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:

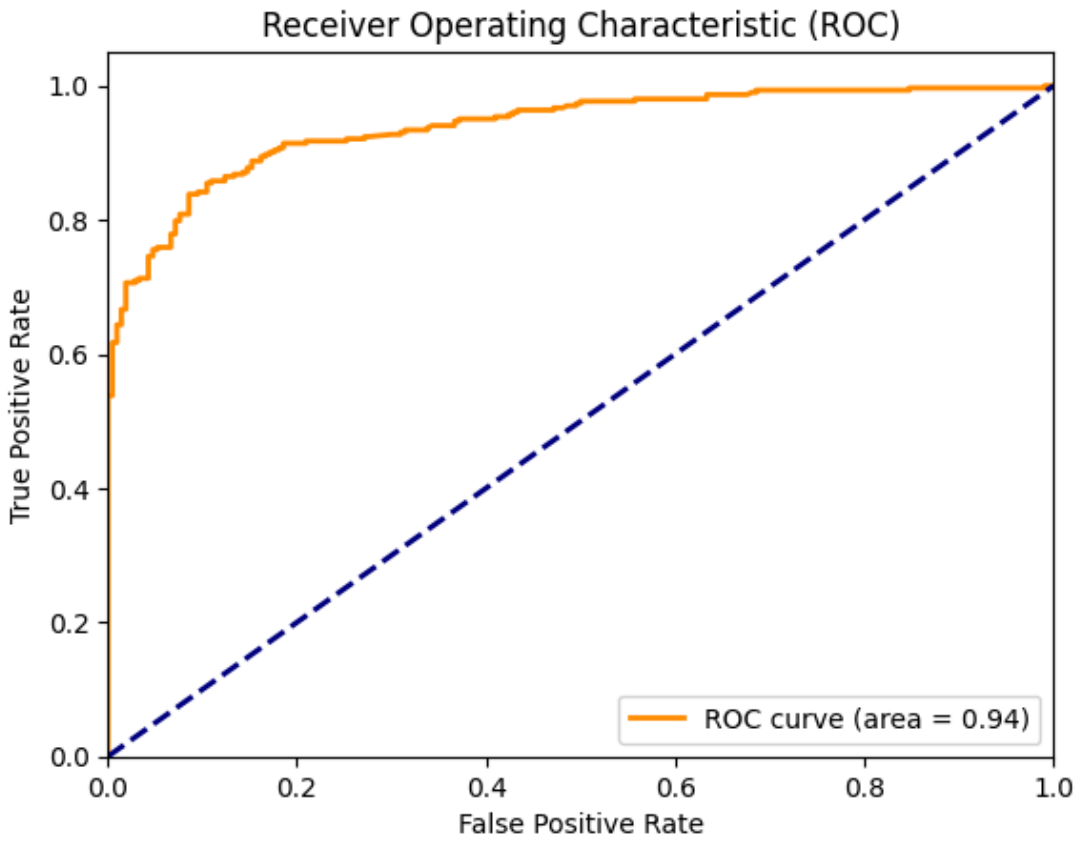
- **Year and Month**: Temporal features representing the observation period.
- **Metadata columns**: Available information related to the material's characteristics.
- **1D Feature Set**: Individual features generated from each type of metadata.
- **2D Feature Set**: Interaction-based features between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

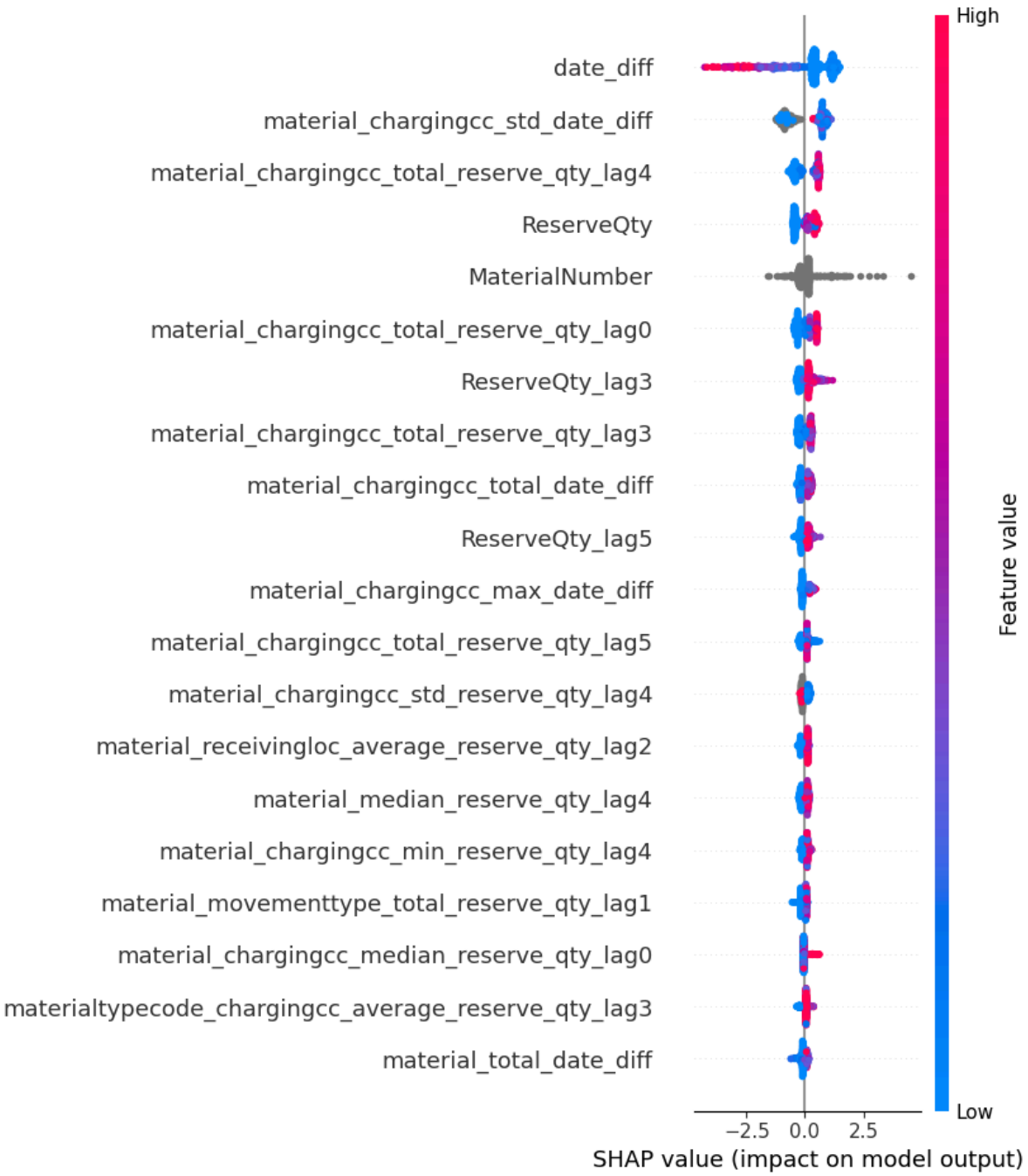
A total of **12,820 features** were generated to support the prediction process.

3. Problem Decomposition

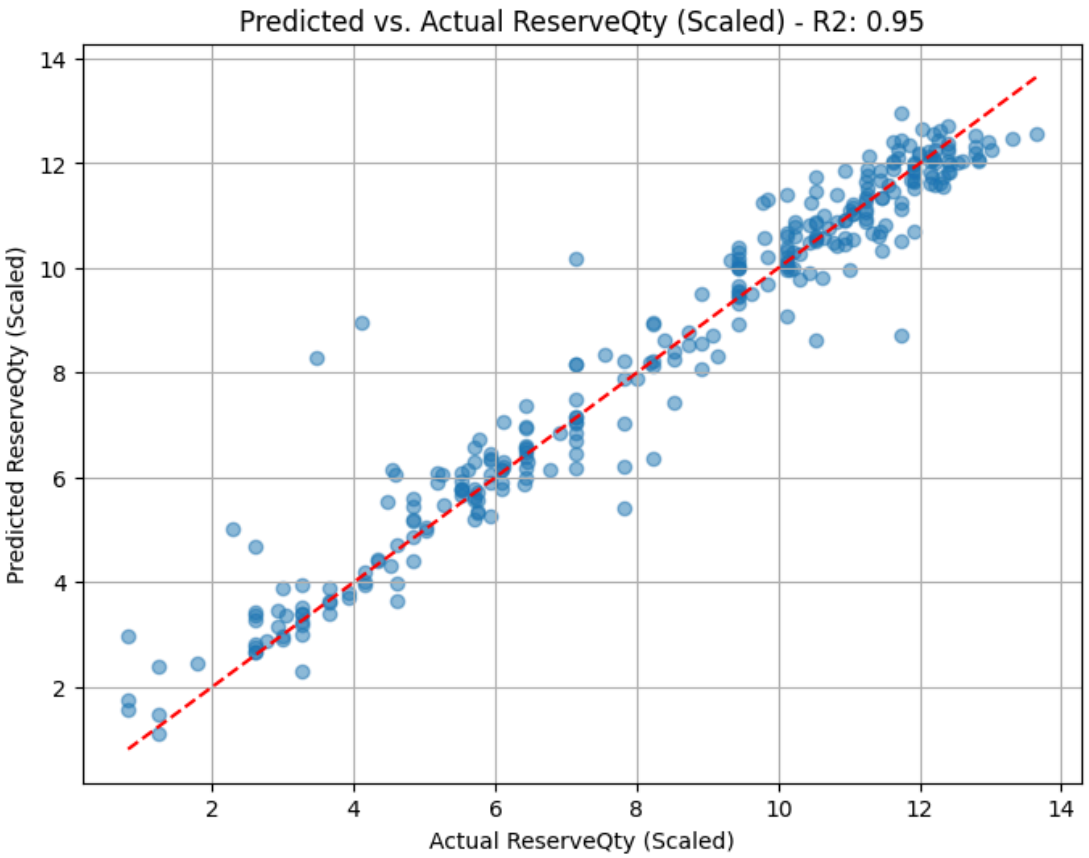
The prediction problem was divided into **two smaller tasks** for efficient handling:

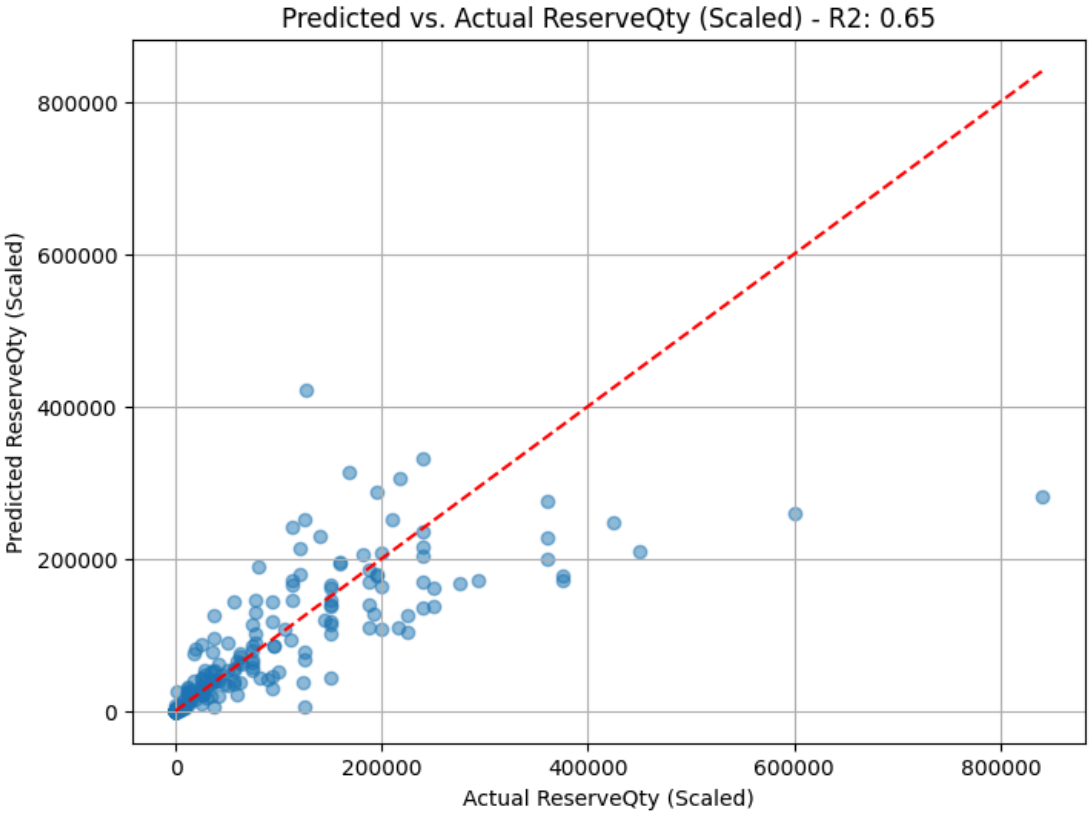
1. **Demand Prediction (Binary Classification)**
    - This sub-task focuses on predicting whether there will be material demand in the next month.
    - It is framed as a **binary classification problem** (Yes/No).
    - The key metric for model performance is the **AUC-ROC**, which achieved a value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

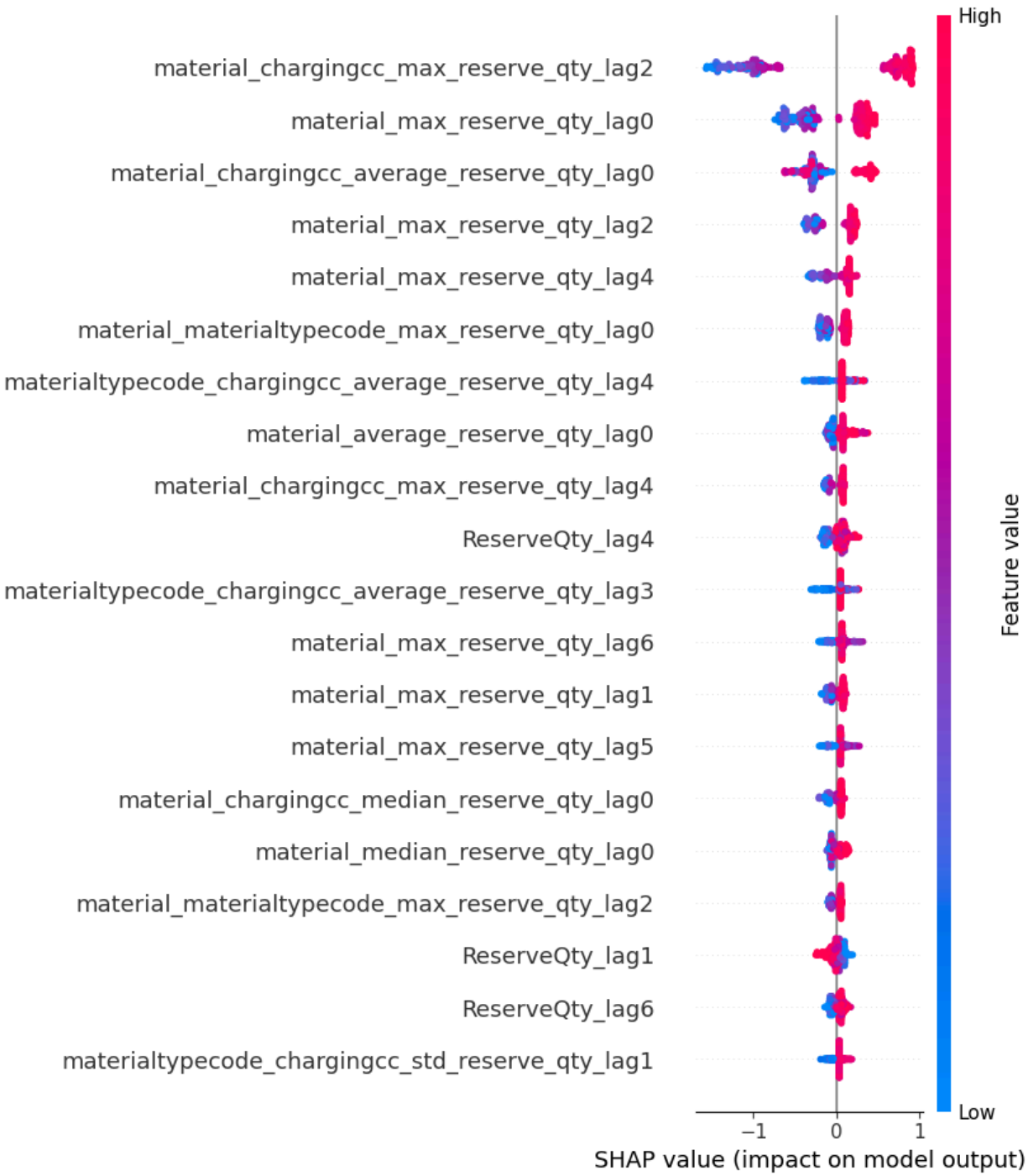




2. **Quantity Prediction (Regression)**
    - In this sub-task, we predict the quantity of material demand for the next month, given that demand exists.
    - It is treated as a **regression problem**.
    - The primary metric is **R^2**, which achieved **0.95** on scaled values and **0.65** on actual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


Predicted vs. Actual ReserveQty (Scaled) - R2: 0.95
Predicted ReserveQty (Scaled)
Actual ReserveQty (Scaled)
2
4
6
8
10
12
14

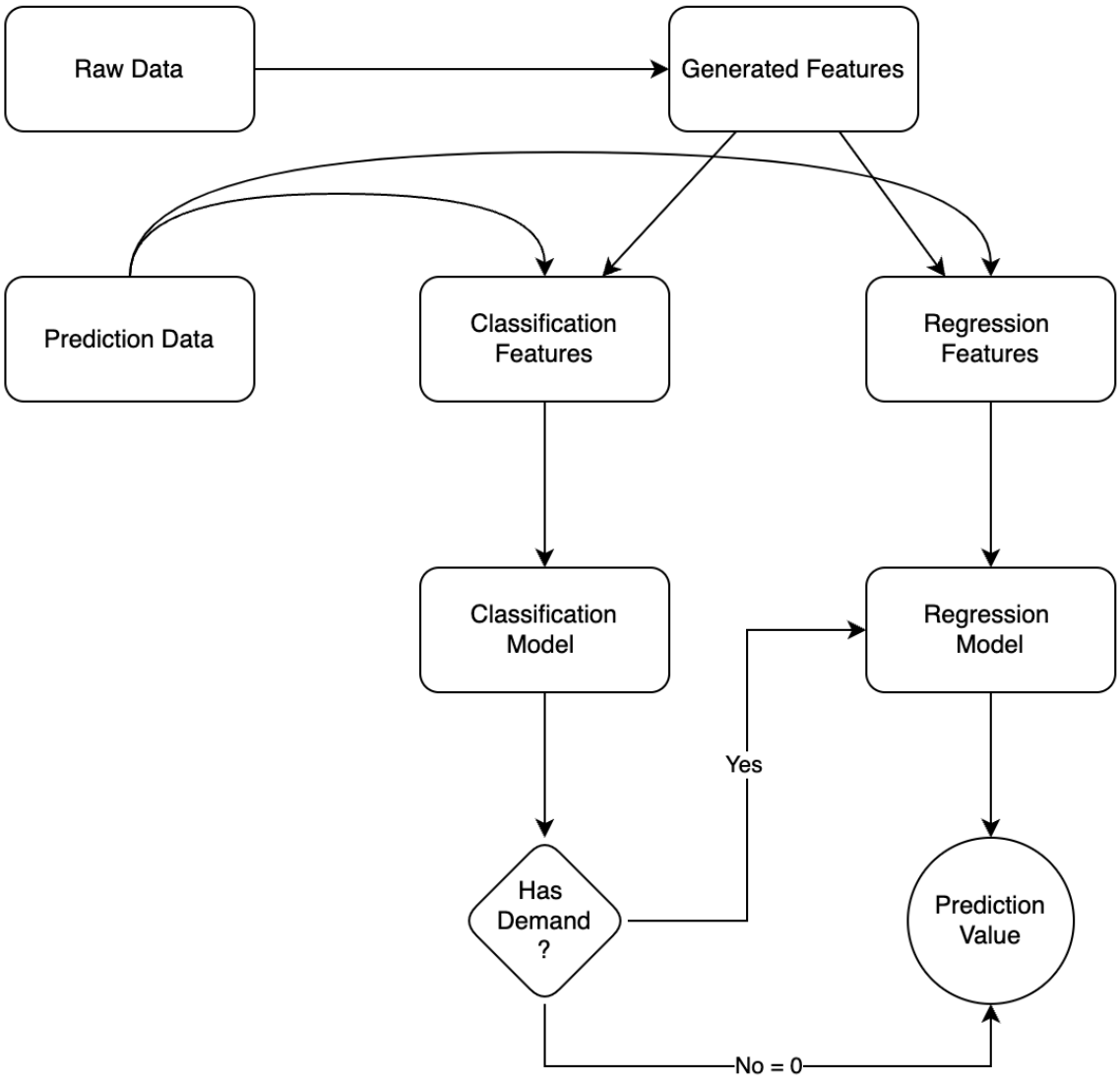

Predicted vs. Actual ReserveQty (Scaled) - R2: 0.65
Predicted ReserveQty (Scaled)
Actual ReserveQty (Scaled)
0
200000
400000
600000
800000


material_chargingcc_max_reserve_qty_lag2
material_max_reserve_qty_lag0
material_chargingcc_average_reserve_qty_lag0
material_max_reserve_qty_lag2
material_max_reserve_qty_lag4
material_materialtypecode_max_reserve_qty_lag0
materialtypecode_chargingcc_average_reserve_qty_lag4
material_average_reserve_qty_lag0
material_chargingcc_max_reserve_qty_lag4
ReserveQty_lag4
materialtypecode_chargingcc_average_reserve_qty_lag3
material_max_reserve_qty_lag6
material_max_reserve_qty_lag1
material_max_reserve_qty_lag5
material_chargingcc_median_reserve_qty_lag0
material_median_reserve_qty_lag0
material_materialtypecode_max_reserve_qty_lag2
ReserveQty_lag1
ReserveQty_lag6
materialtypecode_chargingcc_std_reserve_qty_lag1
−1
0
1
SHAP value (impact on model output)
High
Low
Feature value

4. Prediction Pipeline Overview

The pipeline for predicting material demand follows a structured process:

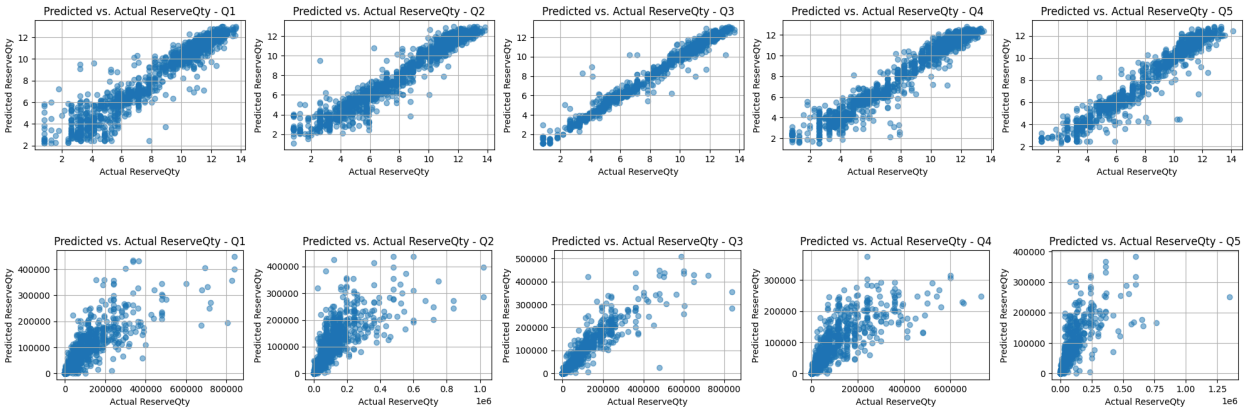
1. Input data is combined with the extracted feature set.
2. The **classification model** first determines whether there will be demand for the next month.
3. If no demand is predicted, the system returns a prediction of **0**.
4. If demand is predicted, the **regression model** is used to estimate the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



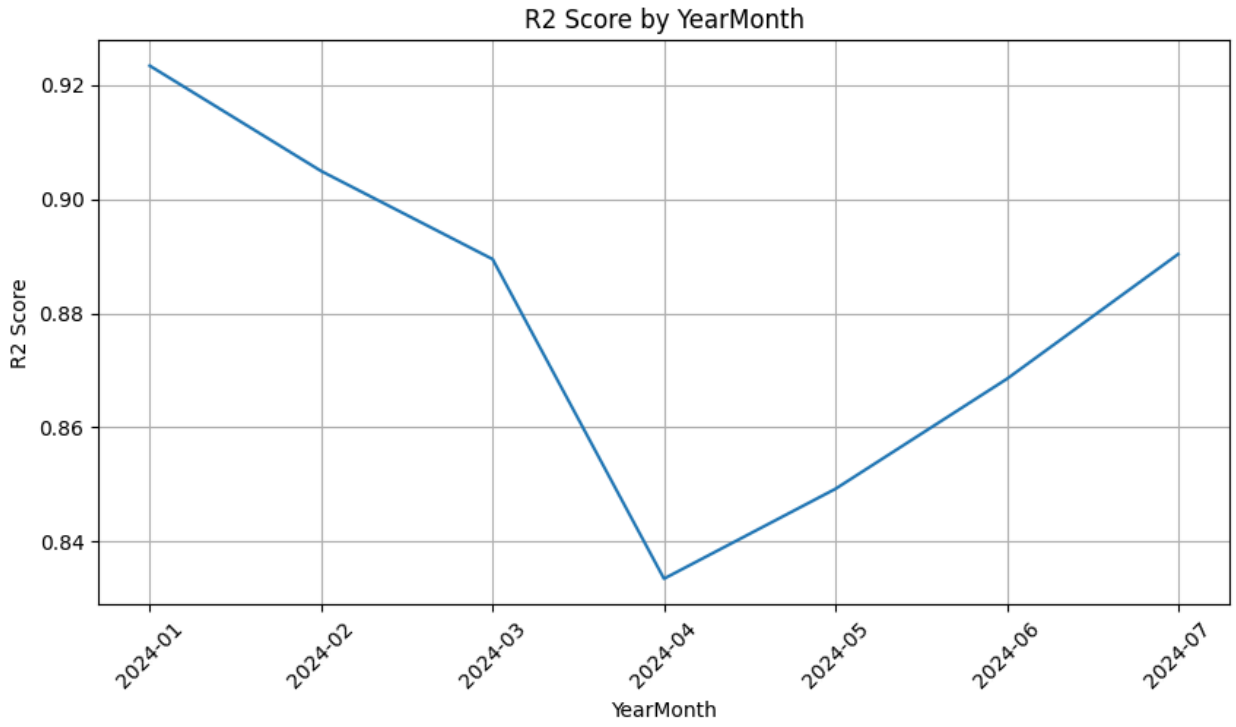
5. Model Training, Testing, and Evaluation

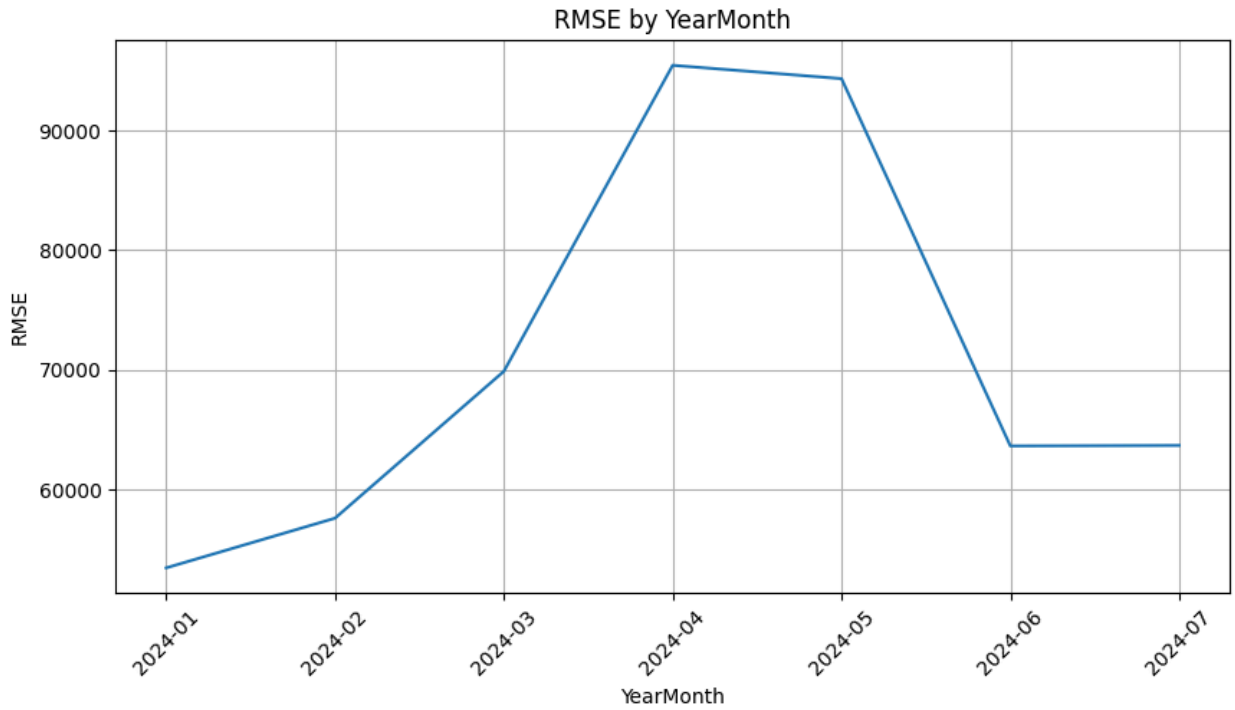
- **Training Phase**: The models were trained using data from **12 months** of 2023.
- **Testing Phase**: The trained models were tested on data spanning from **January to July 2024**.

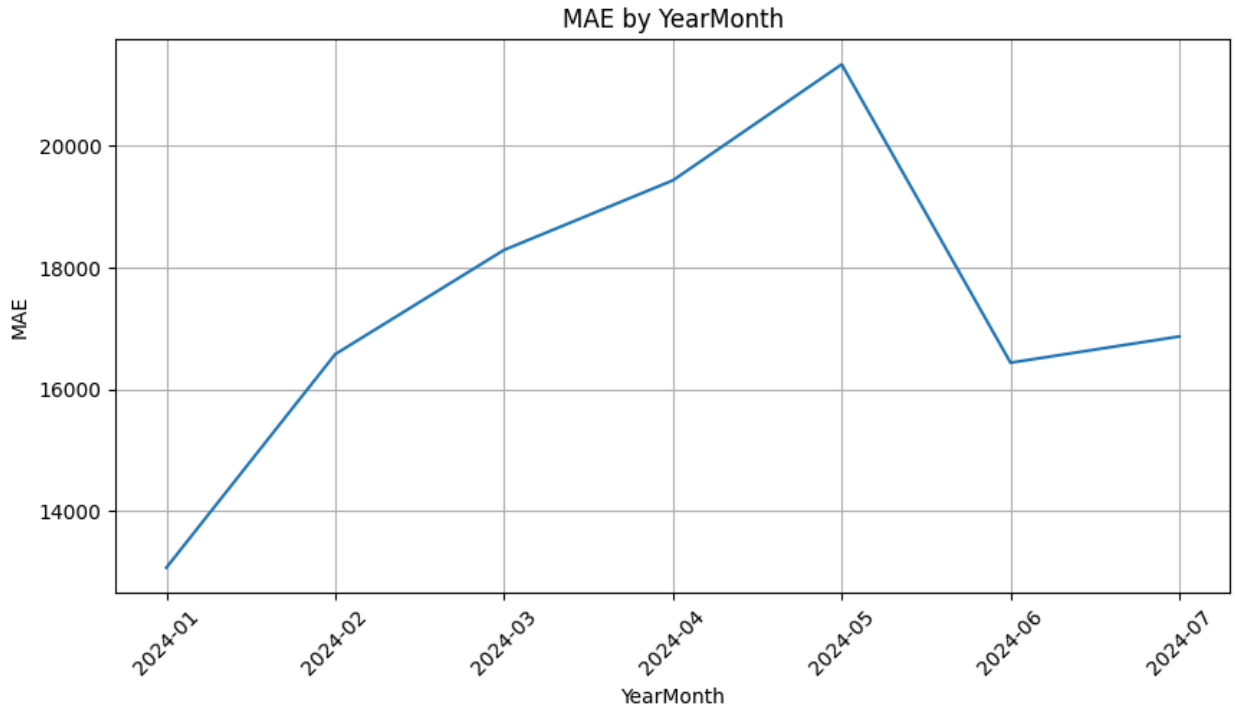
Evaluation results show:



- **R^2** 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.

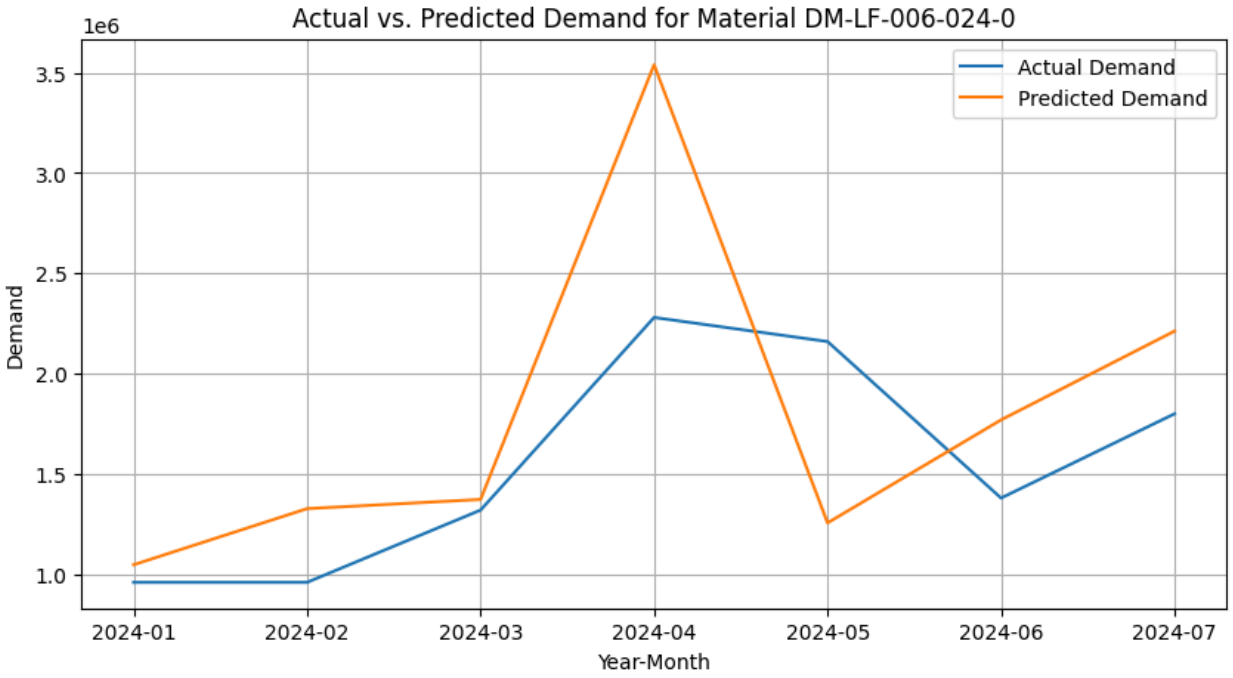


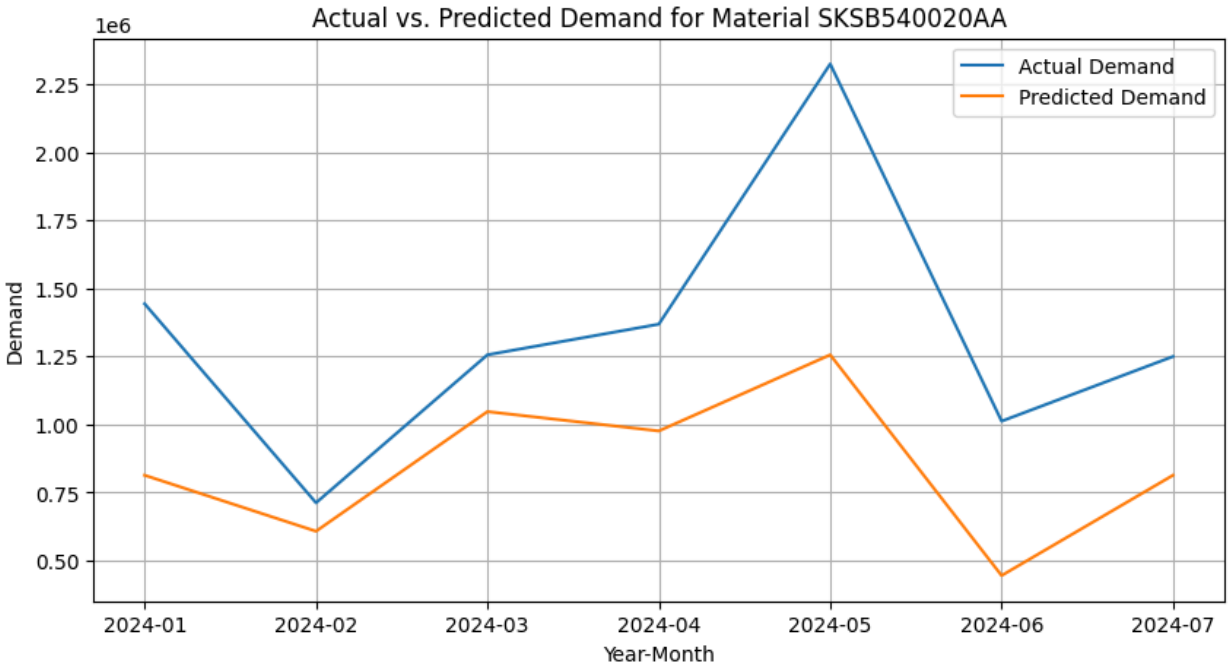

RMSE by YearMonth
RMSE
90000
80000
70000
60000
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
YearMonth

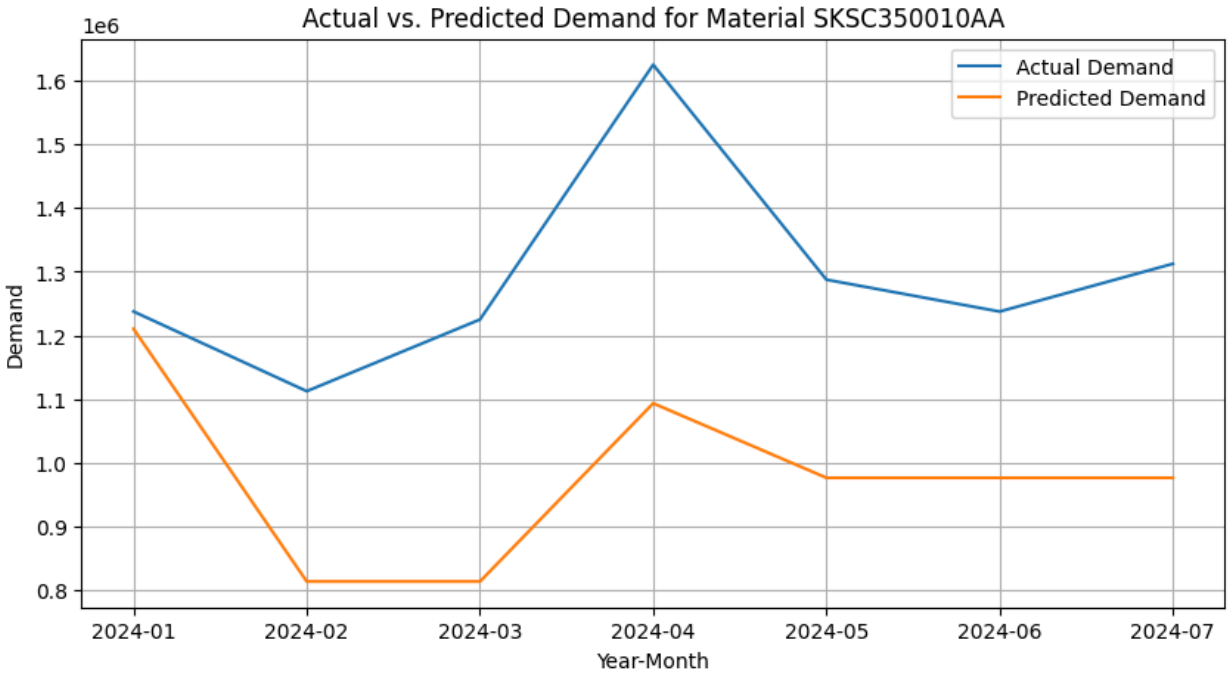

MAE by YearMonth
MAE
20000
18000
16000
14000
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
YearMonth

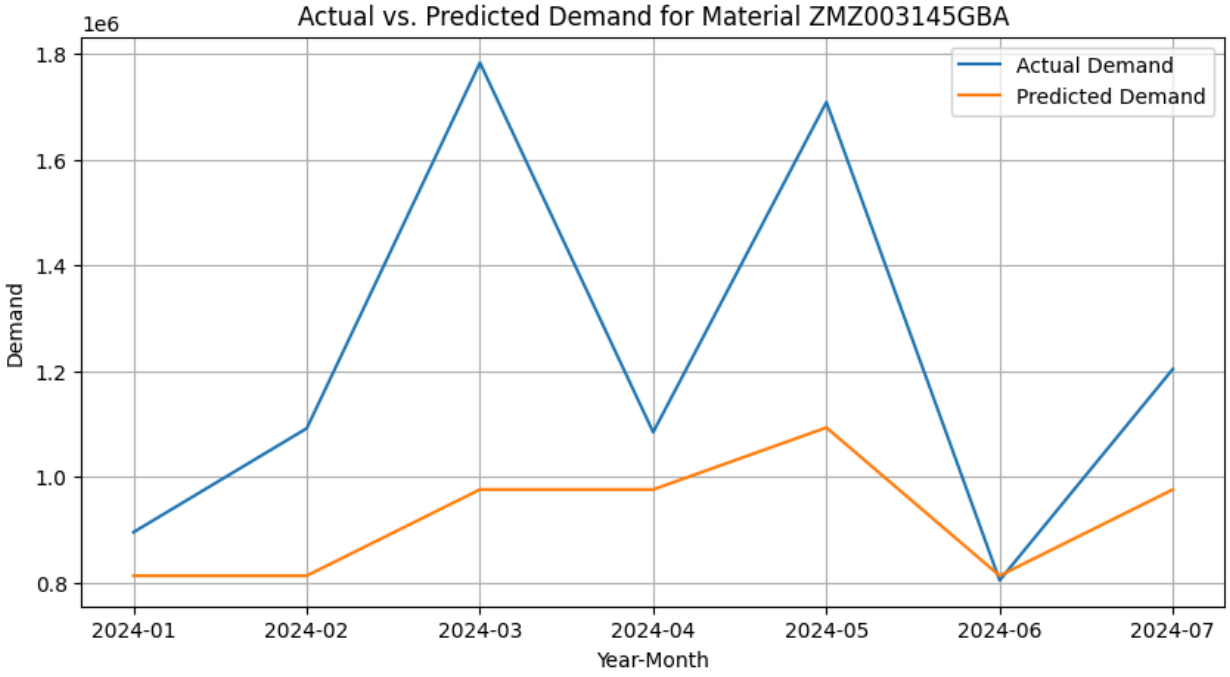
6. Top Demand Prediction

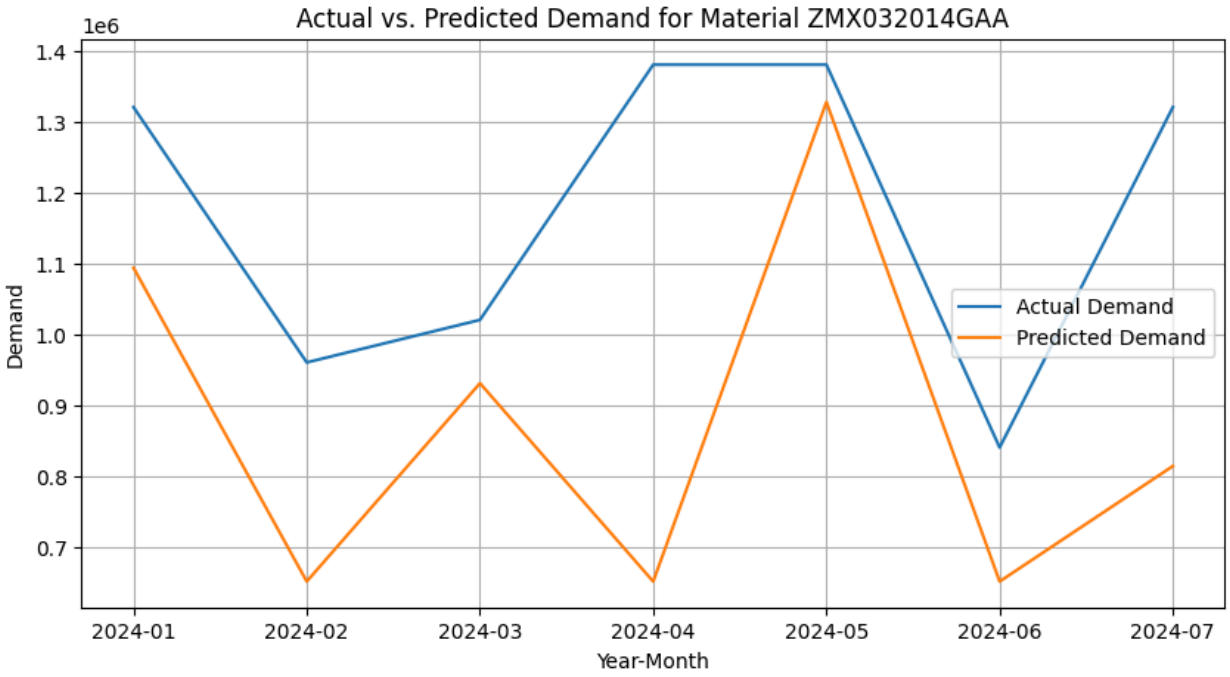
- Top 10 Most Demand Materials

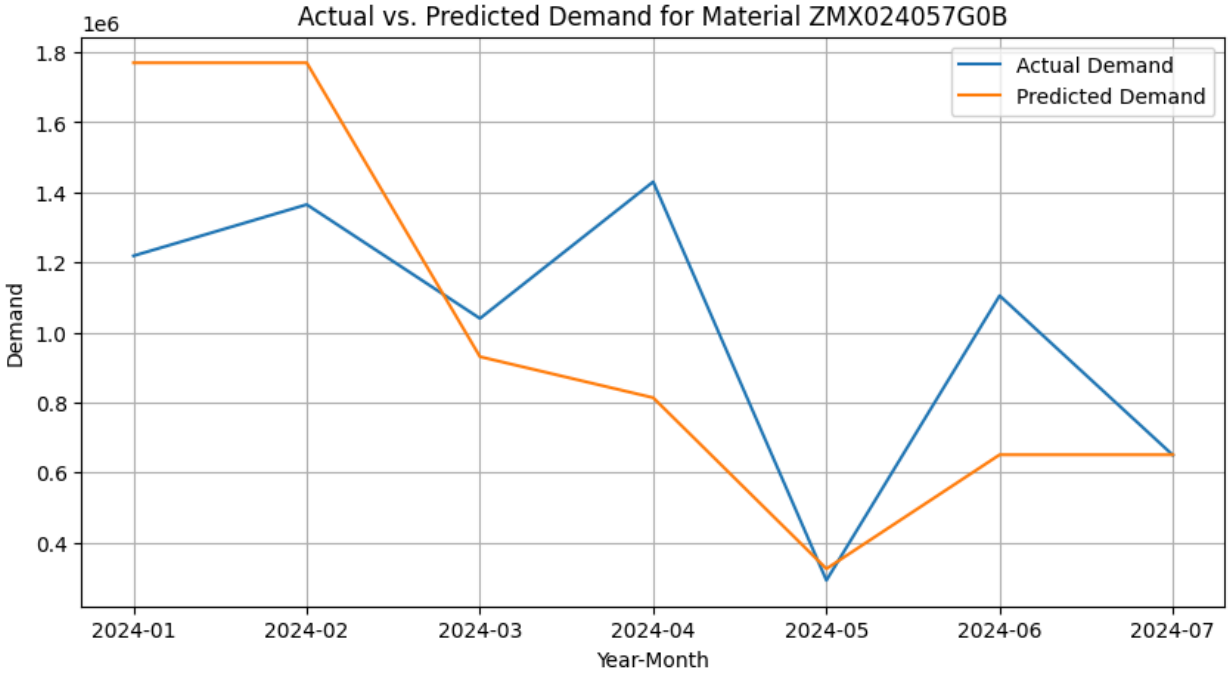


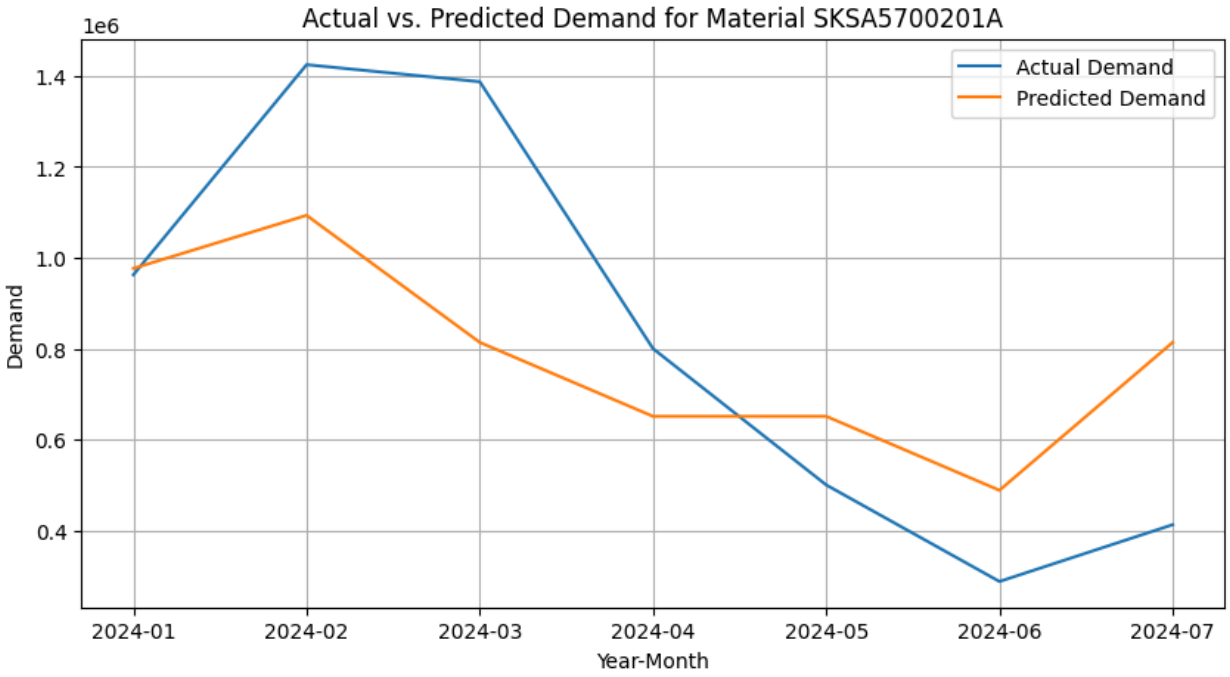


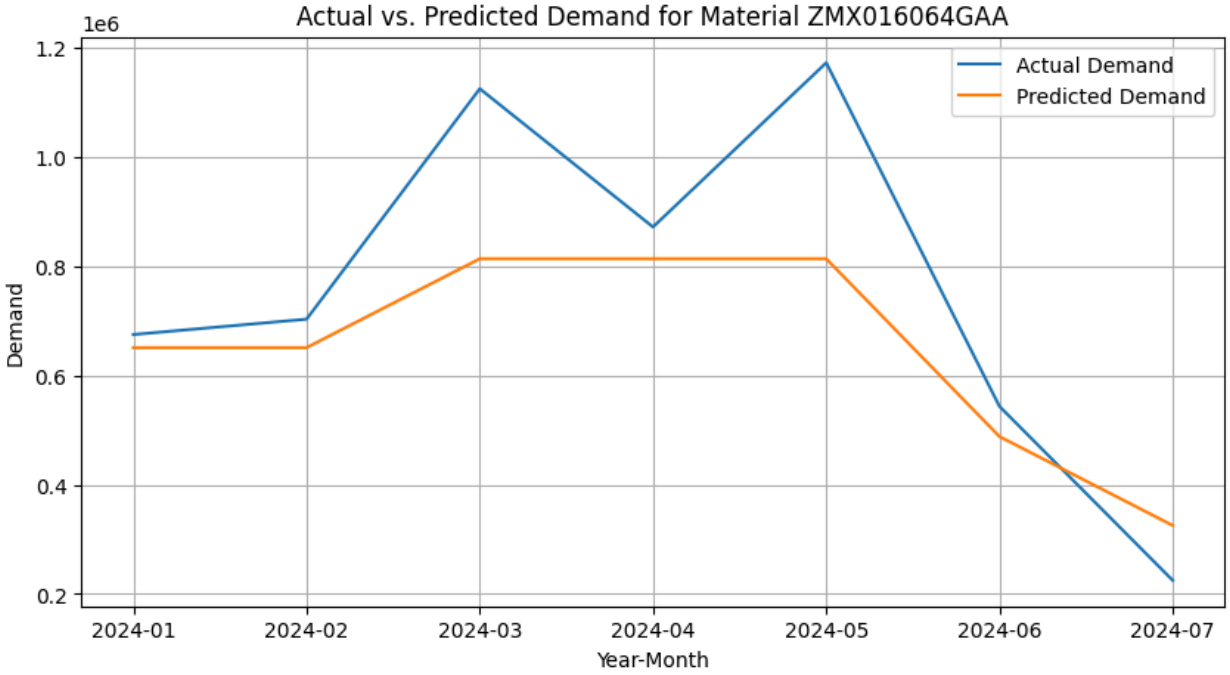

Actual vs. Predicted Demand for Material SKSC350010AA
1e6
Actual Demand
Predicted Demand
Demand
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month

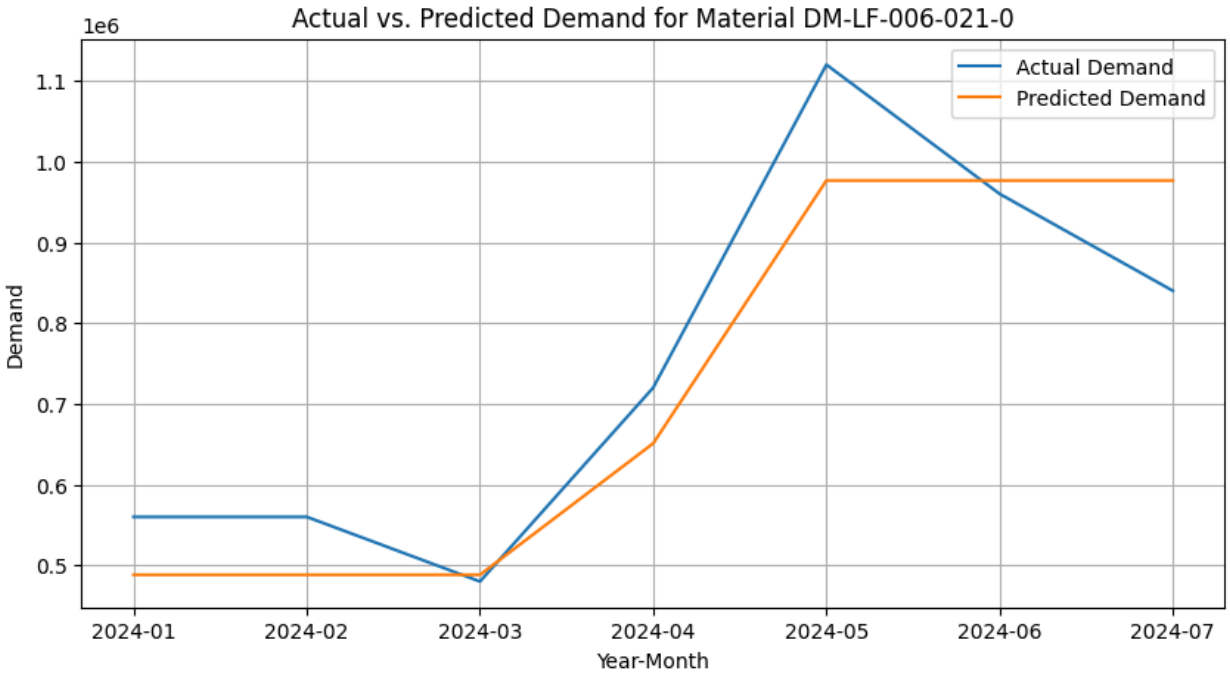

Actual vs. Predicted Demand for Material ZMZ003145GBA
1e6
Actual Demand
Predicted Demand
Demand
1.8
1.6
1.4
1.2
1.0
0.8
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month

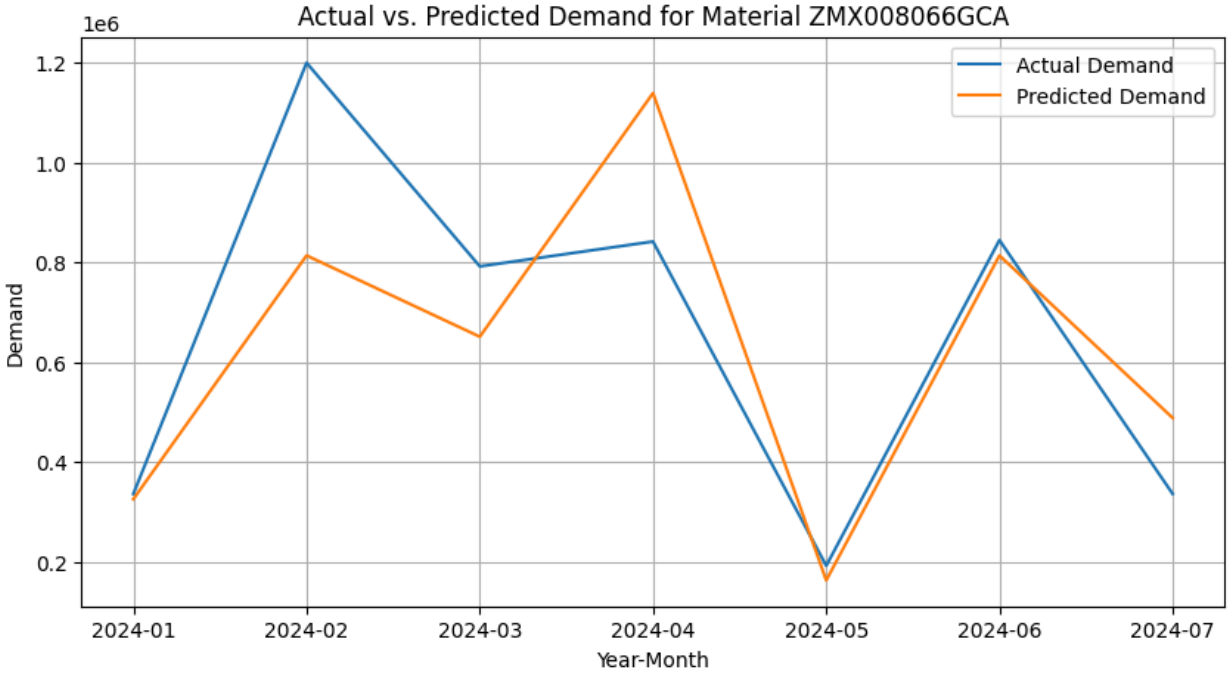

Actual vs. Predicted Demand for Material ZMX032014GAA
1e6
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
Demand
Actual Demand
Predicted Demand
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month


Actual vs. Predicted Demand for Material ZMX024057G0B
1e6
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Demand
Actual Demand
Predicted Demand
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month


Actual vs. Predicted Demand for Material SKSA5700201A
1e6
Actual Demand
Predicted Demand
Demand
1.4
1.2
1.0
0.8
0.6
0.4
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month


Actual vs. Predicted Demand for Material ZMX016064GAA
1e6
Actual Demand
Predicted Demand
Demand
1.2
1.0
0.8
0.6
0.4
0.2
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month


Actual vs. Predicted Demand for Material DM-LF-006-021-0
1e6
1.1
1.0
0.9
0.8
0.7
0.6
0.5
Demand
Actual Demand
Predicted Demand
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month


Actual vs. Predicted Demand for Material ZMX008066GCA
1e6
1.2
1.0
0.8
0.6
0.4
0.2
Demand
Actual Demand
Predicted Demand
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
Year-Month