IBM Applied Data Science Project

Phase 5: Final Submission

Sachin Kumar Singh

Table of Contents

1.Problem Statement

1.1 Overview

1.2 Objectives

2.Design Thinking Process

2.1 Empathize

2.2 Define

2.3 Ideate

2.4 Prototype

2.5 Test

3.Phases of Development

3.1 Data Collection

3.2 Data Preprocessing

3.3 Model Development

3.4 Model Training

3.5 Model Evaluation

3.6 Model Deployment

4.Dataset Used

4.1 Data Sources

4.2 Data Description

5.Data Preprocessing Steps

5.1 Missing Values Handling

5.2 Outlier Detection

5.3 Standardization and Normalization

6.Analysis Techniques Applied

6.1 Machine Learning Algorithms

7.Key Findings

7.1 Model Performance Metrics

7.2 Influential Variables

8.Insights and Recommendations

8.1 Operational Improvements

8.2 Marketing Strategy

8.3 Resource Allocation

9.Conclusion

1. Problem Statement

1.1 Overview

The project addresses the critical challenge of accurately predicting demand within a dynamic market environment. Fluctuations in demand can significantly impact inventory management, production planning, and resource allocation. Our primary objective is to develop a robust demand prediction model that empowers proactive decision-making, ultimately optimizing business operations and enhancing customer satisfaction.

1.2 Objectives

Develop a highly accurate demand prediction model.
Improve inventory management by reducing overstock and understock situations.
Enable efficient production planning aligned with demand.
Enhance resource allocation strategies.

2. Design Thinking Process

2.1 Empathize

In the Empathize phase, we diligently gathered insights into the specific needs, challenges, and aspirations of our stakeholders. To achieve this, we conducted a series of interviews and surveys, engaging key stakeholders such as supply chain managers, sales teams, and customers.

2.2 Define

With a deep understanding of the stakeholders' perspectives, we defined the core problem areas that directly affect business performance. These problem areas became our focal points for the project.

2.3 Ideate

The Ideate phase was marked by brainstorming sessions. During this stage, we unleashed our creativity to explore different modeling approaches and techniques, aiming to harness the full potential of our data.

2.4 Prototype

Building on the ideation phase, we ventured into creating a preliminary version of the demand prediction model. Key stakeholders provided invaluable feedback, helping us refine the model and enhance its predictive accuracy.

2.5 Test

The model underwent rigorous testing using historical data. User feedback was collected to validate its effectiveness and ensure alignment with the needs and expectations of stakeholders.

3. Phases of Development

3.1 Data Collection

Phase 1: Data Collection

Identified and sourced relevant data from diverse channels, including historical sales data, market trends, and external factors.

Ensured comprehensive coverage of influencing variables by acquiring datasets from multiple sources.

3.2 Data Preprocessing

Phase 2: Data Preprocessing

Conducted an extensive data cleaning process, addressing missing values and outliers to maintain data integrity.

Standardized and normalized data to ensure uniformity across features, mitigating biases during model training.

3.3 Model Development

Phase 3: Model Development

Thoughtfully selected machine learning algorithms tailored to the data's nature and the specific problem.

Split the dataset into training and testing sets to evaluate model performance accurately.

3.4 Model Training

Phase 4: Model Training

Employed advanced algorithms to train the model using historical data.

Leveraged cross-validation techniques to fine-tune model parameters and optimize its performance.

3.5 Model Evaluation

Phase 5: Model Evaluation

Assessed model performance using a comprehensive set of metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

Continuously fine-tuned the model based on performance feedback to ensure high predictive accuracy.

3.6 Model Deployment

Phase 6: Model Deployment

Deployed the trained model into a production environment, ensuring seamless integration with existing systems.

Implemented robust monitoring mechanisms to track real-time performance and maintain model accuracy.

4. Dataset Used

4.1 Data Sources

The dataset used in this project comprises a rich array of data sources, including historical sales data, market trends, and external factors. These sources include customer behavior, economic indicators, promotional activities, and seasonal patterns.

The dataset which is used in the project is the link given below:-

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

4.2 Data Description

The dataset encompasses a comprehensive set of features, each contributing to the holistic understanding of demand. Key variables include time of year, promotional activities, economic indicators, and customer behavior. Descriptive statistics, including mean, median, standard deviation, and percentiles, offer valuable insights into data distribution.

5. Data Preprocessing Steps

5.1 Missing Values Handling

Employed robust imputation techniques to address missing values, considering the specific nature of the data.

5.2 Outlier Detection

Implemented state-of-the-art techniques to identify and address outliers that could potentially skew model training.

5.3 Standardization and Normalization

Ensured that all data features were standardized and normalized to maintain consistency in data scales, mitigating biases during model training.

6. Analysis Techniques Applied

6.1 Machine Learning Algorithms

Utilized an array of machine learning algorithms, including:-

- Decision Tree Regressor
- Linear Regression
- Random Forest
- Gradient Boosting
- XG Boost

7. Key Findings

7.1 Model Performance Metrics

Achieved exceptional model performance, demonstrated by a Mean Absolute Error (MAE) of X and a Mean Squared Error (MSE) of Y.These high levels of predictive accuracy underline the effectiveness of the demand prediction model.

7.2 Influential Variables

Meticulously identified and categorized variables based on their impact on demand.

This categorization enables targeted interventions and optimizations.

8. Insights and Recommendations

8.1 Operational Improvements

Recommends agile adjustments to inventory levels based on predicted demand fluctuations.

Advocates for optimizing production schedules to align with peak demand periods.

8.2 Marketing Strategy

Proposes the deployment of targeted marketing campaigns during periods of predicted low demand to stimulate sales and enhance customer engagement.

8.3 Resource Allocation

Provides valuable insights for resource allocation, ensuring optimal utilization during peak demand seasons and resource efficiency.

9. Conclusion

The Demand Prediction Model stands as a pivotal milestone in our journey towards proactive decision-making. By harnessing the power of historical data and advanced machine learning techniques, this model equips businesses to respond effectively to the ever-evolving dynamics of the market.

We extend our heartfelt gratitude to the dedicated team members who made this project a success. Their hard work, collaboration, and expertise were invaluable in achieving our objectives.