

Problem Statement: Predicting Procurement Spend Efficiency and Anomalies

Objective:

Develop a machine learning model to predict whether a purchase order is efficient or exhibits anomalies based on historical procurement data. The goal is to help the procurement team identify areas for cost savings, improve strategic planning, and enhance overall procurement efficiency.

Target Variable:

Efficiency_Label: A binary variable indicating whether a purchase order is efficient (0) or exhibits anomalies (1). This needs to be defined based on specific criteria such as cost overruns, delays, and deviations from expected spending patterns.

[Dataset](#)

Steps to Solve the Problem:

1. Data Preprocessing:

- Handle missing values appropriately.
- Convert categorical variables to numerical using techniques like one-hot encoding.
- Normalise/standardise numerical features.

2. Feature Engineering:

- Create new features based on domain knowledge (e.g., cost per unit, delivery time, number of changes to a purchase order).
- Aggregate features to summarise information (e.g., average spend per supplier, average delivery time).

3. Exploratory Data Analysis (EDA):

- Analyse the distribution of the target variable.
- Identify correlations between features and the target variable.
- Visualise important features using plots.

4. Modelling:

- Split the data into training and testing sets.
- Train different classification models (e.g., Logistic Regression, Decision Trees, Random Forest, Gradient Boosting).
- Evaluate models using appropriate metrics (e.g., accuracy, precision, recall, F1-score).

5. Model Selection and Tuning:

- Use cross-validation to select the best model.
- Tune hyperparameters using techniques like Grid Search or Random Search.

6. Model Evaluation:

- Evaluate the final model on the test set.
- Analyse feature importance to understand which factors contribute most to inefficiencies.

7. Deployment:

- Deploy the model for real-time predictions.
- Set up a monitoring system to track model performance over time.