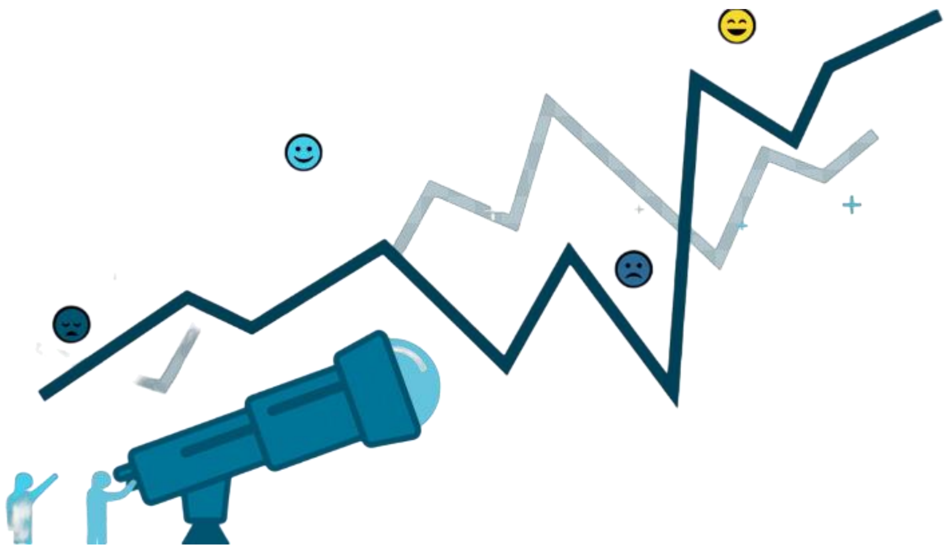
Product Demand Prediction with Machine Learnings

PHASE - 2

EXPLANATION:

Product demand prediction with machine learning involves using historical sales data, customer behaviour, and external factors to forecast a product's future demand. Through data analysis and advanced algorithms, machine learning models identify patterns and correlations to make accurate predictions. This process aids businesses in optimizing inventory management, pricing strategies, and overall operational efficiency. By leveraging predictive insights, organizations can enhance customer service by ensuring product availability, reduce overstock and stockouts, and adapt to changing market conditions, ultimately improving competitiveness and profitability.



PROCESS INVOLVED:

- Data Collection
- Data Preprocessing
- Feature Engineering
- Model Selection
- Training and Validation

- ❖ Model Tuning
- ❖ Prediction and Deployment
- ❖ Monitoring and Updating

DETAILS ABOUT DATASET:

The dataset used in the project is obtained from a website called Kaggle. Kaggle is a popular online platform for data science and machine learning enthusiasts. It provides a collaborative environment where data scientists, analysts, and machine learning practitioners can access datasets, participate in competitions, and share their insights and code. Users can explore real-world data challenges, develop predictive models, and compete for prizes or recognition. Kaggle also offers a wide range of tutorials, courses, and tools to support learning and development in the field of data science. It serves as a vibrant community hub where data professionals can collaborate, learn, and showcase their skills while solving complex data-related problems.

COLUMNS IN THE DATASET:

- ❖ **ID**: An identifier unique to each record in the dataset, used to distinguish and reference individual data points within the dataset.
- ❖ **Store ID**: A specific identification code or number assigned to each store, enabling the tracking of sales data for different retail locations.
- ❖ **Total Price**: The cumulative monetary value associated with a transaction or sale, including any discounts or additional charges applied.
- ❖ **Base Price**: The initial or standard price of a product or service before any discounts, promotions, or modifications are considered.
- ❖ **Units Sold**: The quantity or number of products or items purchased by customers in a particular transaction, indicating demand for a specific product.

LIBRARIES USED:

- ❖ **Numpy**: A Python library for efficient numerical operations and array manipulation.
- ❖ **Pandas:** Python library for data manipulation and analysis using Data Frame structures.

- ❖ **Matplotlib:** Python 2D plotting library for creating static, animated, or interactive visualizations.
- ❖ **Plotly:** Python 2D plotting library for creating static, animated, or interactive visualizations.
- ❖ **Sklearn:** Python library for machine learning and data analysis tools.

HOW TO DOWLOAD THE LIBRARIES:

We can download the libraries by using the “pip” commands in the command prompt by using the following steps:

1. Open the command prompt
2. Select the path where you need to install the libraries
3. “Pip install < Library Name >” replace the library name with what the library you need to install.
4. If there is any error like older version of pip then you need to upgrade your pip by using the below command

   “Python -m pip install —upgrade pip”

TRAIN AND TEST THE DATASET:

- ➢ To train data for product demand prediction using machine learning, we first prepare a historical dataset of product sales and associated features. Utilizing algorithms like ARIMA, we select a portion (e.g., 80%) of this data for training. During training, the model learns patterns and dependencies, enabling it to make predictions.

- ➢ For testing, we reserve the remaining data (e.g., 20%) as a separate dataset. The model predicts demand values for this test set, and we compare these predictions against the actual demand values to assess the model's accuracy and evaluate its performance in real-world scenarios. This rigorous testing ensures the model's reliability and effectiveness in forecasting future product demand.

ARIMA ALGORITHM:

ARIMA, or Auto Regressive Integrated Moving Average, is a popular time series forecasting algorithm used to analyse and predict temporal data patterns. It's widely employed in various fields like finance, economics, and weather forecasting.

ARIMA consists of three key components:

1. Auto Regression (AR): AR captures the linear relationship between a data point and its past values. This component examines how past observations influence the current one. By identifying the optimal number of lags (p), AR determines the order of autoregressive terms.

2. Integration (I): The "I" component focuses on making the time series data stationary. Stationarity means that the statistical properties of the data (e.g., mean, variance) remain constant over time. Differencing the data (d) is often performed to achieve this stationarity.

3. Moving Average (MA): MA models the relationship between a data point and past error terms (residuals). This component helps account for short-term fluctuations that aren't explained by the autoregressive component.

- ➢ Training an ARIMA model involves selecting the appropriate values for the parameters p, d, and q (order of the moving average terms). This is often done through a combination of statistical techniques, like autocorrelation and partial autocorrelation plots, and domain expertise.
- ➢ Once trained, the ARIMA model can be used for forecasting. It employs historical data to predict future values of the time series. Testing involves evaluating the model's accuracy by comparing its predictions to actual data. Common metrics for evaluation include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).
- ➢ In summary, ARIMA is a versatile algorithm for time series forecasting that encompasses autoregressive, integration, and moving average components. Training entails selecting appropriate parameters, and testing assesses the model's accuracy using various metrics to ensure it provides reliable predictions for future data points.

CONCLUSION:

In conclusion, Phase 2 of the "Product Demand Using Machine Learning" project lays the groundwork for an ambitious journey to enhance inventory management and production planning. Our mission is to create a machine learning model that predicts product demand by utilizing historical sales data and external factors. This phase prioritizes the essential steps of data collection, preprocessing, and feature engineering, ensuring a solid foundation for subsequent phases. The focus on model selection, training, and evaluation reflects our commitment to delivering accurate forecasts. As we move forward, we remain dedicated to optimizing business operations, minimizing costs, and meeting customer needs efficiently through the power of data-driven insights.

Member: MOHAMMED VAZHIL KHAN K
Reg No: 310521104061