Enhancing Supply Chain Efficiency With Predictive Supplier Modeling

Purdue University, Daniels School of Business

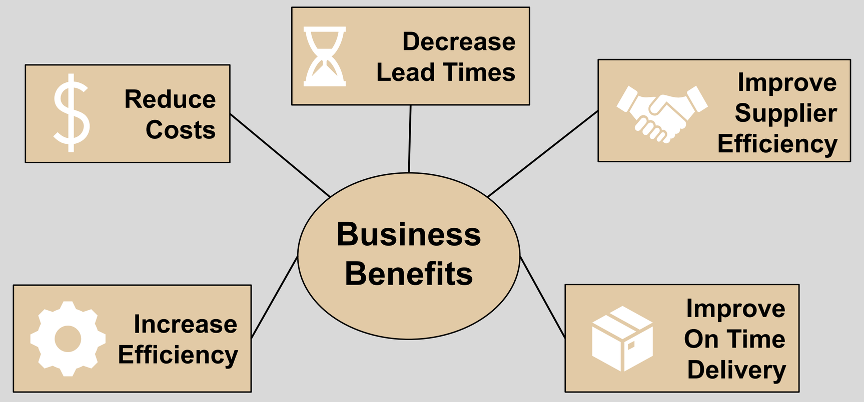
Wentong Zhang zhan4721@purdue.edu | **Christos Kaiafas** ckaiafas@purdue.edu | **Ace Setiawan** setiawa@purdue.edu | **Daniel Gallagher** gallag97@purdue.edu | **Rileigh Dethy** rdethy@purdue.edu | **Kailey Leuer** kleuer@purdue.edu | **Xing Wang** wang5719@purdue.edu | **Wenyong Xu** xu1443@purdue.edu

BUSINESS PROBLEM

Business Problem: The company is currently facing high lead times and variability in supplier performance, leading to inefficiencies and increased costs in the supply chain

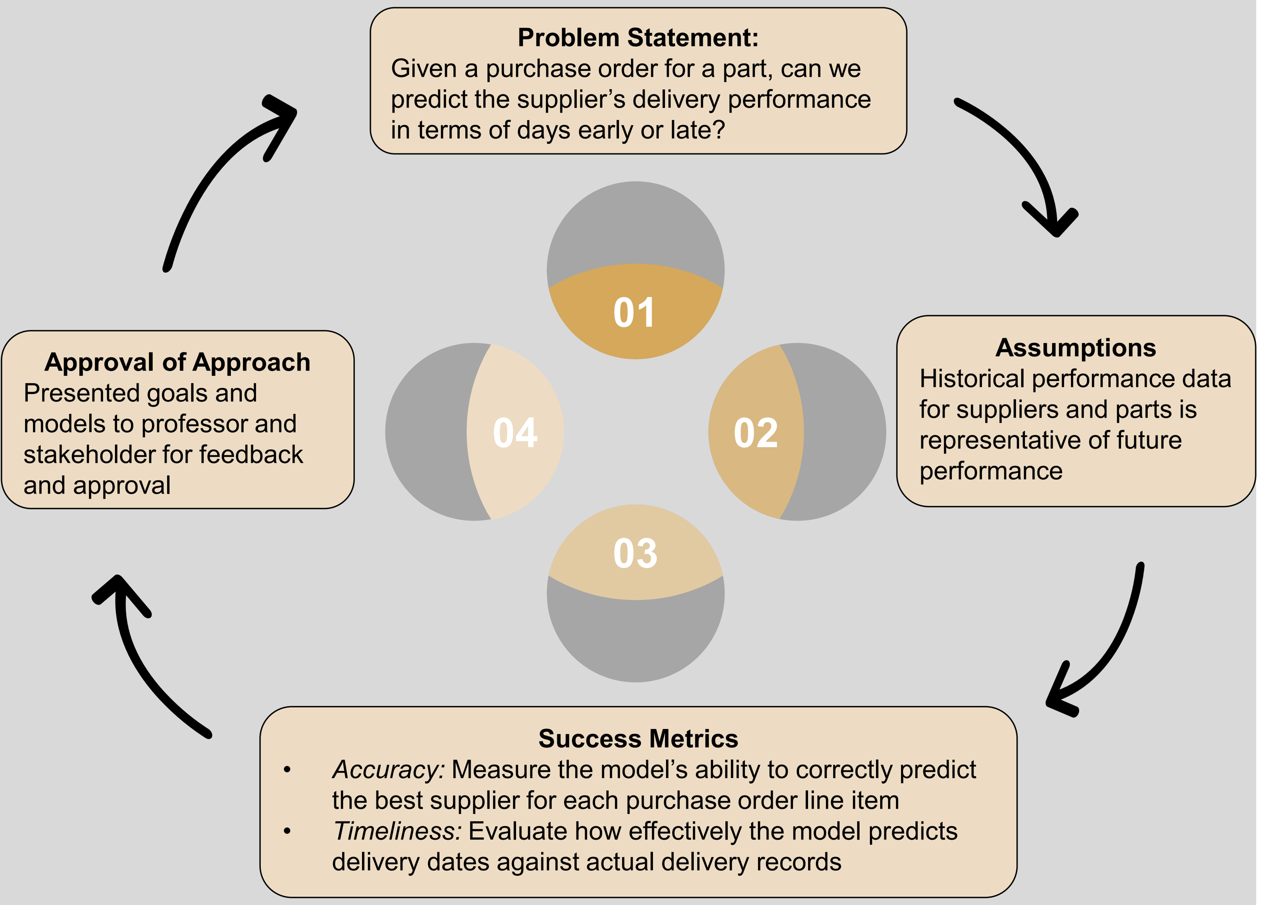
Importance/Motivation: Improving the supplier selection process can lead to more reliable delivery schedules, reduced costs, and increased overall business efficiency

Audience/Stakeholder: Manufacturing Company



ANALYTICAL FRAMING

Analytical Goal: Develop a predictive model to enhance supplier selection for each part order based on historical performance metrics and clustering analysis.



DATA

Data Source: The data comes from the Company and contains procurement and supply chain information, including purchase order details, supplier information, engineering requirements, and delivery dates (see sample below)
- **25,000+** Rows of Data
- Date Range: 11/17/2015 – 05/29/2024

Data Preprocessing:
- ***Data Selection:*** Removed redundant and repetitive columns
- ***Missing Values:*** Imputed missing numeric values; removed highly sparse columns
- ***Duplicate Handling:*** Verified no duplicates existed
- ***Data Type Standardization:*** Unified date formats, numeric values, and categorical variables
- ***Added Columns:*** Feature engineering historical metrics

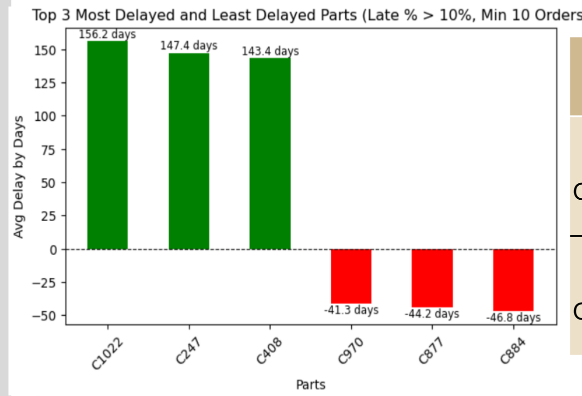
PO Number	PO Line Number	PO Amount USD	PO Qty	Unit Price USD	Supplier	Part Number	PO Placed Date	Promised Date	Actual Delivery Date
P10788	10	12445.22	1	12445.22	S110	C104	12/5/2022	12/3/2024	12/17/2024
P12469	30	23955.59	1	23955.59	S342	C819	6/8/2023	12/18/2023	1/1/2024

DATA UNDERSTANDING

Parts: 1,308 Unique Part Numbers
Suppliers: 527 Unique Suppliers

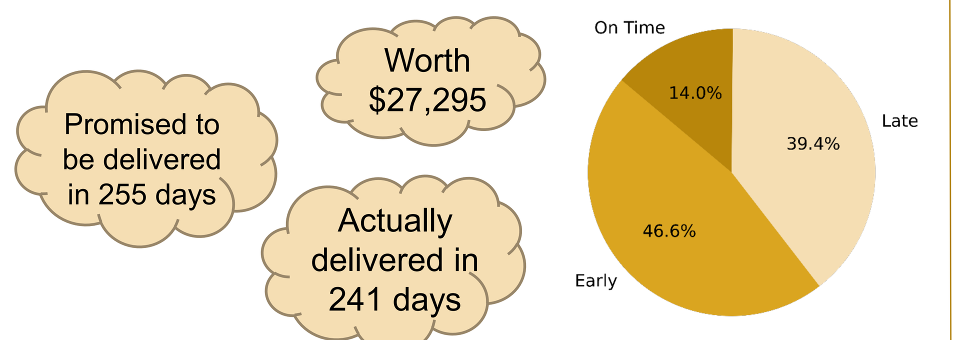
Parts & Orders

Problematic Parts (high % of early/late orders, high order frequency)

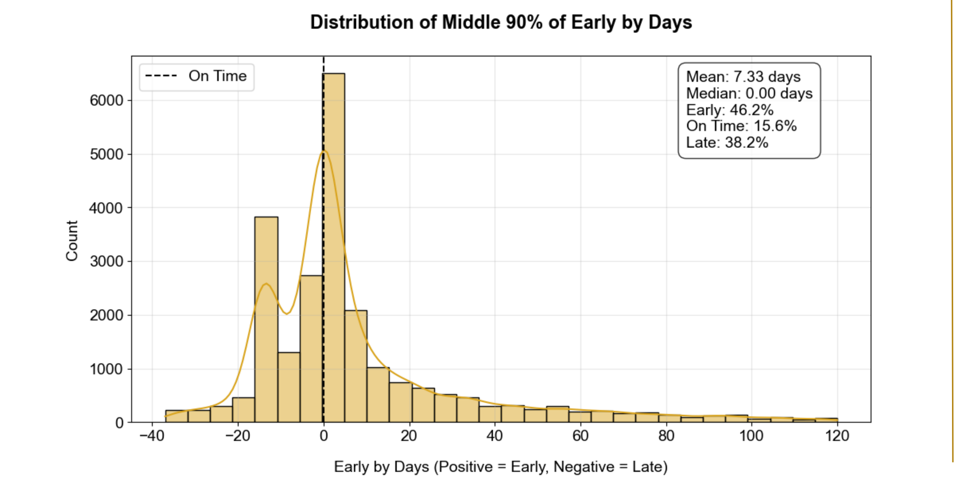


	Part	Total Orders	% of Orders Early/Late
Early Orders	C1022	65	24.62%
	C247	11	18.18%
	C408	11	27.27%
Late Orders	C970	18	66.67%
	C877	35	74.29%
	C884	191	80.63%

On average, an order is:



But distribution of timeliness shows it is centered around on-time, with **17,060** orders within 2 weeks of being on time.

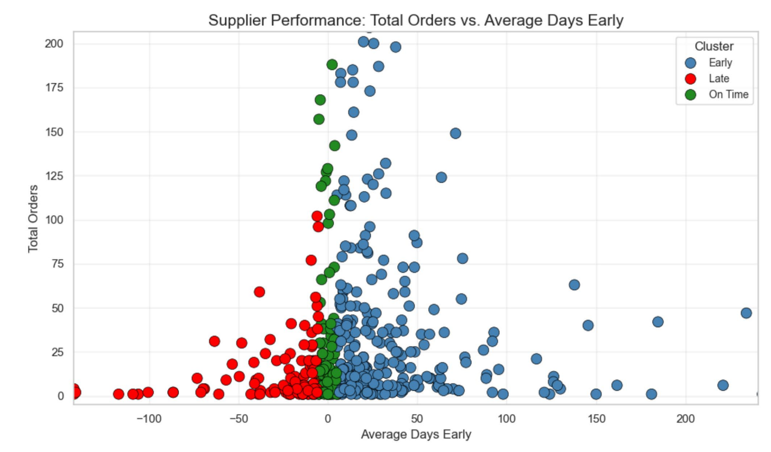


Suppliers

136 suppliers are on average at least 75% early/on-time

283 suppliers ordered from 10+ times

Supplier Clusters:



**Early = Promised Delivery Date – Actual Delivery Date

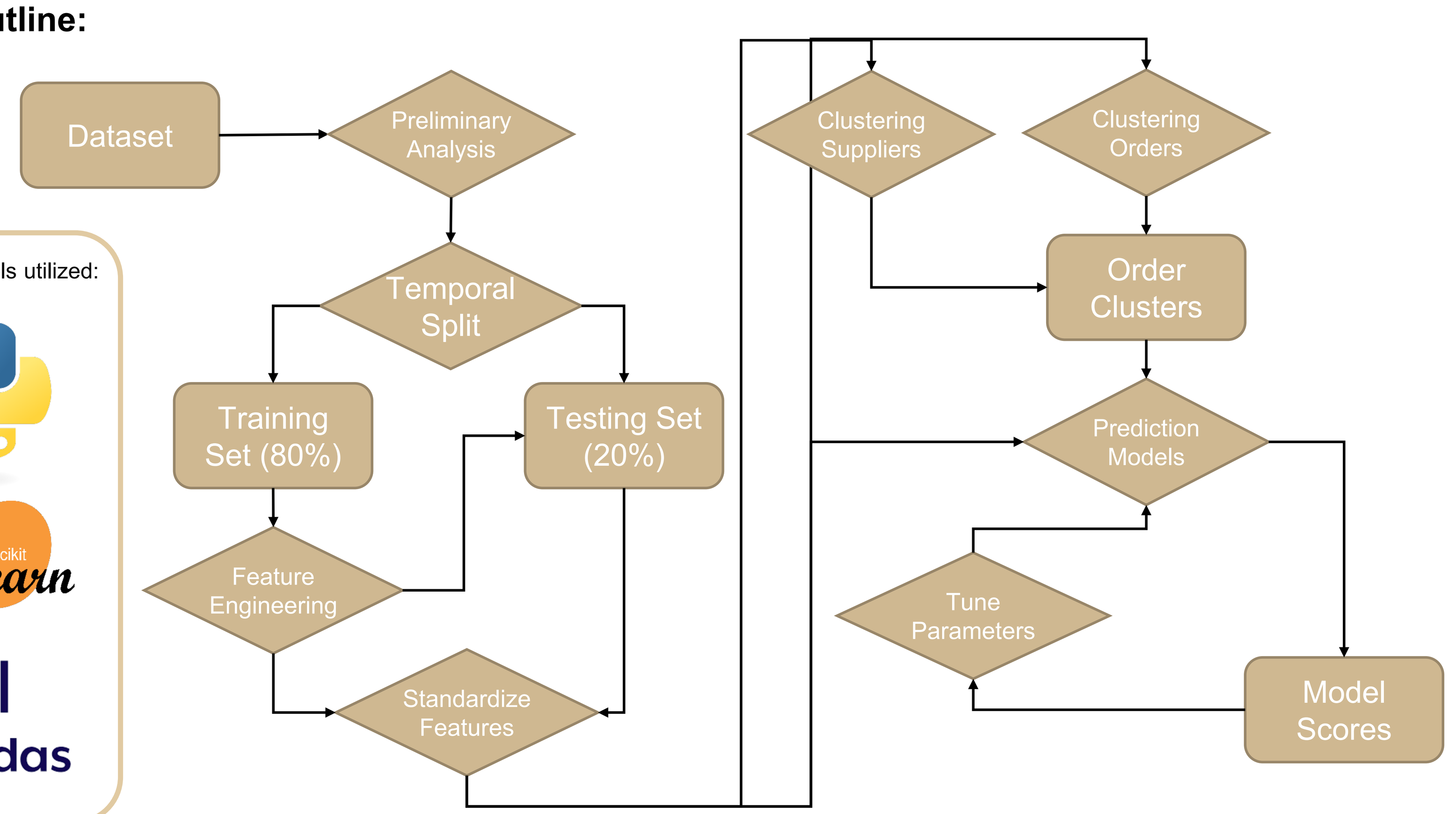
Cluster Information:

Cluster	Conditions	Mean Days Early	# Suppliers
0: On Time	± 5 Avg Days Early	-0.29	113
1: Early	Avg Days Early > 5	41.82	290
2: Late	Avg Days Early < -5	-34.13	124

PURDUE UNIVERSITY | Mitch Daniels School of Business

METHODOLOGY

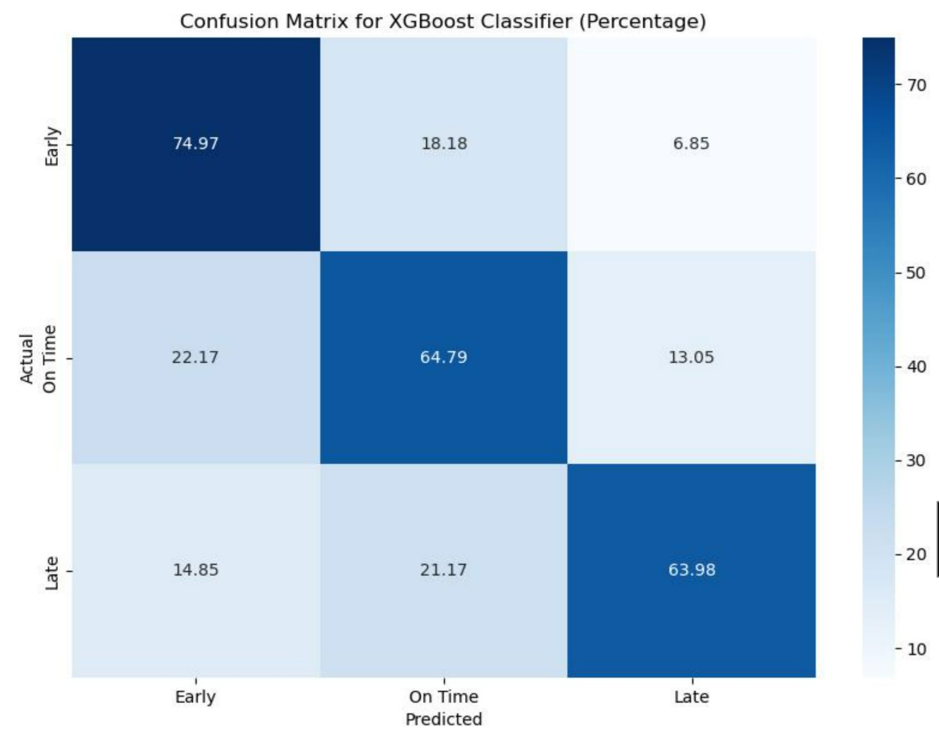
Process Outline:



MODEL BUILDING

1. **Classification**
   - Split parts into 3 classes
     - Training supervised models to classify parts on likelihood to be late/early/on-time
   - Selecting Model
     - Compare accuracy scores and determine best parameters, XGBoost performs best

Class	Conditions
On Time	Days Early +-5
Early	Days Early >5
Late	Days Early <-5



Model	Early Class Accuracy	OnTime Class Accuracy	Late Class Accuracy
Random Forest	61.99%	42.05%	10.44%
Gradient Bossting	65.31%	36.81%	8.81%
XGBoost	74.97%	64.79%	63.98%

2. **Forecasting & Predictive Models**
   - Predicting "Early by Days"
     - The difference between promised and actual delivery date
   - Models/Algorithms Tested
     - Random Forest, XGBoost, Quantile Regression, Two-Step Model, LGBM Classifier
   - Temporal Splitting
     - Split 80/20 based on order placed date, make sure engineered features strictly uses training set

Key Features for Part Classification & Regression
1. Promised Lead Days
2. PlacedDate Days Since 2015
3. PromisedDate Days Since 2015
4. PO Amount USD
5. PartSupplier Early by Days Min
6. PartSupplier Early by Days Mean

RESULTS

Business Validation: By creating a model that successfully identifies good performing suppliers as well as predicting order timeliness, supply chain processes will become more efficient. The processes will have better information for initial planning and therefore, reduce costs and unforeseen disruptions to an already complicated chain.

Impact of Model:
- Improved supply chain reliability by reducing late deliveries and enhancing supplier performance.
- Identified top-performing suppliers to support strategic decision-making and optimize resource allocation.
- Leveraged machine learning techniques to uncover valuable data features and increase predictive efficiency.

Model Comparison

Model	Train MAE	Test MAE
Base Random Forest	25.95	30
Quantile Regression	31.93	30.27
Two Step without Supplier Cluster Codes (Classification + Gradient Boost)	12.77	12.83
Two Step with Supplier Cluster Codes (Classification + Gradient Boost + Clustering)	12.28	11.54

Order Classes	Train MAE	Test MAE
On Time: Two Step (Classification + Gradient Boost)	1.36	1.25
Early: Two Step (Classification + Gradient Boost)	25.05	25
Late: Two Step (Classification + Gradient Boost)	12.50	6.88

* The models have been pruned and retrained
** Pruning effectively reduced number of parameters and improved accuracy

CONCLUSION

- From the table above, the **two-step** model with Supplier Cluster Codes (**XGBoost** for classification and **GradientBoost** for regression) showed the best results.
- Based on the MAE of each model, we can conclude that the model performed decently well in predicting the different order classes. On average the model is **12.8** days off compared to the true value.
- By incorporating supplier information, we can further improve the model. The result justifies that with a better supplier selection, we can have a better predictions for the delivery time.

FUTURE WORK

- Explore additional features to improve the model's accuracy, such as the shipment date.
- Research for better models for better prediction.
- Build a pipeline to provide the user with concrete conclusions such as which supplier to pick for a given part number

ACKNOWLEDGEMENTS

We sincerely appreciate Professor Xing Wang, our TA Wenyong Xu, and our industry partner for providing us with this opportunity and for their invaluable guidance and support throughout the course of this project.

informs BUSINESS ANALYTICS CONFERENCE