# Predictive Quality Control: Merging Machine Learning with Lean Six Sigma

**Harsimran Kaur**

Master of Science in Engineering Management

**Abstract:** ***This paper proposes a conceptual framework that integrates Lean Six Sigma with machine learning to enable predictive quality control in modern manufacturing systems. The framework embeds machine learning models within the DMAIC methodology to support early defect prediction, improved process capability, and proactive decision making. Random Forest and Support Vector Machine based approaches are discussed as representative predictive techniques for enhancing quality management (Breiman, 2001; Cortes & Vapnik, 1995). The paper also examines the theoretical foundations of implementation challenges related to data quality, organizational readiness, and model governance while highlighting implications for engineering managers leading Industry 4.0 initiatives (Sordan et al., 2024). The proposed framework provides a structured foundation for future empirical validation and practical adoption in smart manufacturing environments.***

**Keywords:** Lean Six Sigma; Machine Learning; Predictive Quality Control; DMAIC; Process Capability; Predictive Analytics; Smart Manufacturing; Industry 4.0

## 1. Introduction

The increasing complexity of modern manufacturing environments demands quality control systems that go beyond traditional reactive approaches. Lean Six Sigma (LSS) has long served as a structured framework for reducing process variability and eliminating waste (George, 2002); however, its reliance on historical data and human-driven analysis limits its predictive potential. Traditional LSS implementations are largely retrospective, meaning corrective action is initiated only after quality failures have already occurred or been detected (Pyzdek & Keller, 2018).

The rise of Industry 4.0 technologies- including the Internet of Things (IoT), cloud computing, and advanced analytics- has significantly expanded the volume and velocity of data available on the factory floor (Sordan et al., 2024). This presents an unprecedented opportunity to shift quality management from a corrective posture to a predictive one. Machine learning, a subset of artificial intelligence, enables systems to learn patterns from historical and real-time data, predict future outcomes, and support decision-making with minimal human intervention (Breiman, 2001).

This paper argues that embedding ML capabilities within the LSS-DMAIC structure can create a powerful hybrid framework that preserves the structured discipline of Six Sigma while significantly enhancing its predictive and adaptive capabilities. The paper is organized as follows: Section 2 reviews the relevant literature; Section 3 presents the proposed integration framework; Section 4 discusses the research methodology and future validation strategy; Section 5 addresses implementation challenges; and Section 6 concludes with directions for future research.

## 2. Literature Review

Lean Six Sigma has been extensively studied across sectors including healthcare, semiconductor manufacturing, food processing, and automotive production (Womack & Jones, 2003). The methodology is widely recognized for its ability to reduce defect rates, improve process cycle efficiency, and lower operational costs (Pande et al., 2000). The DMAIC cycle in particular provides clear phase gates- Define, Measure, Analyze, Improve, Control- that guide teams through structured problem-solving (Pyzdek & Keller, 2018).

Recent scholarship has begun exploring the intersection of LSS with digital technologies. Sordan et al. (2024) demonstrated that integrating AI with Lean Six Sigma in Industry 4.0 contexts can substantially augment the Analyze phase of DMAIC by uncovering non-linear relationships among process variables that traditional statistical tools such as regression and control charts may miss. Furthermore, ML models have demonstrated effectiveness in anomaly detection and real-time process monitoring, both of which are critical for quality assurance in high-throughput manufacturing environments (Mansour et al., 2025).

Breiman's (2001) foundational work on Random Forests established the basis for ensemble-based predictive modeling, which has since been applied widely in industrial quality applications. Support Vector Machines, introduced by Cortes and Vapnik (1995), have similarly been applied to binary classification problems in manufacturing defect detection due to their robustness in high-dimensional feature spaces. More recently, Mansour et al. (2025) integrated Six Sigma, machine learning, and real-time defect prediction in a manufacturing context, demonstrating that ML-augmented quality systems can reduce defect rates while improving inspection efficiency. Despite this growing body of work, most existing studies apply ML and LSS independently or in loosely coupled ways. A tightly integrated framework- where ML models are embedded at specific DMAIC phases and governed by LSS quality targets- remains an underexplored area (Sordan et al., 2024). This gap motivates the present study.

## 3. Proposed Integration Framework

The proposed framework, termed the ML-Enhanced DMAIC (MLE-DMAIC) model, maps machine learning capabilities onto each phase of the traditional DMAIC cycle (Pyzdek & Keller, 2018) as follows:

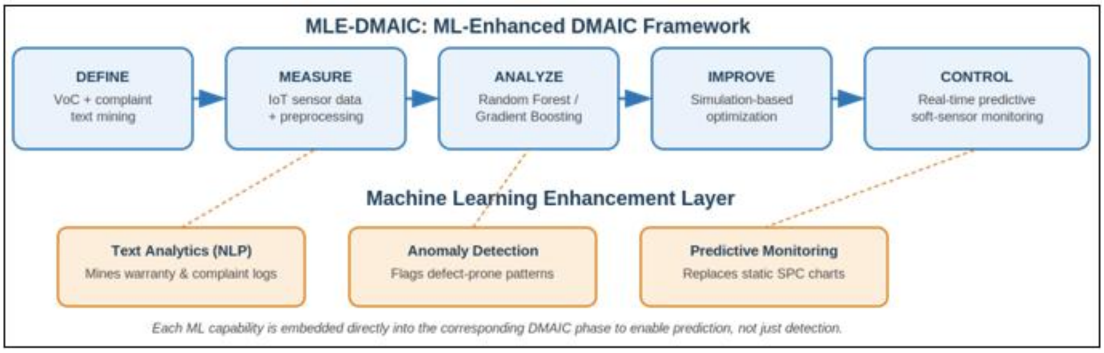


**Figure 1:** The MLE-DMAIC Framework- Machine Learning capabilities embedded within each DMAIC phase.

### 3.1 Define Phase

In this phase, engineering managers use voice-of-the-customer (VoC) data and process failure mode analysis to establish Critical-to-Quality (CTQ) characteristics (George, 2002). ML-based text analytics can be employed to mine customer complaint databases, warranty records, and maintenance logs to accelerate the identification of quality-critical parameters, providing a data-driven prioritization of root causes.

### 3.2 Measure Phase

IoT sensors embedded in machinery collect real-time process data including temperature, pressure, vibration, and dimensional measurements. This data feeds into ML preprocessing pipelines that handle noise reduction, missing value imputation, and feature engineering- producing a clean, structured dataset ready for analysis (Mansour et al., 2025).

### 3.3 Analyze Phase

Supervised ML models, particularly Random Forest (Breiman, 2001) and Support Vector Machine (Cortes & Vapnik, 1995) classifiers, are trained on historical defect data to identify which process parameters most strongly predict product failures. Feature importance scores generated by these models complement traditional tools such as fishbone diagrams and Pareto charts, providing a data-driven prioritization of root causes that supports more targeted corrective action.

### 3.4 Improve Phase

Simulation-based optimization, guided by ML-predicted outcomes, is used to test process adjustments virtually before physical implementation. This reduces the time and cost associated with traditional design-of-experiments (DOE) trials (Pande et al., 2000) while enabling a broader search across the parameter space.

### 3.5 Control Phase

Real-time predictive models are deployed as soft sensors that continuously monitor process outputs and issue alerts when predicted defect probabilities exceed established thresholds (Mansour et al., 2025). This replaces conventional Statistical Process Control (SPC) charts with a dynamic, adaptive monitoring system that reduces false positives and responds faster to process drift.

## 4. Research Methodology

### 4.1 Conceptual Research Approach

This study adopts a conceptual research methodology, which is appropriate for proposing integrated frameworks that synthesize existing theories and practices from two or more established domains (Sordan et al., 2024). The MLE-DMAIC framework is developed through a structured review of peer-reviewed literature in the areas of Lean Six Sigma, machine learning, and smart manufacturing, followed by a systematic mapping of ML capabilities onto the DMAIC cycle phases. Conceptual frameworks of this nature serve as foundational contributions that precede and guide empirical investigation (Pande et al., 2000). The authors acknowledge that the proposed MLE-DMAIC model has not yet been validated through controlled experiments, industrial case studies, or simulation studies, which represent important and necessary directions for future work.

### 4.2 Future Validation Strategy

To transition the MLE-DMAIC framework from conceptual proposal to evidence-based practice, a structured pilot validation study is proposed. The recommended validation environment is a mid-sized discrete manufacturing facility- such as an automotive components supplier or electronics assembly plant- where process sensor data is already captured via IoT infrastructure and historical defect records are maintained in a structured database.

The validation dataset would consist of multivariate time-series process readings (e.g., temperature, pressure, vibration, torque) linked to binary defect outcome labels (conforming / non-conforming). A minimum dataset of 5,000 labeled observations per process line is recommended to support robust model training and cross-validation. Random Forest and Support Vector Machine classifiers (Breiman, 2001; Cortes & Vapnik, 1995) would be trained on 70 percent of the data, with the remaining 30 percent reserved for hold-out testing.

Model performance would be evaluated using the following metrics: Accuracy (percentage of correctly classified observations); Precision (proportion of predicted defects that are actual defects); Recall (proportion of actual defects correctly identified); F1 Score (harmonic mean of precision and recall, particularly informative for imbalanced defect datasets); and Process Capability Index (Cpk) before and after ML integration to quantify improvements in process control. A statistically significant improvement in Cpk alongside F1 scores exceeding 0.85 would constitute evidence in support of the framework's practical effectiveness (Pyzdek & Keller, 2018). Cross-sector replication across at least two manufacturing environments would further establish generalizability.

## 5. Implementation Challenges and Managerial Implications

While the MLE-DMAIC framework offers significant theoretical advantages, its practical implementation raises several important challenges for engineering managers.

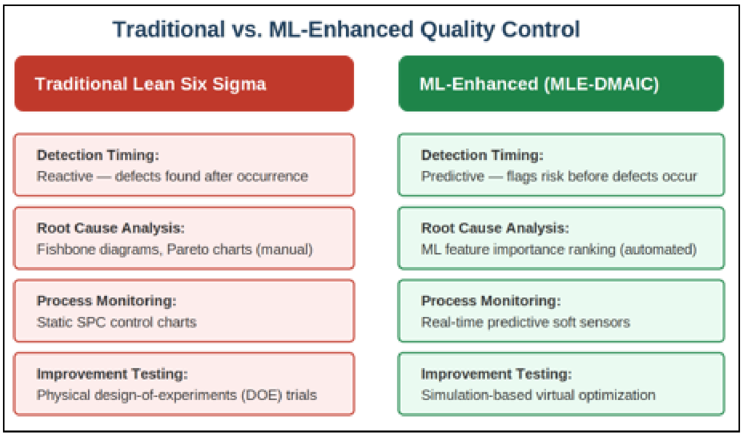


**Figure 2:** Comparison of traditional Lean Six Sigma practices versus the ML-enhanced (MLE-DMAIC) approach.

### 5.1 Data Quality and Infrastructure

Machine learning models are only as reliable as the data they are trained on (Breiman, 2001). Many manufacturing facilities, particularly small and medium enterprises, lack the sensor infrastructure and data governance practices necessary to support high-quality ML pipelines. Investments in IoT hardware, data storage, and cleaning protocols are prerequisites for successful deployment (Sordan et al., 2024).

### 5.2 Change Management

Lean Six Sigma practitioners accustomed to traditional statistical tools may be resistant to ML-driven approaches they perceive as opaque or difficult to interpret (Pande et al., 2000). Engineering managers must invest in cross-functional training programs that build ML literacy among quality engineers and process technicians. Explainable AI (XAI) techniques- such as SHAP (SHapley Additive exPlanations) values — can help bridge this gap by making model predictions interpretable to non-technical stakeholders (Mansour et al., 2025).

### 5.3 Model Maintenance and Governance

Unlike static quality standards, ML models require periodic retraining as process conditions change over time. Without a defined model governance protocol embedded within the Control phase, predictive accuracy can degrade silently, undermining the reliability of the quality system. Organizations must establish clear ownership, retraining schedules, and performance benchmarks for deployed models (Sordan et al., 2024).

### 5.4 Regulatory and Compliance Considerations

In regulated industries such as aerospace and pharmaceuticals, quality management systems must meet strict documentation and validation requirements (Pyzdek & Keller, 2018). Integrating ML into DMAIC workflows requires demonstrating that predictive models meet the same evidence standards as traditional statistical methods- a non-trivial organizational and technical challenge.

## 6. Conclusion

This study presented the MLE-DMAIC framework as a conceptual approach for integrating machine learning with Lean Six Sigma to support predictive quality control in manufacturing. The framework illustrates how predictive analytics can strengthen each phase of the DMAIC cycle and contribute to more proactive quality management, representing a transition from reactive quality management toward proactive defect prevention. Although the proposed model has strong theoretical and managerial value, empirical validation is required before widespread implementation. Future studies should evaluate the framework using real industrial datasets, compare its performance across manufacturing sectors, and investigate advanced artificial intelligence techniques for adaptive process optimization (Sordan et al., 2024; Breiman, 2001).

## References

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[2] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
[3] George, M. L. (2002). Lean Six Sigma: Combining Six Sigma Quality with Lean Speed. McGraw-Hill.
[4] Mansour, H., Abohashima, H., Elkhouly, H., & Harraz, N. (2025). Smart quality control: Integrating Six Sigma, machine learning and real-time defect prediction in manufacturing. International Journal of Lean Six Sigma, 16(6), 1415–1447. https://doi.org/10.1108/IJLSS-05-2024-0089
[5] Pande, P. S., Neuman, R. P., & Cavanagh, R. R. (2000). The Six Sigma Way: How GE, Motorola, and Other Top Companies are Honing Their Performance. McGraw-Hill.
[6] Pyzdek, T., & Keller, P. (2018). The Six Sigma Handbook (5th ed.). McGraw-Hill Education.
[7] Sordan, J. E., Oprime, P. C., Pimenta, M. L., & Chiabert, P. (2024). Integration of AI with Lean Six Sigma in the Fourth Industrial Revolution. Journal of Manufacturing Technology Management, 35(2), 210–229. https://doi.org/10.1108/JMTM-03-2023-0091
[8] Womack, J. P., & Jones, D. T. (2003). Lean Thinking: Banish Waste and Create Wealth in Your Corporation (2nd ed.). Free Press.