

Product Demand Prediction with Machine Learning

Phase - 3

- **Data Collection:**First, you need to collect historical sales data and any relevant external factors that might influence product demand.
- This can include factors such as economic indicators, seasonality, marketing campaigns, competitor data, or any other relevant data sources.
- **Data Preprocessing:**Data preprocessing is a crucial step in preparing the dataset for analysis. This typically involves the following steps:
- **Handling Missing Values:**Check for missing values in the dataset and decide on an appropriate strategy to handle them, such as imputation or removal.

- **Data Cleaning:** Clean the data by removing any irrelevant or redundant features that do not contribute to the analysis.
- **Data Transformation:** If required, perform data transformations such as normalization or standardization to bring the data to a common scale.
- **Feature Engineering:** Create new features that might help improve the predictive performance of the model. This could involve generating lag features, creating new variables from existing ones, or extracting relevant information from timestamps.

- **Handling Categorical Variables:** Convert categorical variables into a format suitable for analysis, such as one-hot encoding or label encoding.
- **Data Splitting:** Split the dataset into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance.
- **Time Series Analysis (if applicable):** If you're dealing with time series data, perform time series analysis to understand any patterns or trends in the data. This can involve decomposing the time series, checking for seasonality, and examining autocorrelation.

- **Data Integration:** Integrate the external factors data with the historical sales data. This may involve merging datasets based on common features or time stamps.
-
1. **Data Scaling:** Scale the data if necessary, especially if you're using algorithms that are sensitive to the scale of the features.
 2. **Feature Selection:** Select the most relevant features that contribute to the prediction task. This can help in reducing the complexity of the model and improving its generalization capability.