IBM APPLIED DATA SCIENCE PROJECT

Phase 1: Problem Definition and Design Thinking

Sachin Kumar Singh

Problem Definition:

The problem at hand is to develop a machine learning model that can accurately forecast product demand based on historical sales data and external factors. This project aims to assist businesses in optimizing their inventory management and production planning processes to efficiently meet customer demands. To achieve this, we will undertake several key steps, including data collection, data preprocessing, feature engineering, model selection, model training, and evaluation.

Design Thinking:

To effectively address this problem, we need to follow a structured approach. The following steps outline our design thinking process:

1. Data Collection:

- Our first step is to gather historical sales data and external factors that influence product demand. These external factors may include marketing campaigns, holidays, economic indicators, and any other relevant information. This data will serve as the foundation for our demand forecasting model.

2. Data Preprocessing:

- Clean and preprocess the collected data to ensure its quality and reliability. This process involves handling missing values, outliers, and other data anomalies. We will also convert categorical features into numerical representations to make them suitable for machine learning algorithms.

3. Feature Engineering:

- Feature engineering is a critical step to enhance the predictive power of our model. We will create additional features that capture seasonal patterns, trends, and external influences on product demand. These engineered features will provide valuable insights to our model during training.

4. Model Selection
- Choosing the right regression algorithms is crucial for accurate demand forecasting. We will evaluate various regression techniques, including Linear Regression, Random Forest, and XGBoost, among others. The selection will be based on their performance and suitability for our specific problem.

5. Model Training:

- Once the model is selected, we will train it using the preprocessed data. Training involves feeding the historical data into the chosen algorithm to enable it to learn patterns and relationships between variables. The model will adjust its parameters to make accurate demand predictions.

6. Model Evaluation:

- After training, we will assess the model's performance using appropriate evaluation metrics. This step is essential to ensure that the model meets our accuracy and reliability requirements. If necessary, we will fine-tune the model or explore other algorithms.

7. Deployment and Integration:

- In later project phases, we will explore deployment options and integration with business systems. This will involve creating a user-friendly interface for businesses to interact with the model and receive demand forecasts.

8. Monitoring and Maintenance:

- Ongoing monitoring and maintenance of the model will be essential to ensure its accuracy over time. We will establish procedures to update the model with new data and adapt it to changing business conditions.