Developing a Scalable Data and Modeling Framework for Stamping Anomaly Detection

Joshua Cai

Operations Research Center, Massachusetts Institute of Technology

Host Organization: Ford Motor Company | Faculty Advisor: Prof. Bruce Lawler

1. INTRODUCTION

- Ford's stamping process produces large volumes of sensor and process data.
- This project leverages data to detect anomalies early and reduce downtime and rework.
- Operators still rely heavily on manual checks and reactive investigation.

Build a scalable data-processing and modeling framework that transforms raw press tonnage data into actionable model artifacts for reliable anomaly detection.

BUSINESS IMPACT

- Fewer bad parts and escapes
- Faster issue discovery and response
- Improved press uptime
- Reduced manual inspection burden

2. METHODS

- Data ingestion
- Stroke validation
- Stroke normalization
- Empty-stroke filtering
- Feature engineering
- Press-specific modeling
- Real-time scoring
- Drift monitoring

3. DATA & SETUP

2	61	762	1M+
Plants	Recipes	Runs	Signals

MODELING STRATEGY

2-Step modeling approach:

- Press 1: Mahalanobis distance on features
- Presses 2-5: Isolation Forest

DATA HIERARCHY

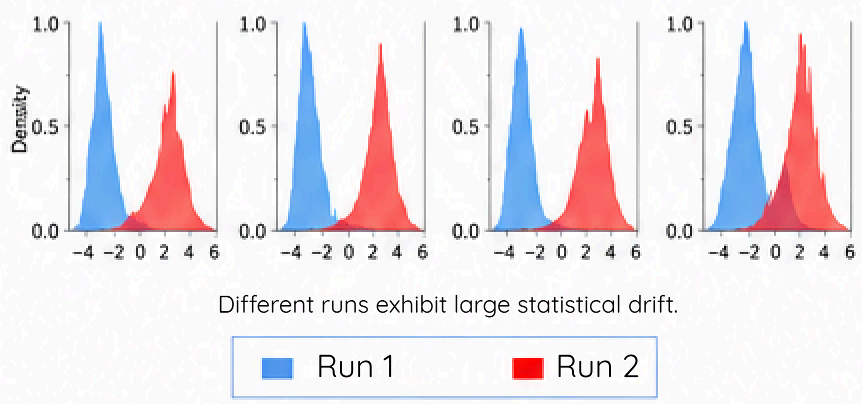
Plant → Line → Press → Recipe → Run → Cycle

RECIPE ELIGIBILITY

- ✔ Include recipes with sufficient volume, low drift, and valid signals.
- ✖ Exclude recipes with unstable sensors, ambiguous data, or high drift.

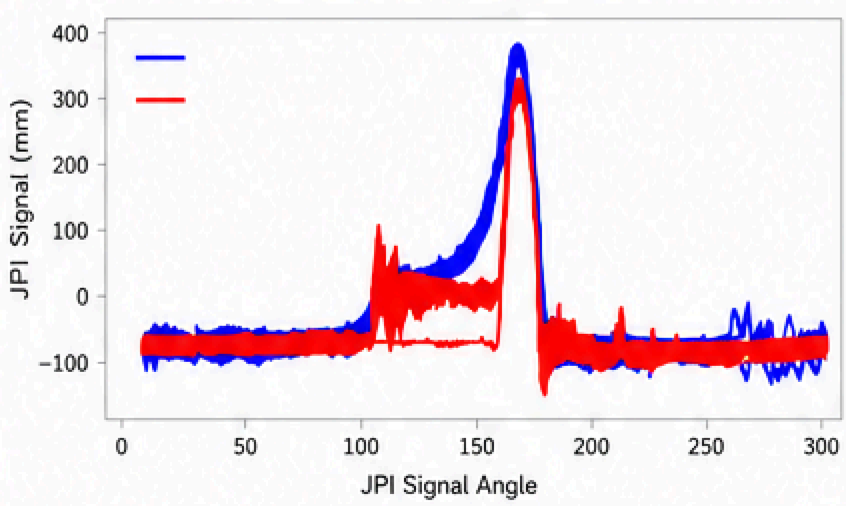
4. RESULTS

A. Distribution Shifts Across Runs



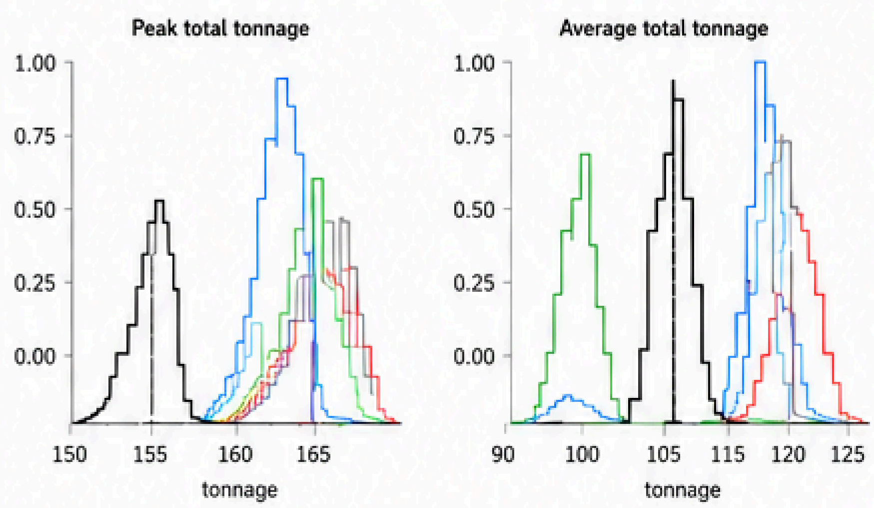
Different runs exhibit large statistical drift.

B. Low-Load vs Production Cycles



Non-production "dry-run" cycles are filtered through training.

C. Multiple Operating Regimes



A single recipe can contain multiple stable operating regimes.

SYSTEM-LEVEL OUTCOMES

Outcome	Status
Cross-plant data ingestion and standardization	Achieved
Data quality filtering and validation	Achieved
Recipe scoping and eligibility	Achieved
Model artifact generation	Achieved
Results interpretation framework	Achieved
Real-time scoring and monitoring	Achieved

- Non-production and low-load cycles must be removed before scoring.
- Multiple regimes within a run are the norm.
- Model reliability depends on clean scope and training logic.
- A scalable framework enables consistent, productive model deployment.

5. CONCLUSIONS

Scalable Foundation

Built an end-to-end framework that turns raw tonnage data into clean, model-ready artifacts.

Clear Modeling Scope

Defined which recipes and signals to model and established eligibility criteria.

Multi-Plant Extensions

Abstracted plant-specific differences through adapters and configuration.

Operational Readiness

Prepared the system for operationalization with integrated components and clear documentation.

6. FUTURE WORK

- Connect alerts with maintenance and quality outcomes.
- Validate false-alert burden in shadow mode before enabling operator notifications.
- Refine gating and extend to multitask/recipe blending.
- Expand to more plants, press lines, and operator dashboards.

7. DELIVERABLES

- Data ingestion and cleaning adapters
- Presses 2-5: validation and filtering modules
- Press 1: Mahalanobis training/scoring pipeline
- Presses 2-5: Isolation Forest training framework
- Recipe eligibility and configuration system
- Model artifact management and validation tools
- Real-time scoring service and integration components.
- Documentation, EDA notebooks

8. ACKNOWLEDGMENTS

Special thanks to MIT, Ford, and the project members and mentors, including Balpreet Singh, Al Hogan, Parinaz Farajiparvar, Brenton Hershner, Rayna Carter, Bruce Lawler, the staff at Valencia and Woodhaven stamping plants, and the rest of the Ford team, for their collaboration and domain expertise throughout the project.