

Machine Learning-Based Duration Prediction in Production Processes

Cagri Erhan Aygordu, Mustafa Kirik, Kemal Guneri, Firat Alma and Sena Aydoğan

Department of Industrial Engineering, Faculty of Engineering

Gazi University, Ankara, Türkiye

cagrierhanunigazi@gmail.com, mustafa.kirik1@gazi.edu.tr, kemalguneribusiness@gmail.com,
almafirat61@gmail.com

Abstract

This study presents a machine learning-based decision support system designed to predict welding operation durations in the galvanized steel pole production line of Mitaş Poligon, utilizing the geometric and operational characteristics of manufacturing components. Traditional manual scheduling approaches suffer from experience-dependent subjectivity, leading to production floor imbalances and capacity plan mismatches. To overcome these constraints, a 1,892-row raw enterprise dataset was extracted from the factory's ERP system and comprehensively filtered to remove missing dimensions or non-standard entries, establishing a core baseline of 1,395 production records. To support model generalization and ensure structural completeness across all 54 product configurations (Part Type × Welding Type × Position), a 605-row synthetic dataset was generated via a physical arc-welding speed model coupled with a 6% Gaussian noise factor, forming a robust 2,000-row hybrid validation dataset. Following the structured CRISP-DM lifecycle, seven advanced machine learning architectures were cross-validated. The CatBoost regressor outperformed alternative models, yielding a Coefficient of Determination (R^2) of 0.9584, a Mean Absolute Error (MAE) of 0.4490 min (~27 seconds), and a Root Mean Squared Error (RMSE) of 0.6421 min. The final optimization pipeline was compiled into a standalone Python/Tkinter desktop graphical user interface (.EXE) for localized, offline floor utilization without structural database dependencies. This data-driven model standardizes planning times, mitigates manufacturing bottlenecks, and provides an objective operational framework for sustainable capacity analysis. Furthermore, the system supports the United Nations Sustainable Development Goals by advancing Industry, Innovation and Infrastructure (SDG 9), Responsible Consumption and Production (SDG 12), and Climate Action (SDG 13).

Keywords

Machine Learning, CatBoost, Production Scheduling, Hybrid Dataset, Decision Support Systems, Sustainability

Acknowledgments

This research project was supported by the Department of Industrial Engineering at Gazi University. The authors would like to thank our advisor, Assoc. Prof. Sena Aydoğan, for her academic supervision and insightful feedback. Additionally, we extend our gratitude to Mitaş Poligon for their technical collaboration and data allocation.

Biographies

Cagri Erhan Aygordu was born in Ereğli, Konya, Türkiye. He is currently a senior undergraduate student in the Department of Industrial Engineering at the Faculty of Engineering, Gazi University, Ankara, Türkiye. He has completed professional long-term internship programs focusing on production planning, operational optimization, and enterprise resource planning (ERP) management systems. Outside of his academic pursuits, he is an amateur basketball player and enthusiast. His professional goals and research interests include the practical deployment of machine

learning algorithms, industrial database structures, and data-driven decision support mechanisms. He aspires to build integrated industrial applications that minimize execution variances and enhance shop-floor planning metrics.

Mustafa Kirik was born in Ankara, Türkiye. He is a final-year undergraduate student in the Department of Industrial Engineering at the Faculty of Engineering, Gazi University, Ankara, Türkiye. He has completed practical manufacturing and field-level internship programs focusing on time studies, assembly line balancing, and plant-floor layouts. His personal hobbies include playing strategic board games. His professional ambitions are centered on Industry 4.0 systems integration, smart mobility frameworks, and predictive optimization models to streamline logistics scheduling and resolve capacity limitations in structured manufacturing environments.

Kemal Guneri was born in Bursa, Türkiye. He is currently a senior undergraduate student in the Department of Industrial Engineering at the Faculty of Engineering, Gazi University, Ankara, Türkiye. He has actively engaged in undergraduate projects focused on supply chain logistics, operations research, and enterprise resource planning integrations within the manufacturing sector. Outside of his academic responsibilities, he is an avid outdoor enthusiast and enjoys trekking and nature hiking. His professional goals are centered on utilizing operations management frameworks, data analytics, and modern continuous improvement methodologies to optimize resource utilization and streamline shop-floor execution parameters.

Firat Alma was born in Ankara, Türkiye. He is a final-year Industrial Engineering undergraduate student at the Faculty of Engineering, Gazi University, Ankara, Türkiye. Throughout his undergraduate tenure, he has participated in applied field studies and internship programs concentrating on production scheduling, facility layout design, and statistical process control applications. His personal hobbies include playing football and participating in team sports. His professional interests are focused on Industry 4.0 applications, digital transformation metrics, and the integration of predictive analytical toolsets to resolve complex manufacturing constraints and enhance structural planning efficiency.