Machine Learning Techniques to Predict Process Time in Operations with High Variability

Kenny-Jesús Flores-Huamán, María-Luisa Muñoz-Díaz, María Rodríguez-Palero, and Alejandro Escudero-Santana(✉)

Dpto. de Organización de Industrial y Gestión de Empresas II, Escuela Técnica Superior de Ingenierías, Universidad de Sevilla, Camino de los descubrimientos s/n, 41092 Sevilla, Spain
alejandroescudero@us.es

Abstract. This study focuses on the exploration and application of machine learning techniques to predict the process time of operations with high variability. The proposed methodology includes the identification of significant parameters, the individual prediction of every operation and the integration of individual prediction to determine the lead time of the chain. The study is conducted in wind tower manufacturing plants located in Spain and Brazil. The proposed approach overcomes the limitations of other current techniques such as direct formulation or linear programming. The results indicate that, overall, gradient boosting models, such as XGBoost or LightGBM, achieve better performance in a significant portion of the operations, although other models like NODE also demonstrate superior results in certain specific operations.

Keywords: Industrial Machine Learning · Deep Learning · regression · process time · prediction

1 Introduction

The revolution in Industry 4.0 is radically transforming the way factories and services operate worldwide. At the core of this transformation is the promotion of advanced sensors and connected machines, enabling to continuously collect vast amounts of operation-related data, thereby opening new opportunities to optimize and enhance processes.

Machine learning has emerged as an ideal technology for extracting insights from collected data. This technology has the capability to provide predictions and discern complex patterns, leading to the development of intelligent systems.

These systems are crucial in various manufacturing and logistic tasks, such as predictive maintenance, quality improvement, supply chain management, task scheduling, process optimization, among others (Rai, Tiwari, & Dolgui, 2021).

Focusing on the case of study, one critical indicator in production planning and control within a factory is lead time. This time refers to the interval between issuing a product order and it being available to meet customer demands. Estimating this time is

R. Carrasco-Gallego et al. (Eds.): CIO 2024, LNDECT 239, pp. 350–355, 2025.
https://doi.org/10.1007/978-3-031-82334-3_61

important as it allows for more efficient sequencing of activities, allocation of appropriate resources, and setting realistic delivery deadlines, thereby avoiding delays and bottlenecks in the production flow, ultimately enhancing process quality and efficiency. In addition, the correct prediction of each process time is necessary for an adequate cost estimate.

Currently, factories employ techniques such as direct formulation or linear programming to estimate lead times. However, these methodologies have certain limitations. Direct formulation, although widely used, may not realistically reflect the times or costs involved in processes. On the other hand, linear programming, while useful, may not be fully updatable over time, making it difficult to apply in dynamic production environments. Therefore, the ability of machine learning algorithms to analyse large sets of historical data, identify patterns, and make accurate predictions makes them invaluable for optimizing lead times and improving operational efficiency in the wind tower industry.

Regarding literature on wind turbine manufacturing from an operational perspective, focusing on fundamental operations and activities for the production and manufacturing of these components, it is quite limited. Sainz (2015) outlines the various stages of the process and highlights several automation technologies that could increase industry production capacity. There is a study examining variability in delivery times for the bending operation in wind tower manufacturing using machine learning, aiming to understand how different factors affect these times and improve plant planning and control (Lorenzo-Espejo et al. 2022). However, no research has been found that focuses exclusively on wind tower production or employing machine learning techniques to predict execution time in each operation. Thus, predicting execution time is considered a novel approach in this field.

Therefore, the objective of this research work focuses on the study and application of different machine learning techniques to predict times. The study was developed using a dataset from different operations within the wind tower manufacturing process in factories in Spain and Brazil. This approach aims to overcome the limitations of other current techniques. The remainder of the work is structured as follows: a brief summary of the most relevant operations in wind tower construction is described in Sect. 2; the methodology employed in this study is detailed in Sect. 3; finally, Sect. 4 presents and discusses the results obtained and provide a brief conclusion of the study.

2 Identification of Key Process

The manufacture of wind turbine towers is a complex process involving the handling of large-sized products and the execution of numerous manual operations. It begins with the selection of suitable materials, mainly steel plates, due to their strength and durability. These plates are cut according to design specifications and prepared for welding by bevelling.

Then, they are bent to form rings called "cans" and longitudinally welded. Several cans are joined together to form tower sections, which are manufactured separately for easy transportation, and cans are added for assembly.

The assembled sections undergo quality testing and surface treatments, such as painting and anti-corrosion coatings, to ensure longevity. Finally, the sections are transported to the installation site where they are assembled to complete the tower structure. This process ensures the robustness and durability of wind turbine towers.

Due to the wide array of processes involved in a wind turbine tower factory, this research primarily concentrates on key procedures. Cutting entails slicing steel plates into raw pieces to form strips for cans; Bevelling involves cutting steel edges at angles other than 90 degrees for sturdy welding joints; Bending shapes steel pieces into cylindrical forms, creating tower cans; Longitudinal welding joins steel ends to form cans; Rebending corrects deformations in stored cans; Fit-up and circular welding connect cylindrical steel cans via circular welding to construct tower sections, adding cans to the streamline assembly.

3 Methodology

The methodology behind this study has been divided into 3 steps: data collection and preprocessing, and the application of machine learning algorithms to individual process, and the estimation of the integrated lead time.

The data were gathered from various sources, including the company's ERP and technical data obtained from each project. Due to the diversity of data sources and the need to analyse each process individually, a process of integrating all this information independently was undertaken in each of the relevant processes. This data was collected throughout the year 2022 at the facilities of the factories located in Seville (Spain) and Pernambuco (Brazil). In Seville, a total of 20,938 temporary entries were recorded in the ERP system, while in Pernambuco, 210,740 temporary records were accumulated.

Once the data had been collected, an initial exploration was conducted to identify patterns, trends, and possible outliers in each operation. It is important to note that much of the information on waiting times and machinery usage is entered by workers during the operation, which can result in errors. Furthermore, it should be considered that the workflow can vary significantly between the two countries.

To improve the quality of the information and facilitate the development of predictive models, a process of cleaning and transformation was carried out on each dataset. This included correcting errors in the texts, imputing missing values, and normalizing attributes.

Additionally, feature engineering techniques were applied to optimize predictive quality by removing irrelevant attributes and adding attributes based on the knowledge provided by the company's engineering department.

The main objective of this research is to implement a reliable and maintainable predictive time operation model based on actual hours obtained. Therefore, to address this regression problem, machine learning models can be employed.

Since the No Free Lunch theorem highlights that there is no universal optimal solution for all problems (Wolpert & Macready, 1997), this implies the need to explore and evaluate various models and approaches to find the best solution for each of the different operations under study.

Various algorithms are employed for different operations: Decision Trees use decision rules to assign values; Random Forest employs multiple trees to improve accuracy and reduced overfitting; Lasso Regression, Ridge Regression, and ENET are linear regression models that control overfitting; Support Vector Regression extends SVM for regression, by finding a hyperplane that maximizes the margin; Gradient Boosting is a sequential ensemble method that corrects errors using decision trees; XGBoost (XGB) and LightGBM (LGBM) are optimized implementations of Gradient Boosting for improved speed and accuracy; Multilayer Perceptron (MLP) are based on neural networks that capture nonlinear relationships; TabNet is a deep neural network using sequential attention for feature selection; NODE is a tabular data architecture with differentiable oblivious decision trees trained end-to-end.

Once the data has been prepared, the optimal model for relevant operations in the study is selected. It begins by training and evaluating models without adjusting hyperparameters, using them as baseline models for comparison with tuned models. The Hold-Out technique is employed for unbiased evaluation and to prevent overfitting by splitting the dataset into training and testing sets. Subsequently, the most suitable model for each relevant operation in the factory is sought using Random Search with cross-validation to explore hyperparameters due to the vast search space of certain models such as XGBoost. This combination reduces bias and provides a more accurate estimate of performance on unseen data.

The model selection process has prioritized choosing the model that has demonstrated the most prominent metrics, giving precedence first to the RMSE metric. RMSE is a measure of the accuracy of a predictive model and is mathematically defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum\nolimits_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{y_i} - \widehat{\mathrm{y_i}}\right)^2}$$

where n is the number of observations, y_i is the actual operation time, and $\widehat{y_i}$ is the predicted operation time by the model. A lower RMSE indicates a better fit of the model to the observed data, meaning the model's predictions are closer to the actual values. In case of a tie in RMSE, other metrics will be used to decide.

4 Results

This section presents and discusses the results obtained by applying machine learning techniques to address our regression problem in each of the operations. Table 1 provides a summary of the most prominent models, together with the RMSE they have obtained for each of the relevant operations in different factories.

It is crucial to highlight that the data captured from Spain and Brazil were analyzed separately due to significant differences in their entry data. These differences arise because the factories operate differently, impacting the predictions. Within the set of operations evaluated in each factory, there is no single optimal model for all situations. Instead, a limited variety of models that perform well in different contexts is observed. For example, in Spain, gradient boosting-based models such as XGBoost or LightGBM proved to be effective in problem resolution, followed by SVR. In contrast, in Brazil, where some attributes differ, boosting-based models continued to show satisfactory results, although NODE achieved better results in some specific operations.

When analyzing the RMSE values, it is noted that several operations, such as bending and recirculating, present low RMSEs, indicating a trend towards accuracy in the predictions for these datasets. However, the circular welding operation shows a significantly

Table 1. Performance assessment of the model that offers the best fit to the data across the diverse relevant operations of the two analyzed factories.

Operation	Spain		Brazil	
	Best model	RMSE (hours)	Best Model	RMSE (hours)
Cutting	NODE	0.5721	LGBM	0.1325
Beveling	SVR	2.8263	LGBM	0.094
Bending	XGB	0.3073	NODE	0.2633
Long. Welding	LGBM	0.7608	XGB	0.3211
Rebending	LGBM	0.384	NODE	0.1795
Fit-up	SVR	1.516	XGB	3.40
Circ. Welding	LGBM	12.41	RF	4.8448

higher RMSE compared to the other operations. This disparity could be related to both the dataset size and the long execution times associated with that operation. Therefore, it is suggested to investigate and improve the effectiveness of models specifically for this operation.

Once the best model for each operation was obtained, these models were integrated into a web system developed with the aim of improving the accessibility of these models to users with no experience in programming or Artificial Intelligence (Abid et al., 2019). The system was developed using Gradio, an open-source Python library used to develop demos and web applications for Artificial Intelligence.

The web application consists of two tabs, the deploy tab, which is responsible for training the loaded models to perform lead time calculations, while the train tab is responsible for training the models as they degrade over time. Figure 1 shows an example of the deploy tab.

In conclusion, this research work is pivotal in two key aspects: firstly, it pioneers an innovative approach to using machine learning for lead time calculations in the wind tower manufacturing industry. Secondly, it showcases a web prototype that demonstrates the practical implementation of these models, along with a system for training new models as existing ones degrade over time. This initiative promises substantial enhancements in operational decision-making, marking a significant leap in efficiency and accuracy. Moreover, beyond its immediate implications for wind tower manufacturing, the methodologies developed here hold the potential to be adapted and applied across various industrial sectors facing similar challenges in operational decision-making.

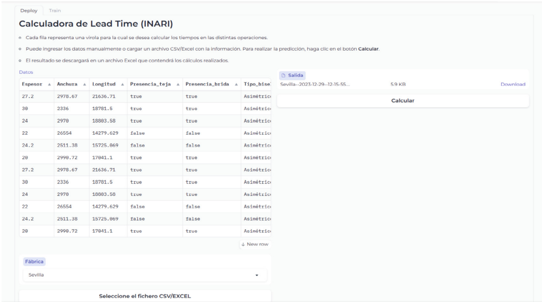

Fig. 1. Representation of the web application

Acknowledgements. This research was part of INARI project (2021/C005/00144319) funded by the agency Red.es belonging to the ministry for digital transformation and public service and cofounded by Next Generation program and Recovery plan transformation and resilience.

References

Abid, A., et al.: Gradio: hassle-free sharing and testing of ML models in the wild. arXiv preprint arXiv:1906.02569 (2019)

Lorenzo-Espejo, A., et al.: Machine learning-based analysis of a wind turbine manufacturing operation: a case study. Sustainability **14**(13) (2022). https://doi.org/10.3390/su14137779

Rai, R., Tiwari, M.K., Ivanov, D., Dolgui, A.: Machine learning in manufacturing and industry 4.0 applications. Int. J. Prod. Res. **59**(16), 4773–4778 (2021). https://doi.org/10.1080/00207543.2021.1956675

Sainz, J.A.: New wind turbine manufacturing techniques. Procedia Eng. **132**, 880–886 (2015). https://doi.org/10.1016/j.proeng.2015.12.573

Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. IEEE Trans. Evol. Comput.Comput. **1**(1), 67–82 (1997). https://doi.org/10.1109/4235.585893