PROJECT DEMAND PREDICTION WITH MACHINE LEARNING

Phase 1:problem definition and design thinking

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. This project involves data collection, data preprocessing, feature engineering, model selection, training, and evaluation.

1.Data collection:-

Gathering historical sales data, which serves as the foundation for forecasting. Additionally, external factors such as economic indicators, seasonality, marketing campaigns, and competitor activities are collected to provide a holistic view of the demand-driving factors.

Datasetlink:- https://www.kaggle.com/datasets/chakradharmattapalli/product-demand-prediction-with-machine-learning

2.Dataprocessing:-

The collected data often requires cleaning and preprocessing to handle missing values, outliers, and inconsistencies. Proper preprocessing ensures that the data is ready for model training and evaluation.

Handling Missing Value:-

Data collected from various sources may have missing values, which can adversely affect model performance. These missing values need to be addressed before proceeding.

Techniques for handling missing values include imputation, where missing values are replaced with estimated or calculated values (e.g., mean, median, or mode of the feature), or deletion, where rows or columns with missing values are removed if they do not contain critical information.

Outlier Detection and Treatment:-

Outliers are data points that significantly deviate from the majority of the data. These outliers can distort statistical analyses and model predictions.

Various methods, such as statistical tests or visualization techniques like box plots, are employed to identify outliers. Once identified, outliers can be handled by either removing them, transforming them, or using robust modeling techniques that are less sensitive to outliers.

Dealing with Inconsistencies:-

Data inconsistencies can arise due to errors in data collection, entry, or formatting. These inconsistencies must be identified and resolved.

This might involve standardizing date formats, ensuring consistent units of measurement, and rectifying inconsistent labeling or categorization of data.

3. Feature Engineering:

Feature engineering is a crucial step to extract meaningful information from the data. It involves creating new features or transforming existing ones to better represent the underlying patterns and relationships in the data. Time-based features, lag features, and statistical aggregates are common techniques used to enhance the dataset.

4.Prediction Models:

We explore a range of predictive models suitable for demand forecasting, such as time series models (e.g., ARIMA, SARIMA), regression models, ensemble methods (e.g., Random Forest, Gradient Boosting), and deep learning models (e.g., LSTM, Transformer). Model selection is based on their suitability for addressing the specific forecasting challenge.

5.Model Training and Validation:

The selected model is trained on the preprocessed data. To assess model performance, we partition the dataset into training and validation sets and fine-tune hyperparameters. Regular validation ensures the model's reliability in generating accurate predictions.

6.Evaluation Metrics:

We employ established evaluation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), to quantify the accuracy of our predictions. These metrics enable us to measure the alignment between predicted and actual demand.

7.conclusion:-

This project represents a significant step toward more efficient and data-driven inventory management and production planning. It equips businesses with the tools they need to stay competitive, reduce waste, and meet customer demands effectively.