ESTIMATING STOCK KEEPING UNIT USING ML

Model Selection Report

Goal:

To select the best-performing model for SKU-level demand prediction, based on validation metrics and deployment suitability.

Models Compared:

- Random Forest Regressor
- XGBoost Regressor

Comparison Metrics:

Model	R² Score	MAE
Random Forest	0.8169	13.63
XGBoost Regressor	0.8267	13.61

Qualitative Comparison:

✓ **Accuracy**: XGBoost showed higher R² and slightly lower MAE.

✓ **Speed**: XGBoost trained faster due to efficient boosting structure.

✓ **Scalability**: XGBoost supports distributed learning and handles large datasets well.

✓ **Robustness**: Both models handled outliers and variable importance effectively.

Final Model Chosen: XGBoost Regressor

Reasons for Selection:

● Best overall test set performance

● Slight improvement in MAE can significantly affect operational decisions in retail

● Supports regularization, which helps prevent overfitting

● Easier integration into lightweight deployment environments like Flask

Limitations:

● Slightly more complex to tune compared to Random Forest

● Requires careful feature scaling in certain use cases (not needed here)

Deployment Readiness:

● Model serialized using `pickle`

● Fully integrated with Flask backend

● Accepts 4 input fields and computes all required features for real-time prediction

This report confirms that the XGBoost model meets the project goals in accuracy, efficiency, and production-readiness.