

Assignment Report: Material Reservation Prediction System

https://github.com/trietnm2/utac_assignment

1. Time Difference Analysis in Material Demand

A comprehensive analysis was conducted to evaluate the time difference in material demand over daily, weekly, and monthly intervals. The key findings are as follows:

- **90%** of materials experience demand within **1 month**.
- **5%** of materials have demand after **2 months**.
- **4%** of materials exhibit demand after **7 months**.

	date_diff	week_diff	month_diff
count	7392.000000	7392.000000	7392.000000
mean	19.795049	2.434524	0.408279
std	38.181163	5.443406	1.302237
min	0.000000	0.000000	0.000000
0%	0.000000	0.000000	0.000000
10%	1.000000	0.000000	0.000000
20%	2.000000	0.000000	0.000000
30%	4.000000	0.000000	0.000000
40%	6.000000	0.000000	0.000000
50%	7.000000	1.000000	0.000000
60%	11.000000	1.000000	0.000000
70%	16.000000	2.000000	0.000000
80%	25.000000	3.000000	0.000000
90%	47.000000	6.000000	1.000000
95%	74.000000	10.000000	2.000000
97%	111.270000	15.270000	3.270000
99%	201.000000	28.000000	7.000000
max	567.000000	81.000000	20.000000

Based on these insights, we utilized data from the preceding **7 months** to forecast the material demand for the upcoming **1 month**.

2. Feature Selection and Engineering

ChargingCC	MovementType	ReceivingLoc	Material Number	CreatedOn	ReserveQty	MaterialTypeCode
83070101	311	LSI-LSI	ZMZ003145GBA	1/2/23 1:28	105000.0	XZDRMY
83070101	311	LSI-LSI	ZMZ003145GBA	1/2/23 1:28	105000.0	XZDRMY
83070101	311	LSI-LSI	ZMZ003145GBA	1/2/23 1:28	105000.0	XZDRMY
83070101	311	LSI-LSI	ZMZ003147GBA	1/2/23 1:28	14000.0	XZDRMY
83070101	311	LSI-LSI	ZMZ003147GBA	1/2/23 1:28	14000.0	XZDRMY

The prediction model is based on a comprehensive set of features, including:

- **Year and Month:** Temporal features representing the observation period.
- **Metadata columns:** Available information related to the material's characteristics.
- **1D Feature Set:** Individual features generated from each type of metadata.
- **2D Feature Set:** Interaction-based features between different metadata variables.

A total of **12,820 features** were generated to support the prediction process.

The Meta columns:

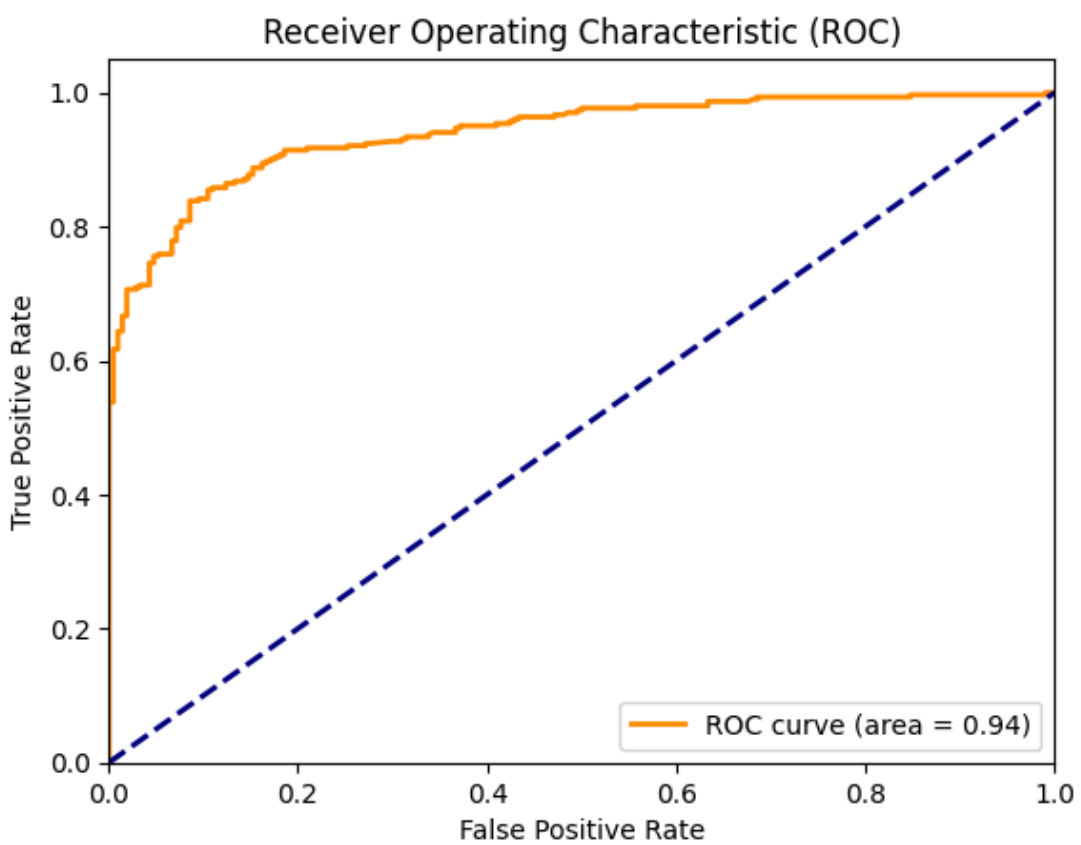
- **MaterialNumber**
- **MaterialTypeCode**
- **ChargingCC**
- **MovementType**
- **ReceivingLoc**
- **Year_month**
- **Date_diff**
- **Week_diff**
- **Month_diff**
- **ReserveQty** (Current Demand)
- **ReserveQty_lag1**
- **ReserveQty_lag2**
- **ReserveQty_lag3**
- **ReserveQty_lag4**
- **ReserveQty_lag5**
- **ReserveQty_lag6**
- **ReserveQty_next** (Demand in next month. Target value of Regression Model)
- **Has_next** (Does it have demand next month or not? Label value of Classification Model)

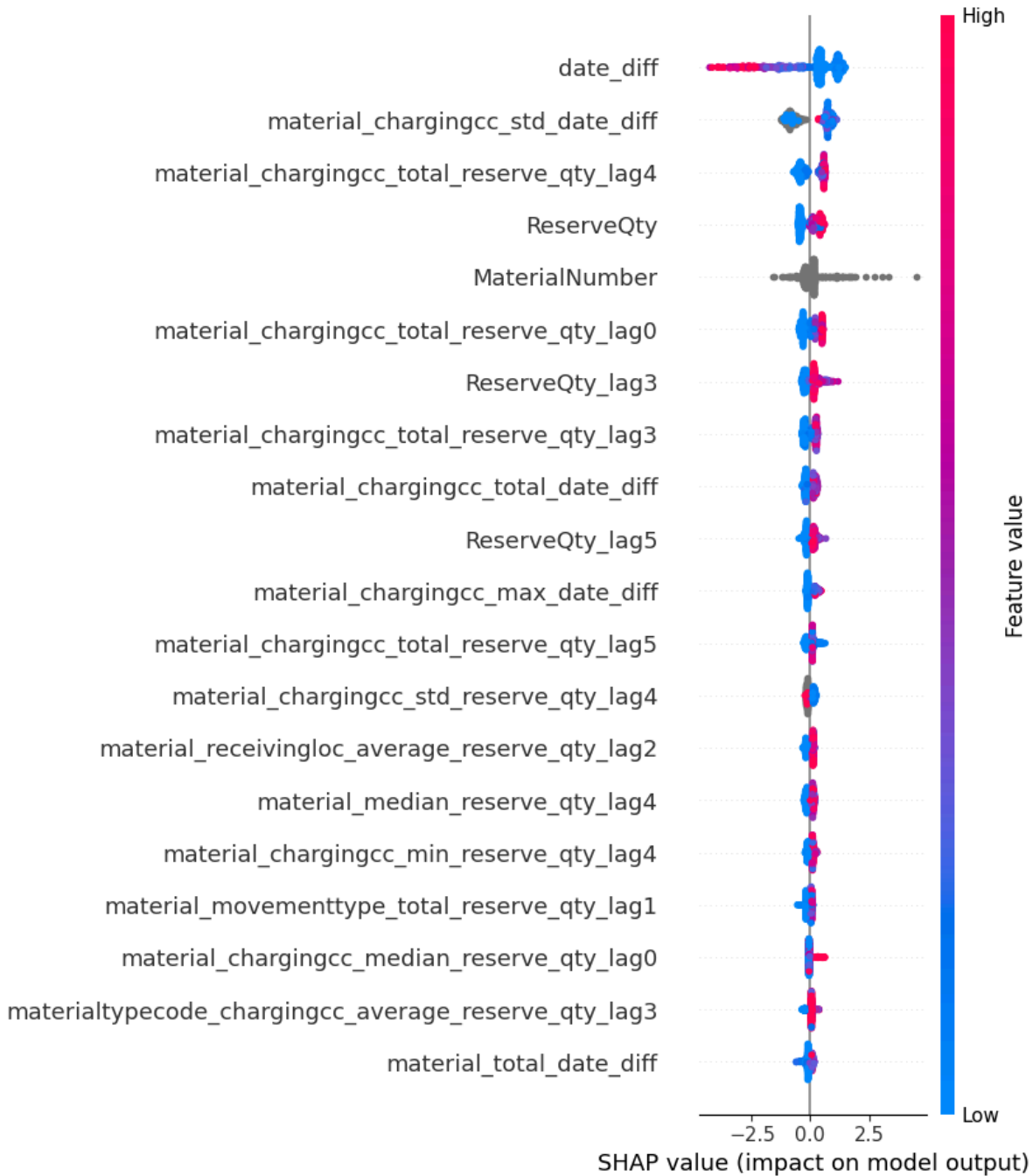
3. Problem Decomposition

The prediction problem was divided into **two smaller tasks** for efficient handling:

1. Demand Prediction (Binary Classification)

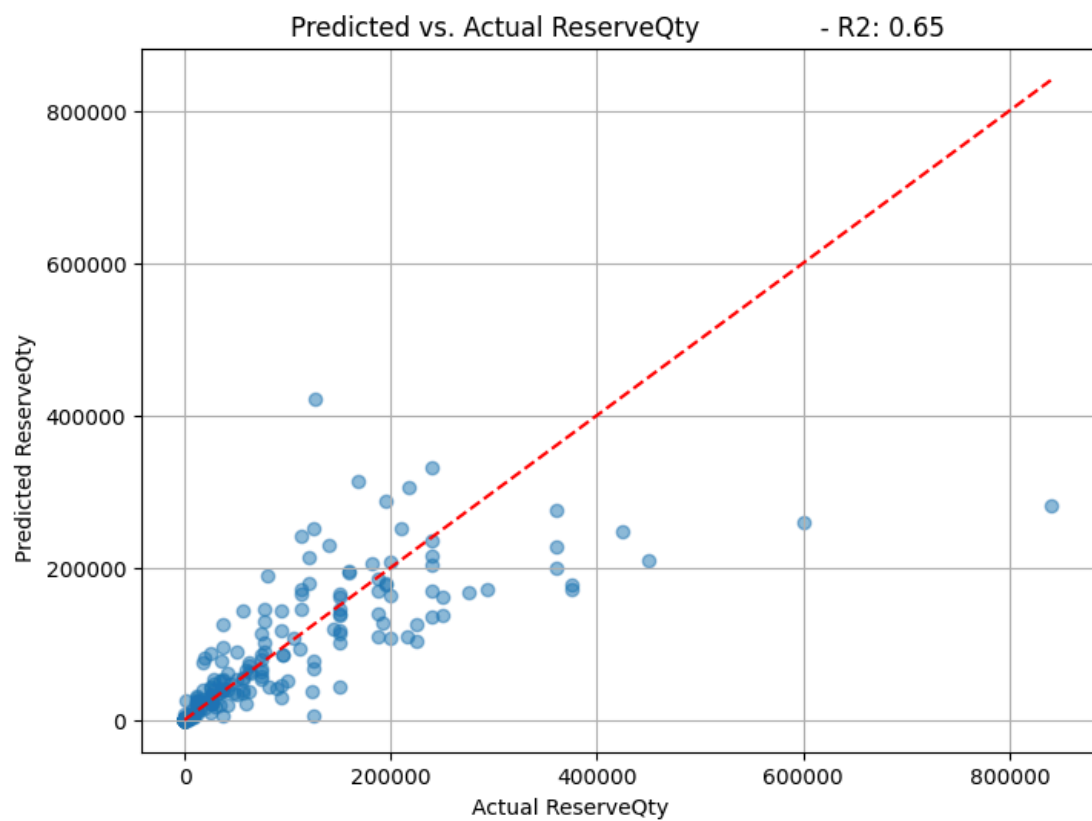
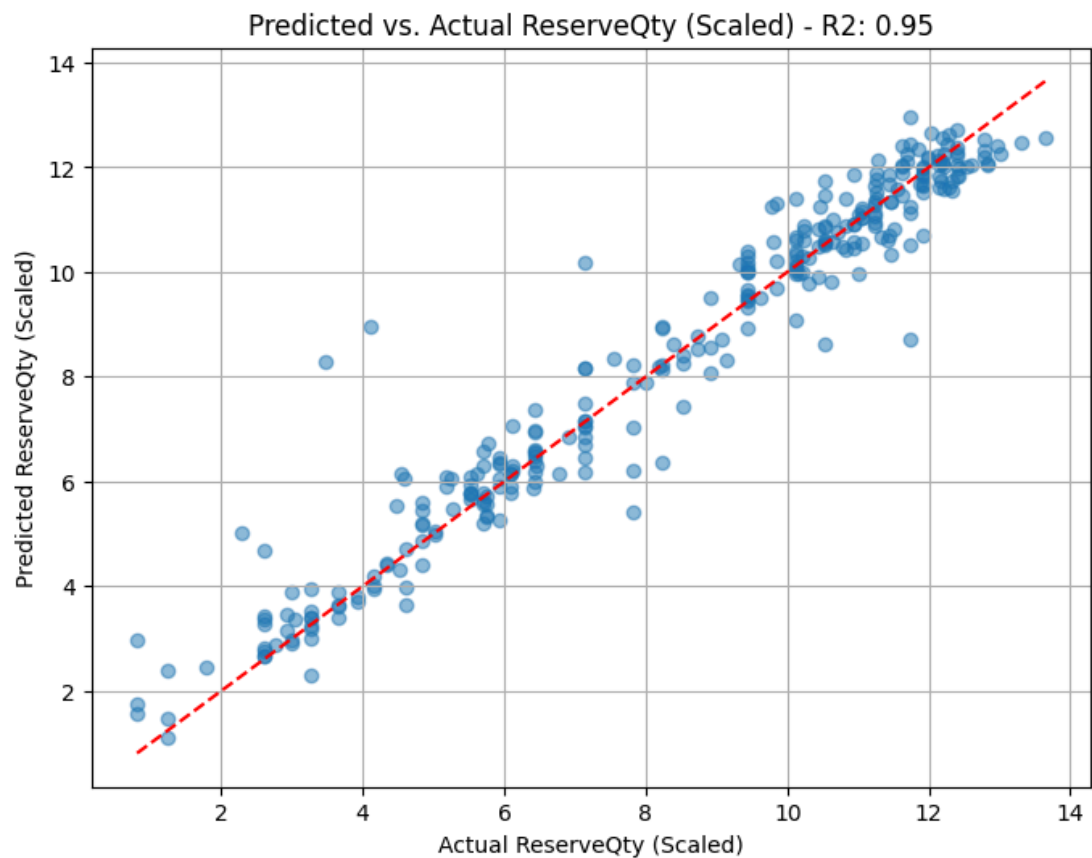
- This sub-task focuses on predicting whether there will be material demand in the next month.
- It is framed as a **binary classification problem** (Yes/No).
- The key metric for model performance is the **AUC-ROC**, which achieved a value of **0.94** (where values closer to 1 are ideal).
- **107 features** were selected based on feature importance for the classification model.

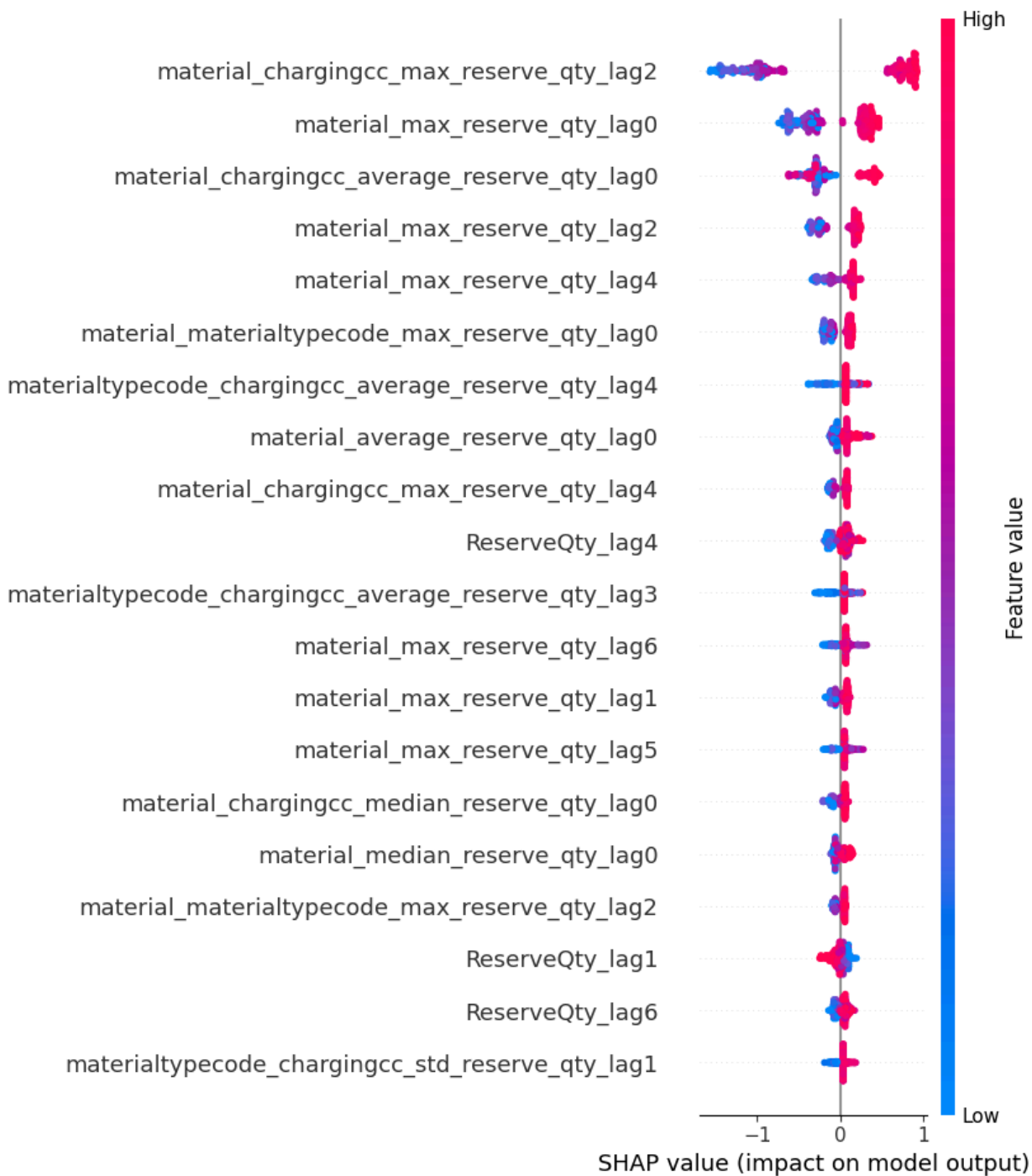




2. Quantity Prediction (Regression)

- In this sub-task, we predict the quantity of material demand for the next month, given that demand exists.
- It is treated as a **regression problem**.
- The primary metric is **R^2** , which achieved **0.95** on scaled values and **0.65** on actual values (closer to 1 indicates better model performance).
- **107 features** were selected for the regression model, also based on feature importance.

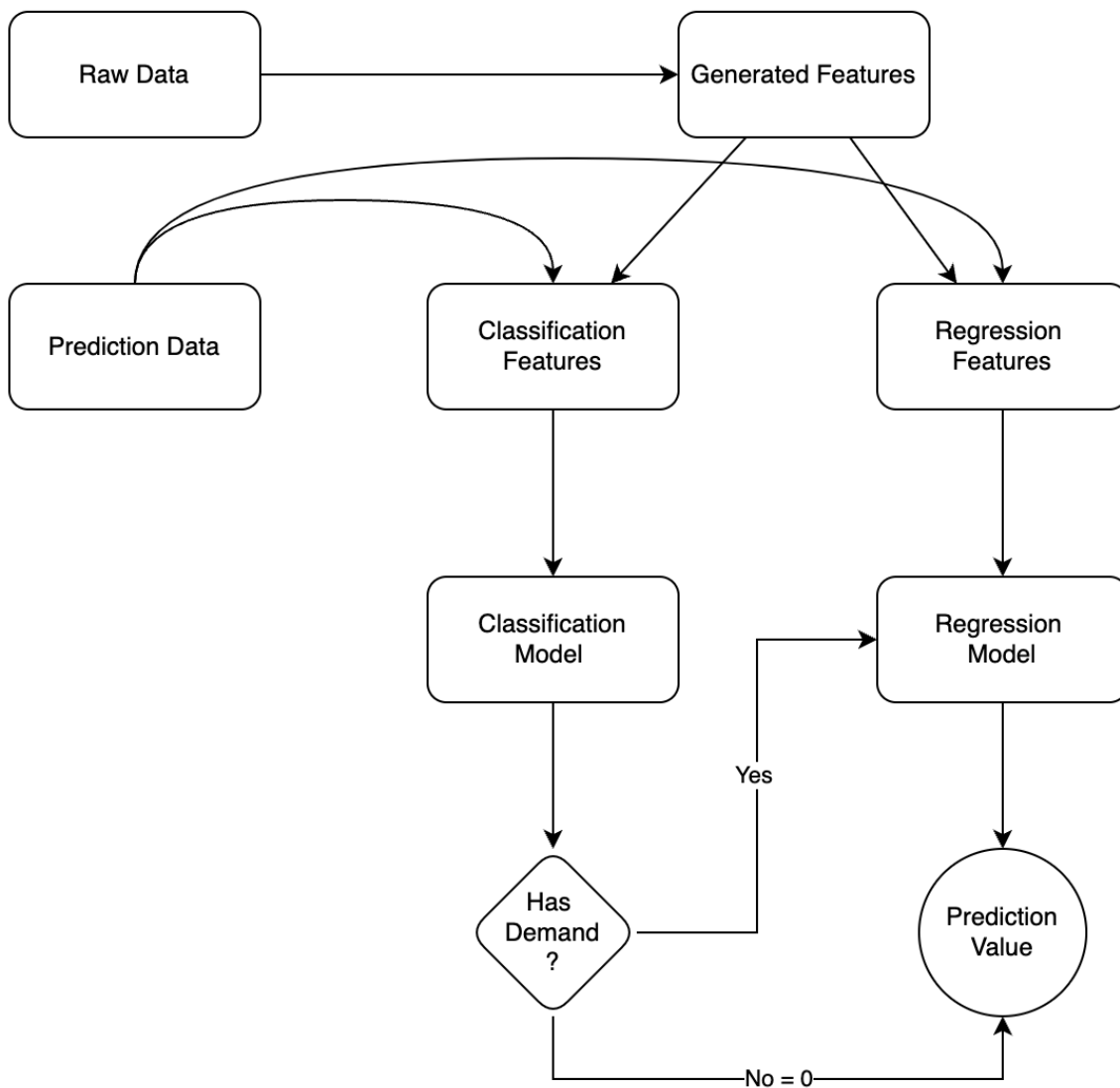




4. Prediction Pipeline Overview

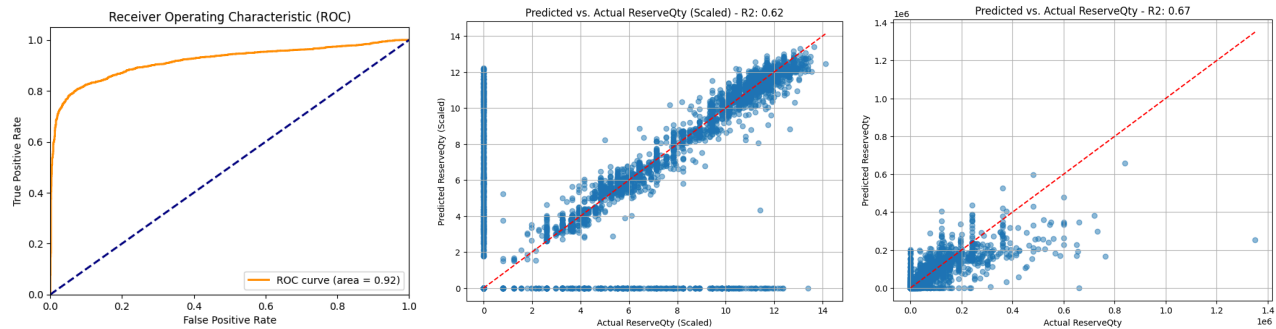
The pipeline for predicting material demand follows a structured process:

1. Input data is combined with the extracted feature set.
2. The **classification model** first determines whether there will be demand for the next month.
3. If no demand is predicted, the system returns a prediction of **0**.
4. If demand is predicted, the **regression model** is used to estimate the quantity of material required for the next month.



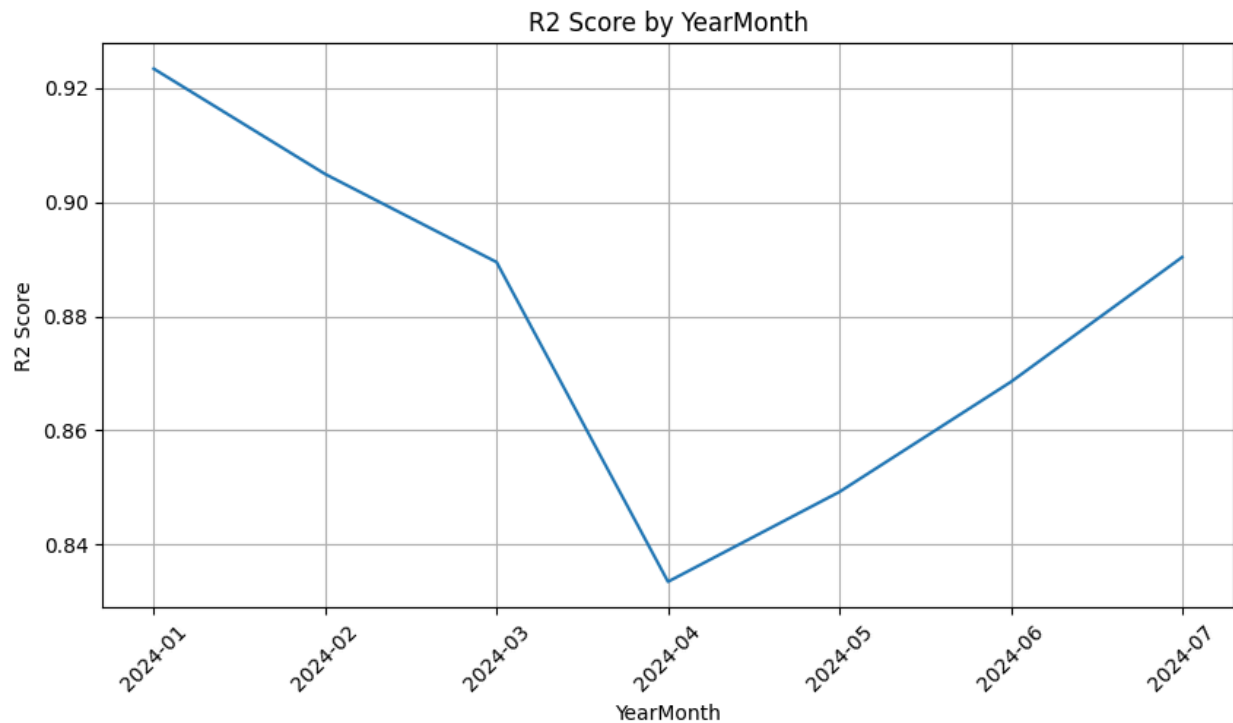
5. Model Training, Testing, and Evaluation

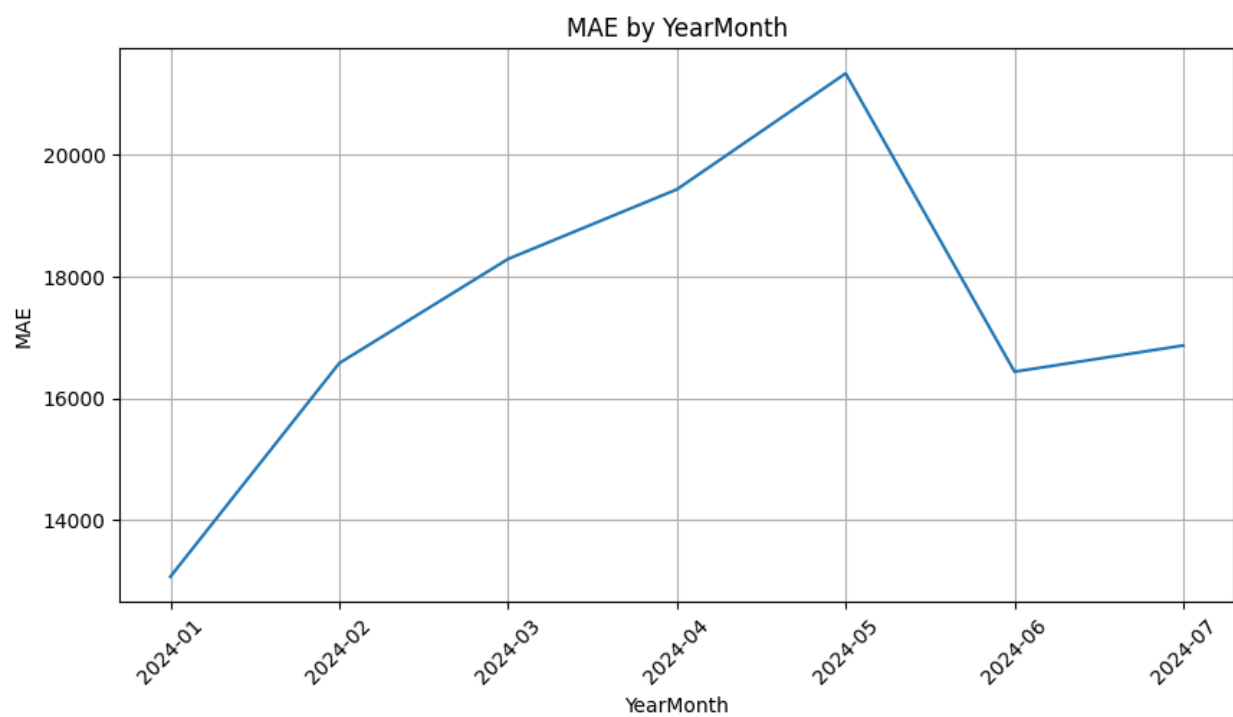
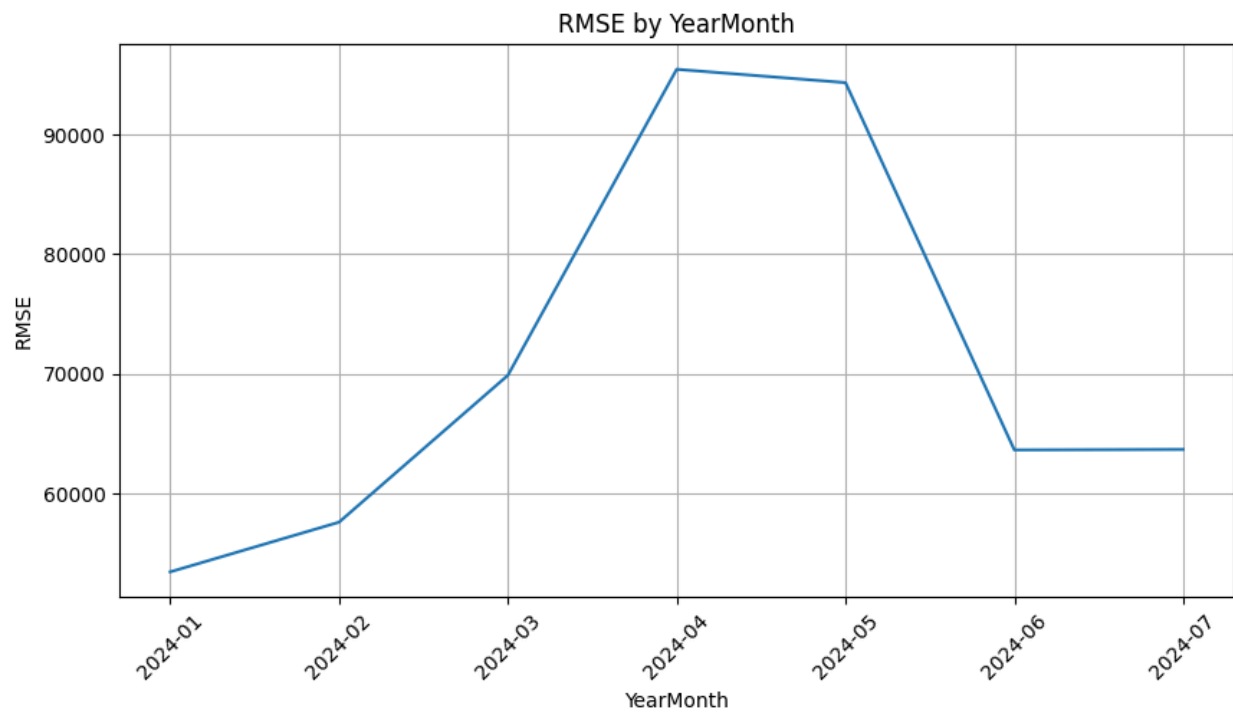
- **Training Phase:** The models were trained using data from **12 months of 2023**.
- **Testing Phase:** The trained models were tested on data spanning from **January to July 2024**.



Evaluation results show:

- **R²** values ranged from **0.83 to 0.93**, with the highest performance occurring in **January 2024** and the lowest in **April 2024**.
- **RMSE** (Root Mean Square Error) and **MAE** (Mean Absolute Error) demonstrated stable performance across the months.





6. Top Demand Prediction

- Top 10 Most Demand Materials

