

Data-driven Learning Factories: Integrating Tulip MES for enhanced operator training

Abstract. Learning Factories are essential for engineering education by replicating realistic production settings where students develop operational and decision-making skills. To reflect modern manufacturing, these environments must emphasize traceability and systematic process data analysis. In Industry 4.0 contexts, a Manufacturing Execution System (MES) enables this by integrating operators, equipment signals, and production records to provide time-stamped, end-to-end visibility across the execution layer. This foundation is critical for analyzing variability, ensuring compliance, and driving continuous improvement. This paper validates and extends a smart manufacturing pipeline within a Learning Factory by integrating enterprise planning, MES-based execution tracking, PLC-controlled stations, and collaborative robotics. Building on prior work establishing order flow and robotic assembly, this contribution focuses on how MES-enabled traceability supports human-centered analysis of operator interactions in automated and semi-automated tasks. Key performance indicators such as *task completion time*, *workflow deviations* and *operational efficiency* are captured, analyzed, and visualized to identify bottlenecks, reveal learning curves, and quantify process adherence. Industrially, results highlight MES as a practical digital transformation enabler, accelerating data integration and converting execution events into actionable insights. Educationally, embedding MES-driven traceability cultivates data-literate engineers who interpret operational data, validate improvements, and align training with digitally enabled manufacturing demands.

Keywords: Learning Factories, Manufacturing Execution System, Traceability, KPIs

1 Introduction

1.1 Context and Motivation

Learning Factories (LF) bridge the gap between theoretical instruction and industrial practice by simulating realistic manufacturing environments where students develop operational and decision-making skills. As Industry 4.0 reshapes production systems, a critical question emerges: how can LFs objectively measure and improve the human dimension of smart manufacturing training? Traditional approaches rely on subjective instructor observation, lacking the systematic data infrastructure needed to quantify operator learning curves or correlate performance with contextual variables. While industrial settings use Manufacturing Execution Systems (MES) for real-time traceability and data-driven optimization, most LFs lack equivalent execution-layer data capture, rendering

training outcomes difficult to measure systematically. This mismatch between industrial measurement capabilities and educational training environments motivates the problem addressed in this work.

1.2 Problem Statement and Objectives

The core challenge is the absence of objective measurement of operator performance at the execution layer in LFs, which prevents consistent and data-driven training assessment. In this work, operator performance is operationalized primarily as execution time, measured as step completion time and total task duration captured through Tulip. These duration based measures also support an efficiency label derived from whether an execution falls below or above the median duration. To address this gap, this paper integrates Tulip MES within a university’s LF to automatically capture execution data with *timestamps* during an Arduino Uno cable assembly training task, relate contextual variables to performance outcomes, and train machine learning (ML) models for efficiency prediction on a production line and operator classification. Tulip was selected because it supports rapid deployment of digital work instructions and structured data capture using a low-code approach that fits LF constraints while still enabling analytics and traceability.

1.3 Contribution

This work demonstrates that low-code MES platforms can provide execution-layer data capture in educational manufacturing environments without enterprise-scale complexity. It contributes a replicable MES-based data collection pipeline for LF training activities, an analytical approach to link contextual factors with performance outcomes, and baseline ML models to predict efficiency and classify operator behavior. Together, these contributions enable evidence-based, data-driven pedagogy aligned with modern manufacturing demands.

2 Related Work

MES manage production operations in real time by tracking the complete life-cycle of orders [4, 6]. In Industry 4.0 contexts, MES platforms have evolved to integrate AI and ML, generating high granularity datasets for predictive modeling and performance optimization [7]. Modern MES architectures have shifted from monolithic solutions to modular platforms that capture execution-layer data, including operator-level signals that are often not available in traditional Enterprise Resource Planning (ERP) systems [4]. In parallel, LFs provide practical environments where students develop skills through realistic system interactions [1]. Bradley et al. [2] showed that low-code and no-code platforms can democratize smart manufacturing education by enabling students without extensive software backgrounds to deploy IoT enabled MES capabilities such as inventory tracking, machine monitoring, and digital work instructions.

ML has been widely applied to industrial manufacturing challenges [9, 3], including smart reconfigurable systems [5]. However, the use of execution-layer data to support operator performance assessment in LF settings remains comparatively limited. Digital transformation work in LFs [8] has focused mainly

on technological infrastructure, with less emphasis on data-driven human performance analysis and learning assessment. Building on these foundations, our work integrates MES based data capture with ML analytics to support quantitative evaluation of operator behavior. We combine supervised learning for task duration prediction and efficiency classification with unsupervised learning for operator behavioral clustering, positioning LFs as data-driven research platforms for human-centered manufacturing training.

3 System Design and Implementation

3.1 Experimental Setup

The study was conducted at a Workstation from a Learning Factory, composed of an automated assembly cell integrating PLC controlled material handling, collaborative robotics, and HMI. Volunteer undergraduate engineering students participated as operators and performed an Arduino Uno cable assembly task guided by a Tulip MES application deployed on a tablet interface, as illustrated in Figure 1.

The data collection methodology is summarized in Figure 1a. Independent variables related to the training environment and the operator were defined for each run and linked to the corresponding execution logs. During task execution, the Tulip application enforced a stepwise workflow and recorded time-stamped completion events, which were stored in Tulip Tables as raw execution data. These logs were then exported and integrated with the operator and environmental variables, followed by cleaning and transformation to ensure consistency and readiness for analysis. Finally, the curated dataset was constructed for downstream analysis and model training based on the *captured execution times* and *associated contextual variables*.

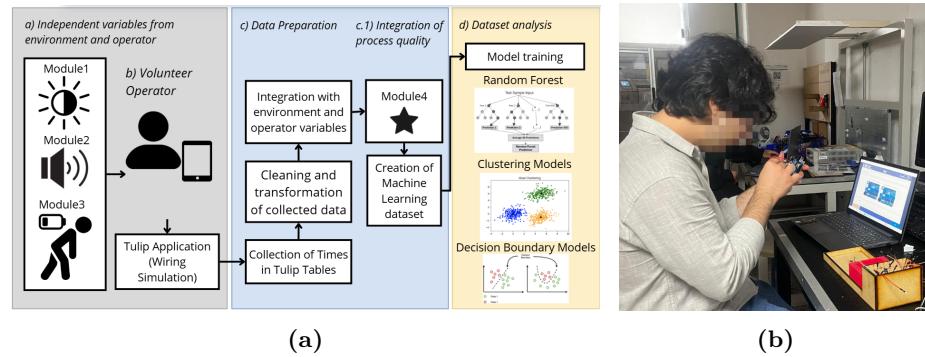


Fig. 1: Experimental infrastructure: (a) MES data collection workflow and (b) operator training session.

3.2 MES Application Design

A low-code Tulip application was developed to guide operators through the assembly workflow while automatically capturing execution data. Figure 2 summarizes the main interaction screens and the resulting time log. The application begins with an introduction screen and pre-task reminders, then guides operators through sequential assembly steps using visual instructions. Each step completion triggers a time-stamped event that is written to Tulip Tables, enabling automatic retrieval of per step durations and overall task time. The time log exported from Tulip Tables serves as the core execution layer dataset used for descriptive analysis and ML.

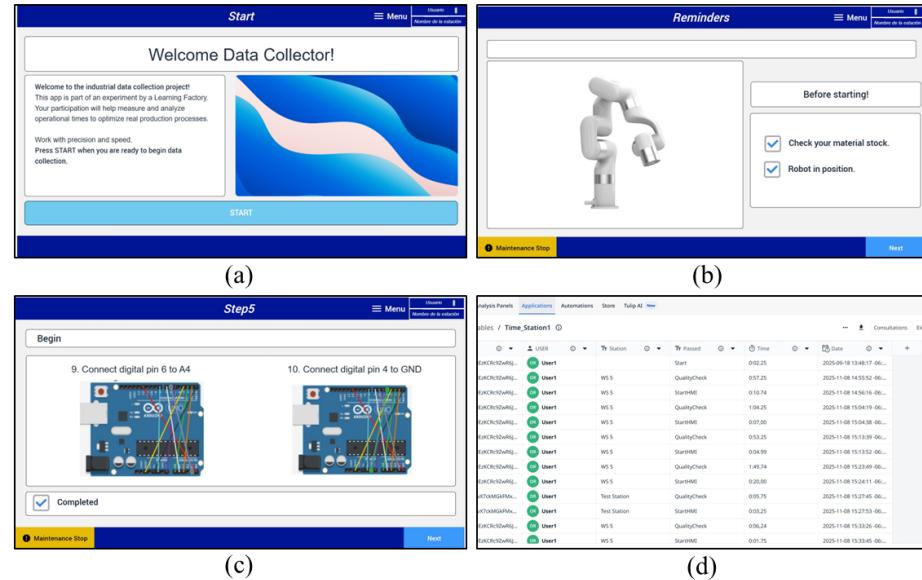


Fig. 2: Tulip MES application overview: (a) introduction screen, (b) pre task reminders, (c) assembly step with visual guidance, and (d) automatic time retrieval per step from Tulip Tables.

3.3 Data Collection Protocol

The MES automatically captured *task completion times* for each assembly step, providing high resolution temporal data without manual intervention. Environmental and contextual variables were recorded manually at the start of each session: *operator fatigue level* (self reported 1–5 scale), *ambient noise* (measured in dB), and *workstation luminosity* (measured in lux). This hybrid approach balanced automated precision with contextual richness, enabling correlation analysis between environmental conditions and performance metrics.

3.4 Analytical Methods

Collected data underwent three complementary analyses. (1) *Descriptive statistics* characterized task duration distributions across operators and identified variability patterns. (2) *Supervised learning models*, including Random Forest regression (task duration prediction), Support Vector Machines, K Nearest Neighbors, Logistic Regression, Gradient Boosting, and Random Forest classifiers (efficiency categorization), were trained and validated using 80/20 train-test splits. (3) *Unsupervised clustering* via K Means with PCA dimensionality reduction identified distinct behavioral profiles based on execution patterns and contextual sensitivity. All analyses were implemented in Python using scikit-learn, with performance evaluated through standard metrics (R^2 , accuracy, silhouette scores).

4 Results and Discussion

4.1 Data Overview and Descriptive Analysis

MES captured execution data from 25 training sessions revealed substantial inter-operator variability in task completion times. Figure 3 shows the distribution of overall task durations across all recorded executions. The distribution is positively skewed, indicating that most sessions were completed quickly while a smaller number of runs required substantially longer times. This variability suggests heterogeneous learning rates and adaptation strategies among participants, patterns that would be difficult to quantify reliably without systematic, automated data capture at the execution layer.

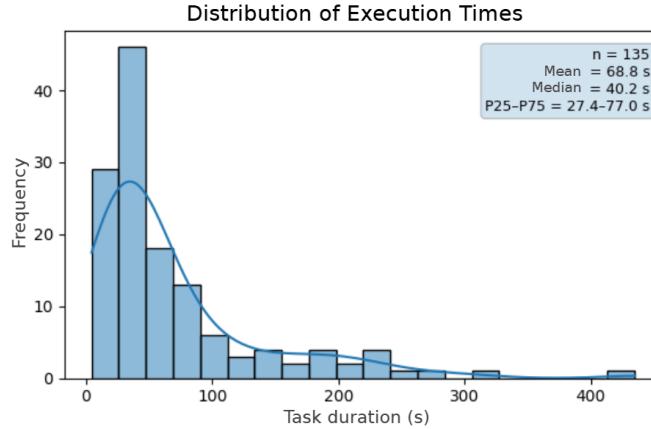


Fig. 3: Distribution of overall task duration across recorded training executions, showing positively skewed completion times and inter-operator variability.

4.2 Predictive Modeling Results

To estimate task duration from MES captured variables, we evaluated multiple regression approaches under the same feature set and data split, using R^2 and mean absolute error as the primary metrics. Across the tested models, tree based methods captured non linear relationships and interaction effects between contextual conditions and execution time more effectively than simpler baselines. We therefore selected Random Forest as it is well suited for heterogeneous tabular MES data, requires minimal feature engineering, and provides stable performance under limited sample sizes while remaining robust to outliers and mixed feature types. Table 1 reports the performance of the selected model.

The Random Forest regressor achieved a strong predictive performance, explaining approximately 92% of the variance in task duration ($R^2 = 0.92$). This indicates that the measured context provides sufficient information to predict execution time with high accuracy. The model also achieved a low mean absolute error (MAE = 3.2 seconds), suggesting practical utility for anticipating training outcomes and identifying conditions that systematically slow down or accelerate task completion.

Table 1: Random Forest classification performance for efficiency prediction

Class	Precision	Recall	F1-score	Support
0	0.73	0.89	0.80	9
1	0.94	0.83	0.88	18
Accuracy				0.85
Macro avg	0.83	0.86	0.84	27
Weighted avg	0.87	0.85	0.85	27

4.3 Operator Efficiency Classification

To support objective evaluation of training executions, we formulated a binary classification task where each run was labeled as efficient when its duration was below the median and inefficient otherwise. We compared several widely used classifiers using accuracy as the primary metric. Figure 4 summarizes the comparative results.

Random Forest achieved the highest accuracy (0.853), while K Nearest Neighbors and Support Vector Machine both reached 0.852. Logistic Regression and Gradient Boosting achieved 0.815. Given the relatively small dataset, these differences should be interpreted as indicative rather than definitive. We selected Random Forest as the reference classifier not only because it achieved the best accuracy under the same evaluation protocol, but also because it offers practical advantages for this setting. As an ensemble method, it is comparatively robust to noise in small samples, can capture non linear interactions among contextual variables with limited feature engineering, and provides feature importance estimates that support interpretability of the factors associated with efficient versus

inefficient executions. These properties make it suitable for scalable efficiency monitoring in LF training sessions.

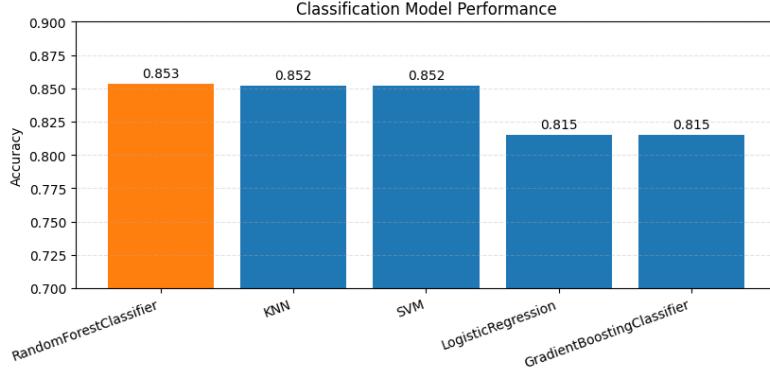


Fig. 4: Classification model performance for efficiency categorization, with Random Forest achieving 0.853 accuracy.

4.4 Behavioral Clustering and Operator Profiles

Beyond supervised tasks, we explored whether execution behavior patterns emerge from the MES derived data without labels. We applied PCA to reduce dimensionality and then performed K-Means clustering on the transformed space. Clustering was conducted on standardized MES-derived variables (*task duration, luminosity, noise, and fatigue*) to uncover distinct execution behavior profiles. The results shown in Figure 5 suggest three distinguishable behavioral profiles.

Cluster 0 ($n = 54$) represents an intermediate behavior, characterized by moderate completion times and relatively stable performance, suggesting a standard or baseline execution regime. Cluster 1 ($n = 63$) which contains the largest number of samples, is associated with shorter and more consistent task durations and can be interpreted as the most efficient execution profile under the evaluated conditions, while Cluster 2 ($n = 18$) forms a smaller and more distinct group with higher variability and longer execution times, which can be interpreted as the least efficient profile and a candidate for targeted feedback or additional training. These clusters suggest that executions can be grouped into repeatable behavioral regimes, enabling differentiated training strategies and targeted interventions based on observed execution signatures, while also complementing the supervised classification results by providing an unsupervised way to flag low-efficiency trials for post-activity reporting and data-driven training decisions without relying solely on instructor observation.



Fig. 5: PCA projection of K-Means clusters, indicating three separable execution regimes with different performance signatures.

4.5 Implications for LF Design

MES-based execution tracking enables objective analysis of training beyond subjective observation. In our study, prediction, classification, and clustering together supported duration forecasting, efficiency monitoring, and identification of distinct execution profiles, showing that Learning Factory activities can generate datasets for human-centered research and evidence-based training improvement.

5 Conclusions

This work shows that MES can function not only as industrial tools but also as enabling infrastructure for LFs. By integrating Tulip MES in a university's LF, we implemented automated traceability at the execution layer and generated structured data suitable for quantitative analysis of training sessions. The resulting pipeline supported task duration prediction with $R^2 = 0.92$, efficiency classification with 85% accuracy, and the identification of three operator behavioral profiles through clustering. Together, these results indicate that a low-code MES can provide LF environments with data centric capabilities comparable to industrial settings, without requiring enterprise scale integration.

A central takeaway is that MES integration shifts training assessment from subjective observation to evidence based evaluation. By capturing execution data systematically and linking it to contextual variables, the study reveals measurable relationships between operator state, workspace conditions, and performance outcomes. This enables instructors to adjust training protocols using

quantitative evidence, identify participants who may require additional support, and evaluate whether instructional changes meaningfully improve outcomes, which directly aligns with the data literacy expectations of Industry 4.0.

5.1 Contributions to Learning Factory Practice

This study contributes a replicable workflow that enables LFs to move from simulation oriented training to measurable operator development. The proposed approach combines MES deployment in a guided task, automated collection of execution events, integration with contextual variables, and downstream statistical and ML analysis. It requires limited additional infrastructure beyond the station setup and tablet based interfaces, yet it supports practical analytics for monitoring training outcomes. The discovered behavioral profiles suggest that different operator groups may benefit from different instructional strategies, and the predictive models support the feasibility of feedback mechanisms that can inform training design

5.2 Limitations and Future Directions

The findings are limited by a modest number of recorded sessions and a single task focused on cable assembly. Future work should expand to longitudinal studies that track individual improvement across repeated sessions and to multi task scenarios that include collaborative human robot operations. Extending data integration toward ERP systems would also enable more complete traceability across planning and execution layers. In addition, causal analysis methods could help distinguish correlation from causation when interpreting the impact of environmental factors. Finally, deploying predictive analytics during live sessions would enable adaptive training interventions based on observed execution signals. Overall, the study positions MES based traceability as a practical foundation for LFs seeking to align training evaluation with modern manufacturing requirements.

References

1. Baena, F., Guarin, A., Mora, J., Sauza, J., Retat, S.: Learning factory: The path to industry 4.0. *Procedia Manufacturing* **9**, 73–80 (2017). <https://doi.org/https://doi.org/10.1016/j.promfg.2017.04.022>, <https://www.sciencedirect.com/science/article/pii/S2351978917301403>, 7th Conference on Learning Factories, CLF 2017
2. Bradley, R., Salim, S.S., Anthony, B.W.: Learning through development of a digital manufacturing system in a learning factory using low-code/no-code platforms. *Manufacturing Letters* **46**, 10–15 (2025). <https://doi.org/10.1016/j.mfglet.2025.09.001>
3. Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., De Felice, F.: Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability* **12**(2), 492 (2020). <https://doi.org/10.3390/su12020492>
4. Kaczmarczyk, V., Zezulka, F., Benešl, T., Arm, J., Marcoň, P., Jirsa, J., Venkrbec, L.: Revisiting the role of manufacturing execution systems in industry 4.0. *IFAC-PapersOnLine* **55**(4), 151–157 (2022). <https://doi.org/10.1016/j.ifacol.2022.06.025>
5. Morgan, J., Halton, M., Qiao, Y., Breslin, J.G.: Industry 4.0 smart reconfigurable manufacturing machines. *Journal of Manufacturing Systems* **59**, 481–506 (2021). <https://doi.org/10.1016/j.jmsy.2021.03.001>

6. Mourtzis, D., Vlachou, E., Milas, N.: Manufacturing execution systems: A review of the state-of-the-art and future trends. *International Journal of Computer Integrated Manufacturing* **29**(7), 677–693 (2016)
7. Shojaeininasab, A., Charter, T., Jalayer, M., Khadivi, M., Ogunfowora, O., Raiyani, N., Yaghoubi, M., Najjaran, H.: Intelligent manufacturing execution systems: A systematic review. *Journal of Manufacturing Systems* **62**, 503–522 (2022). <https://doi.org/10.1016/j.jmsy.2022.01.004>
8. Tole, S., Schmidt, M., Braun, S., Müller, E.: Digital transformation of learning factories: A systematic review. *Procedia CIRP* **106**, 37–42 (2022)
9. Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D.: Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems* **48**, 144–156 (2018). <https://doi.org/10.1016/j.jmsy.2018.01.003>