Anomaly Detection for Product Backorder Prediction

- **Objective:** Developed a system to proactively identify products at risk of going on backorder, shifting the company from reactive firefighting to proactive inventory planning. The goal was to reduce stock-outs, which hurt customer satisfaction and increase costs, without overstocking products.

- **Methodology:** Developed and benchmarked three distinct machine learning pipelines to identify the optimal model for predicting backorders from a dataset of ~1.9 million product observations. The best-performing pipeline consisted of a three-step process:

    1. StandardScaler: Scaled all 22 predictor variables to ensure equal weighting.

    2. PCA (Principal Component Analysis): Reduced feature dimensionality to 15 key components to increase training speed and efficiency.

    3. Random Forest Classifier: Used an ensemble of 400 decision trees (max depth of 20) to vote on the backorder risk.

- **Results & Impact:** The final model, selected after a comprehensive hyperparameter search across all tested pipelines, performed with high accuracy on unseen data. This system allows planners to focus on a targeted list of at-risk items instead of manually monitoring all SKUs.

    o Backorders Caught (Recall): 74%

    o False Alarms (False Positives): 18%

    o Overall Accuracy: 82%