Predicting Procurement Compliance Using KPI-Driven Machine Learning Models

Brittany M.D. Dowdle

Northwest Missouri State University
School of Computer Science and Information Systems
Maryville MO 64468, USA
S574362@nwmissouri.edu

Abstract. Keywords: procurement · compliance · prediction · product

1 Introduction

This project explores procurement performance using real-world data; the goal is to predict whether a purchase order will result in supplier compliance. Order compliance is critical in performance metrics. It reflects whether supplier contracts are meeting agreed-upon delivery schedules, product quality, and pricing terms. By building a predictive model based on features such as quantity, price deviation, or defect rates, the objective is to provide procurement teams with data-driven insights. This could predict risk, improve supplier relationships, and enhance operational efficiency.

1.1 Goal of This Project

The final deliverables from this project will include:

– Machine learning model predicting order compliance (including data visualizations [4])
– A PDF report written in LaTeX via Overleaf
– A fully documented GitHub repository with Jupyter notebooks (including EDA [4])

1.2 Project Resources

– **GitHub Repository:** https://github.com/Bdowdle4/Dowdle$_A$nalytics$_C$apstone
– **Overleaf Report:** https://www.overleaf.com/read/bszyhdxsnrsf
– **Pro Analytics 01:** https://github.com/denisecase/pro-analytics-01
  The guide used to follow a repeatable workflow for professional python projects [1]

2   Collect and Describe Data

The data set analyzed consists of 777 purchase order records with 11 columns. Each row in the data set reflects a unique transaction with attributes that relate to the identity of the supplier, the characteristics of the order, the pricing, the defects, and whether or not the supplier was compliant. The data set was publicly accessible on Kaggle; the link is in Section 2.2 Dataset Resource.

2.1   Dataset Overview

- **Data Type:** Structured (Avg of 10 key features per PO)
- **Source:** Kaggle
- **Size:** 68 KB
- **Rows:** 777
- **Columns:** 11
- **File Extension:** `.csv`
- **Tool for Ingestion:** `pandas` in Python

The data set was downloaded from Kaggle [2] and moved to the project folder. Then it was added to my project repository in the "Data" folder. Finally, it was read into two Jupyter notebooks using the pandas library (Cleaning and Modeling).

2.2   Data Attribute Dictionary

Column Name	Description	Data Type	Example
po_id	Unique identifier for the purchase order	String	PO-10231
supplier	Supplier name or identifier	String	Supplier_A
order_date	Date the order was placed	Date	2024-01-03
delivery_date	Date the order was delivered	Date	2024-01-11
item_category	Category or type of item ordered	String	Raw Materials
order_status	Status of the order (e.g., Delivered, Pending)	String	Delivered
quantity	Quantity of units ordered	Integer	500
unit_price	Price per unit paid	Float	12.75
negotiated_price	Contractually agreed price per unit	Float	12.00
defective_units	Number of defective units in the delivery	Integer	5
compliance	Binary indicator of contract compliance	Integer	1

Table 1. Key data attributes and examples.

2.3   Domain and Professional Description

Domain: Business Operations
Subdomain: Procurement / Supply Chain

This project would be important to:

1. Supply Chain Analysts - to identify patterns in KPI metrics
2. Procurement Managers - for supplier scorecards and vendor decisions
3. Chief Purchasing Officer (CPO) - to support strategic sourcing and policy decisions

This data set falls within the field of procurement analytics. Procurement professionals use analytics to track KPIs such as on-time deliveries, cost savings, and defect rates to measure supplier performance. When a supplier consistently does not meet the expected performance level, they are a financial and operational risk. Suppliers can be considered non-compliant for late deliveries, defective products, and violations of negotiated pricing. By identifying patterns in procurement data and predicting compliance outcomes, organizations can optimize sourcing strategies, negotiate better contracts, and reduce supply-side risk. [3]

2.4   Dataset Resource

Procurement KPI Analysis Dataset:
https://www.kaggle.com/datasets/shahriarkabir/procurement-kpi-analysis-dataset

This data set is anonymized to protect company and supplier identities. It provides real-world transactions of 5 different suppliers from 2022-2023. This data reflects challenges like supplier delays, compliance gaps, defects, and inflationary price trends over time. It is not expected to be updated. [2]

3   Clean and Prepare Data

3.1   Data Formats

3.2   Null Handling

3.3   Outliers

3.4   Jupyter Notebook

4   Exploratory Data Analysis (EDA)

4.1   Trend Line Charts

4.2   Correlation Matrix

4.3   Vendor Comparisons

5   Model and Generate Insights

5.1   Chosen Models

5.2   Parameters

5.3   Metrics

6   Present Results

6.1   Summarize KPIs

6.2   Plot Model Predictions

6.3   Maybe Dashboard Style Visuals?

7   Finalize Deliverables

7.1   GitHub Repo Completeness

7.2   Overleaf Report Completeness

7.3   Limitations

7.4   Future Work

8   Giving Credit Where It's Due

References

1. Case, D.: Git add, commit, push guide. bluehttps://github.com/denisecase/pro-analytics-01/blob/main/03-repeatable-workflow/06-git-add-commit-push.md (2023), accessed: 2025-06-29

2. Kabir, S.: Procurement kpi analysis dataset. bluehttps://www.kaggle.com/datasets/shahriarkabir/procurement-kpi-analysis-dataset (2022), accessed: 2025-06-23
3. Team, T.E.: Why procurement is a good career: Key reasons to consider this path. bluehttps://www.techneeds.com/2025/04/05/why-procurement-is-a-good-career-key-reasons-to-consider-this-path/ (April 2025), accessed: 2025-06-29
4. Team, U.E.: Top 100 data science project ideas for beginners. bluehttps://www.upgrad.com/blog/data-science-project-ideas-topics-beginners/ (2024), accessed: 2025-06-23