

**Data Science & ML Internship – Intern**

Assessment Report

By :

Aadhil Shihabdeen

[aadhillinked@gmail.com](mailto:aadhillinked@gmail.com)

+94752203374

Table of Contents

[**Introduction** 3](#_Toc179850541)

[**Data Description** 3](#_Toc179850542)

[**Methodology** 4](#_Toc179850543)

[**Results** 5](#_Toc179850544)

[**Business Implications** 5](#_Toc179850545)

[**Conclusion** 5](#_Toc179850546)

# **Introduction**

**Objective:**

The goal of this project is to enhance business operations by implementing two key data-driven solutions: Customer Personalization and Demand Forecasting. These solutions aim to improve customer engagement and optimize inventory management respectively, leveraging advanced analytics and machine learning techniques.

**Background:**

In a competitive retail environment, understanding customer preferences and accurately forecasting product demand are crucial for staying competitive. Customer personalization enhances shopping experiences and customer loyalty by tailoring products and offers to individual preferences. Demand forecasting allows businesses to effectively manage inventory, reduce costs, and prevent stockouts or overstock situations.

**Data Description**

**Data Source:**

The data used in this project comes from a comprehensive dataset of supermarket transactions over two years, covering various branches and including details such as product types, sales amounts, and transaction volumes.

**Preprocessing:**

The dataset was cleaned and normalized by:

* Removing irrelevant or redundant features.
* Handling missing values and outliers.
* Standardizing categorical and numerical data to ensure consistency.

# **Methodology**

**Model Description:**

**Customer Personalization Model:**

* **Features Used:** The model incorporates features like purchase frequency, average spending, and most preferred item types. These were chosen to capture distinct customer behaviors and preferences.
* **Machine Learning Technique:** Decision Trees were used for their ability to model non-linear relationships and their interpretability. This allows for easy extraction of rules that can be applied for personalized marketing strategies.

**Demand Forecasting Model:**

* **Approach:** The model aggregates historical sales data weekly. This granularity was chosen to balance between capturing trends and managing computational efficiency.
* **Forecasting Technique:** Gradient Boosting Machines were employed for their robustness and effectiveness in handling various types of data and their ability to improve predictive accuracy through ensemble learning.

**Model Training and Validation:**

* **Training and Validation:** Both models were trained on a split of historical data, with 80% used for training and 20% for validation, ensuring that the models generalize well to new data.
* **Parameter Tuning:** Techniques such as Grid Search were utilized to fine-tune the hyperparameters of the models to optimize performance.

# **Results**

**Model Performance:**

* **Visualization:** Graphs were created to compare actual and predicted values to visually assess the models' performance. Residual plots were also used to identify any systematic errors in the forecasts.

**Interpretation:**

* The results demonstrated that **the Customer Personalization Model** could effectively predict customer preferences, which can be leveraged to enhance marketing efforts. **The Demand Forecasting Model** showed promising accuracy in predicting weekly sales, crucial for inventory planning.

**Business Implications**

**Applications:**

* **Customer Personalization:** This model can be integrated into CRM systems to provide personalized product recommendations and offers, significantly enhancing customer engagement and satisfaction.
* **Demand Forecasting:** The insights from this model can guide inventory stocking decisions, helping to reduce costs associated with overstocking and stockouts, and ensuring that customer demand is met efficiently.

**Recommendations:**

* Further refine the models by incorporating additional data sources, such as customer demographic information and economic indicators, to enhance predictive accuracy.
* Implement continuous learning mechanisms to allow models to adapt to new trends and changes in customer behavior over time.

# **Conclusion**

While the models provide a strong foundation, they are based on historical sales data and assume consistent customer behavior patterns. Future work could explore real-time data integration and more dynamic modeling approaches to keep pace with market changes and customer dynamics. **IF I OFFERED MORE TIME, I CAN BUILD A FULL STACK APPLICATION TO USE THIS MODELS EFFECTIVELY**