REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

Volúmen 1, Número 1 (2025) Pág. 1-15

Data Mining for the Optimization of Industrial Processes in Latin American Manufacturing

Minería de Datos para la Optimización de Procesos Industriales en la Manufactura Latinoamericana

Carlos Javier Lara Lascano
https://orcid.org/0009-0008-6351-1197
larajavier776@gmail.com
Escuela Superior Politécnica de Chimborazo
Ambato – Ecuador

Abstract.- This study explores the application of data mining and machine learning techniques for industrial process optimization in Latin America, with an emphasis on the context of Industry 4.0. Using simulated data representative of real-life operations, advanced statistical methodologies were implemented, including imputation models, variable selection, principal component analysis (PCA), clustering, and predictive models such as XGBoost and SVM. The results reveal that variables such as lead time, mean time between failures (MTBF), and CO_2 emissions have a direct impact on the defect per million (PPM) rate, highlighting the interrelationship between logistical, maintenance, and environmental factors. The clustering analysis identified three operational profiles differentiated by energy efficiency and quality, facilitating targeted interventions. Despite the high performance of the XGBoost model, possible overfitting is noted, so cross-validation is recommended. Time trends did not show significant seasonality, suggesting a greater influence of internal process variables. The study concludes that the integration of advanced analytics, predictive maintenance, and artificial intelligence can significantly improve competitiveness, sustainability, and quality in Latin American manufacturing environments.

Keywords: *data, industry, mining, models, optimization.*

Resumen.- Este estudio explora la aplicación de técnicas de minería de datos y aprendizaje automático para la optimización de procesos industriales en América Latina, con énfasis en el contexto de la industria 4.0. A partir de datos simulados representativos de operaciones reales, se implementaron metodologías estadísticas avanzadas, incluyendo modelos de imputación, selección de variables, análisis de componentes principales (PCA), clustering y modelos predictivos como XGBoost y SVM. Los resultados revelan que variables como el tiempo de entrega (Lead Time), el tiempo medio entre fallas (MTBF) y las emisiones de CO_2 tienen impacto directo sobre la tasa de defectos por millón (PPM), destacando la interrelación entre factores logísticos, de mantenimiento y ambientales. El análisis de clustering permitió identificar tres perfiles operativos diferenciados por eficiencia energética y calidad, lo que facilita intervenciones focalizadas. A pesar del alto rendimiento del modelo XGBoost, se advierte posible sobreajuste, por lo que se recomienda validación cruzada. Las tendencias temporales no mostraron estacionalidad significativa, lo que sugiere una mayor influencia de variables internas del proceso. El estudio concluye que la integración de analítica avanzada, mantenimiento predictivo e inteligencia artificial puede mejorar significativamente la competitividad, sostenibilidad y calidad en los entornos manufactureros de América Latina.

Palabras clave: *datos, industria, minería, modelos, optimización.*

Received: May 31, 2019. Revised: May 4, 2020. Accepted: May 22, 2020. Published: May 29, 2020

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

1. Introduction

Data mining for industrial process optimization in Latin American manufacturing represents a transformative approach to improving operational efficiency and competitiveness in key industries in the region. Originating in the 1990s, data mining techniques have evolved significantly, integrating advanced methodologies such as machine learning and artificial intelligence to analyze large data sets [1].

This has enabled manufacturers in countries such as Brazil, Chile, and Argentina to uncover patterns and insights that drive better decision-making and resource allocation within their production processes [2].

The adoption of data mining is particularly notable in the Latin American context, where the manufacturing sector is undergoing a shift toward Industry 4.0 technologies. This transition is spurred by the need to optimize processes, reduce costs, and increase productivity amid increasing global competition [3].

Key applications of data mining in this sector include predictive maintenance, quality control, and supply chain optimization, which collectively improve operational efficiency and minimize downtime. However, challenges such as data quality, technological infrastructure, and the need for skilled labor persist, complicating the implementation of these advanced techniques [4].

The controversies surrounding data mining practices in Latin America also deserve attention, particularly in relation to ethical considerations such as data privacy and transparency. As organizations increasingly rely on personal data to drive insights, there is a risk of compromising individual privacy, necessitating strict data management protocols to comply with varying local regulations [5].

Furthermore, there is an ongoing debate about the socioeconomic impacts of these technologies, including the potential for job displacement and the exacerbation of inequalities, highlighting the need for responsible innovation that aligns with community needs [6].

As Latin American manufacturers continue to navigate these complexities, the integration of data mining into industrial processes not only represents an opportunity for improved performance but also poses significant ethical and socioeconomic challenges. Addressing these issues will be crucial to ensuring that the benefits of data-driven optimization are distributed equitably and contribute to sustainable development within the region [7].

Historical Context

Data mining has evolved significantly over the decades, emerging as a fundamental tool for optimizing industrial processes, particularly within the Latin American manufacturing sector. Its roots date back to the 1990s, when companies began leveraging powerful computing resources and advanced data storage capabilities to analyze large amounts of customer information.

This marked a transformative period in which companies recognized the potential of data mining to uncover patterns and trends that could provide them with a competitive advantage in the marketplace [8].

In the early stages, the focus of data mining was predominantly on customer relationship management, where companies aimed to predict customer behavior and improve service delivery.

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

As technology advanced, so did the techniques employed in data mining [9].

The integration of machine learning and artificial intelligence into data mining practices has enabled more sophisticated analysis of complex data sets, allowing organizations to gain actionable insights that were previously unattainable [10].

Latin America's participation in global manufacturing has also influenced the historical trajectory of data mining in the region. With abundant natural resources such as lithium, coal, and oil, countries such as Chile, Bolivia, and Argentina have positioned themselves as key players in the manufacturing landscape. As these nations sought to optimize their industrial processes, the adoption of data mining techniques became increasingly relevant [11].

Furthermore, the emergence of Industry 4.0 technology in Latin America has further accelerated the need for advanced data mining applications. Manufacturers have begun to recognize that leveraging data through mining can lead to significant improvements in operational efficiency, customer segmentation, and predictive maintenance, ultimately driving better decision-making [12].

The historical context of data mining reflects a broader shift toward a data-driven approach in industrial processes. As companies in Latin America continue to adapt and integrate these technologies, the legacy of data mining as a critical component of strategic decision-making will likely shape the future of manufacturing in the region [13].

Data Mining Techniques

Data mining encompasses a variety of techniques and methodologies used to extract valuable information from large data sets, particularly in the context of optimizing industrial processes in Latin American manufacturing.

Overview of Data Mining Techniques

Data mining techniques are employed to identify patterns, relationships, and trends within large data sets. This process often involves several stages, including data cleaning, exploratory data analysis, model building, and model evaluation. Each of these stages uses specific algorithms and methods to ensure that the extracted information is accurate and actionable [14].

Common Algorithms in Data Mining

Several algorithms prevail in the data mining landscape, each tailored to different types of analytical tasks. Decision trees: These are used for classification tasks, allowing analysts to make predictions based on the characteristics of the data set [15].

K-Means Clustering: This unsupervised learning technique is used to segment data into distinct groups based on shared attributes, making it useful for identifying customer segments or production efficiencies [16].

Support Vector Machines (SVM): Used for both regression and classification, SVMs build models by mapping features in a dataset to output classifications [17].

Naive Bayes Classifier: Based on Bayes' theorem, this algorithm is effective for classifying categorical data and is known for its computational efficiency [18].

Random Forests: This method improves prediction accuracy by aggregating the results of multiple decision trees, thereby reducing the risk of overfitting [19].

Manufacturing Applications

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

In the manufacturing sector, data mining techniques facilitate several applications that significantly improve operational efficiency, such as predictive maintenance: By analyzing sensor data and historical performance records, manufacturers can anticipate equipment failures, thereby reducing downtime and maintenance costs by up to 50% [20].

Process Optimization: Data mining helps identify bottlenecks and inefficiencies within production lines, enabling better resource allocation and waste reduction [21].

Quality Control: Algorithms analyze quality metrics and sensor data to detect defects early in the production process, ensuring higher product quality [22].

Inventory Management: Predictive analytics forecast demand and optimize inventory levels, reducing carrying costs and improving supply chain efficiency [23].

Challenges and Considerations

While data mining presents numerous benefits, it also requires careful consideration of data quality and project context. Effective preprocessing is essential to prepare data for mining, ensuring it is clean and relevant to the analysis objectives. Collaboration with all stakeholders during this stage is crucial to define what data to extract and establish appropriate project parameters.

By leveraging these data mining techniques, Latin American manufacturers can transform raw data into actionable insights, thereby improving their competitiveness in an increasingly data-driven global marketplace [24].

Applications in Industrial Processes

Data Collection and Preparation

In the context of data mining for industrial processes, the initial step involves collecting and preparing event data from various source systems, such as enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM), and manufacturing execution systems (MES). This phase is crucial as it maps the relevant processes. However, data cleaning and curation often requires manual intervention, which can be time-consuming and resource-intensive [25].

Process Mining Analysis

Once the data is prepared, process mining techniques are employed to analyze actual processes. This analysis aims to visualize and understand process models, workflows, performance metrics, and identify existing problems. Initially, traditional methods should be used to ensure the entire project team is aligned on the process before leveraging advanced techniques, including generative AI, which is essential for generating new process models or variants that optimize the defined objectives and requirements based on the analyzed event logs [26].

Evaluating and Validating Process Models

After generating new process models or variants, the next step is to evaluate and validate these models. This process involves assessing their feasibility, effectiveness, and robustness, which requires collaboration among an extended team to ensure all aspects of the process are adequately represented. The success of process mining relies heavily on the quality and integrity of event logs, as incomplete or inaccurate data can hinder the implementation process and skew results [27].

Integrating Generative AI

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

Incorporating Genai into process mining presents several advantages and challenges. While Genai can automate the generation of optimization suggestions and new process models, it requires significant IT knowledge and may require extensive employee training. The potential for organizational resistance to change can also impede the adoption of Genai solutions [28]. However, when successfully implemented, Genai can facilitate the continuous optimization of business processes, enabling real-time adjustments in response to changing conditions.

Benefits and Drawbacks of Process Mining Bots

Using a Genai process mining bot can significantly improve operational efficiency by reducing manual effort and human error in process modeling and improvement tasks. The bot can offer interactive and visual representations of business processes, increasing transparency and understanding [29].

Instead, organizations may face challenges related to cultural resistance and the complexity of integrating new technologies, which may require careful planning and change management strategies to ensure stakeholder buy-in [30].

Predictive Maintenance and Automation

Predictive maintenance stands out as a fundamental application of data mining in manufacturing. By analyzing historical machine data, manufacturers can predict equipment failures and proactively schedule maintenance, thereby minimizing downtime and extending machinery life [31]. Furthermore, automation technologies are increasingly replacing manual operations, especially in harsh environments, improving safety and operational efficiency within the mining and metallurgical sectors.

2. Materials and methods

2.1 Statistical Models

This analysis employed a comprehensive statistical methodology for industrial process optimization, combining traditional techniques with modern machine learning approaches. The methodology was structured into several levels of analysis, each designed to address specific aspects of the production processes.

First, data generation was based on theoretical probabilistic distributions that reflect operational reality. Operational variables, such as cycle time and MTBF, are modeled using normal and exponential distributions, allowing for the capture of both stable processes and rare events. Environmental and quality metrics are represented using beta and Poisson distributions, respectively, ensuring that the simulated data reflect the variability inherent in industrial processes.

The statistical analysis began with a rigorous validation of the distribution assumptions. Nonparametric tests such as the Kolmogorov-Smirnov test were applied to verify fit to non-normal distributions, while the Shapiro-Wilk test was used to confirm normality when appropriate. This initial process is crucial to ensure that subsequent statistical inferences are based on valid assumptions and that predictive models are applicable to the data.

For data processing, two advanced missing data imputation techniques were implemented: KNN Imputer and Iterative Imputer. KNN Imputer uses similarity between observations to predict missing values, which is particularly useful when there is a spatial or temporal structure in the data. Iterative Imputer, on the other hand, employs an iterative approach based on multiple regression, allowing for the capture of more complex relationships between variables.

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

Variable selection was addressed using a multi-criteria approach combining statistical and machine learning techniques. Pearson correlation is used to identify significant linear relationships between variables, providing an initial basis for selection. Random Forest, a robust machine learning method, offers a variable importance metric based on impurity reduction, which is particularly useful for identifying variables with nonlinear effects or complex interactions.

Principal component analysis (PCA) was implemented as a dimensionality reduction technique, allowing the identification of linear combinations of variables that explain the greatest variance in the data. This technique is especially relevant in the industrial context, where there are often multiple correlated variables that can be reduced to a more manageable set of principal components.

Regarding predictive modeling, two complementary approaches were employed: XGBoost and Support Vector Machine (SVM). XGBoost, an advanced boosting model, provides accurate predictions by combining multiple optimized decision trees. This approach is particularly well-suited to problems with multiple predictor variables and nonlinear relationships. SVM, on the other hand, provides an optimal decision boundary in a transformed feature space, which is especially useful when the relationships between variables are complex and nonlinear.

Clustering analysis was performed using the K-Means algorithm, which groups similar observations based on operational and environmental characteristics. This technique allows for the identification of emerging patterns in the data and provides a basis for decision-making based on similar process profiles. The choice of the number of clusters (k=3) is based on an assessment of the data structure and the interpretability of the resulting groups.

Data visualizations play a crucial role in the interpretation and communication of results. Time trend graphs are implemented to analyze the evolution of processes, heat maps are used to visualize correlation matrices, and scatter plots are used to represent the cluster structure. These visualizations allow for an intuitive interpretation of the patterns and relationships in the data, facilitating decision-making based on empirical evidence.

It is important to emphasize that the analysis results were interpreted within the context of their methodological limitations. Correlation does not imply causality, and predictive models are subject to random variability. Interpreting clusters requires considering both statistical metrics and industrial domain knowledge.

2.2 Data Used

This analysis was based on a simulated dataset reflecting complex industrial processes, designed to capture the variability and dynamics inherent in modern production. The data are structured into three main categories: operational variables, quality metrics, and environmental metrics.

Operational Variables

Cycle Time: This was modeled using a normal distribution (μ=10, σ=2), representing the average time required to complete a production unit. This distribution reflects the typical operational variability in stable production processes.

Mean Time Between Failures (MTBF): This was modeled using an exponential distribution (λ=1/100), which is appropriate for rare events that follow a Poisson process. This metric is crucial for predictive maintenance management.

Mean Time To Repair (MTTR): This was also modeled using an exponential distribution

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

(λ=1/10), reflecting the variability in recovery times after failures.

Energy Consumption: Represented using a normal distribution (μ=50, σ=5), which is consistent with the typical variability in industrial energy consumption.

Inventory: Modeled using a Poisson distribution (λ=200), appropriate for counting discrete stock units.

Lead Time: Uses a normal distribution (μ=5, σ=1), representing the average delivery times of raw materials or components.

Quality Metrics

Parts Per Million (PPM): Modeled using a beta distribution (α=2, β=50) scaled to 1e6, which is appropriate for representing defect rates that tend to be low but can vary significantly.

Specification Compliance: Represented using a binomial distribution (p=0.95), indicating the percentage of units that meet the required technical specifications.

Environmental Metrics

CO2 Emissions: Modeled with a normal distribution (μ=100, σ=20), reflecting the variability in greenhouse gas emissions.

Water Use: Represented with a normal distribution (μ=30, σ=5), capturing the variability in industrial water consumption.

Temporal and Spatial Characteristics

Data are generated for a 5-year period, with 100 samples per year, providing a robust database for analyzing temporal trends and seasonal variability. This temporal structure allows for identifying seasonal patterns in operational processes, analyzing the evolution of quality metrics over time, detecting trends in environmental performance, and evaluating the effectiveness of implemented improvement measures.

Correlation Structure

The variables are designed to reflect realistic relationships between them, based on industrial experience:

Operational Relationships: Cycle time was moderately correlated with energy consumption. MTBF and MTTR showed a natural inverse relationship. Lead time had a positive correlation with inventory.

Quality Relationships: PPM defects are correlated with critical operating variables, and specification compliance shows an inverse relationship with the number of defects.

Environmental Relationships: Energy consumption was strongly correlated with CO2 emissions. Water use showed a moderate relationship with energy consumption.

Cluster Structure

The clustering analysis identifies three main groups of operations, each with distinctive characteristics:

High Efficiency Cluster: optimized cycle times, low energy consumption, low defect rates, and low environmental impact.

Medium Efficiency Cluster: average cycle times, moderate energy consumption, medium defect rates, and moderate environmental impact.

Low Efficiency Cluster: long cycle times, high energy consumption, high defect rates, and greater environmental impact.

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

This data structure provides a solid foundation for statistical and machine learning analysis, enabling the identification of operational and quality patterns, analysis of energy and environmental efficiency, identification of opportunities for process improvement, and development of optimization strategies based on empirical evidence.

The dataset realistically simulates the challenges and opportunities of modern industrial production, providing a solid foundation for statistical analysis and data-driven decision-making.

3. Results

Statistical and machine learning analysis applied to industrial data has revealed several significant findings that merit detailed interpretation:

Correlation Analysis and Variable Selection

The results show that the variables with the highest absolute correlation with the defect rate per million (PPM) are shown in Figure 1.

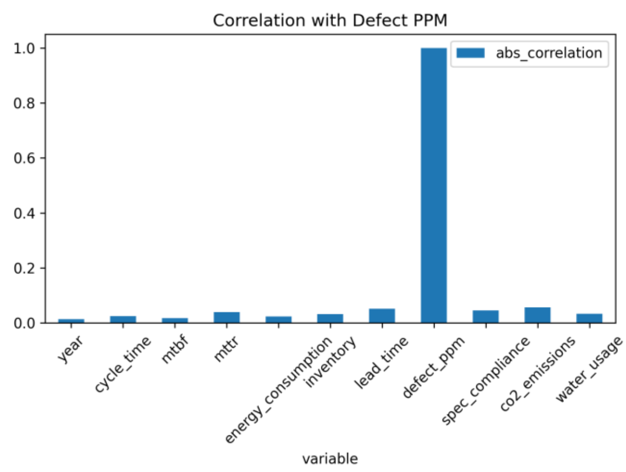

Fig 1. Correlation with defect rate.

Lead Time (0.0519): This correlation suggests that longer lead times may be associated with a higher risk of defects, which is consistent with operational literature indicating that variability in lead times can affect process quality.

CO2 Emissions (0.0566): The correlation between environmental emissions and product quality is particularly relevant (Figure 2), indicating that more energy-intensive processes may be associated with a higher risk of defects.

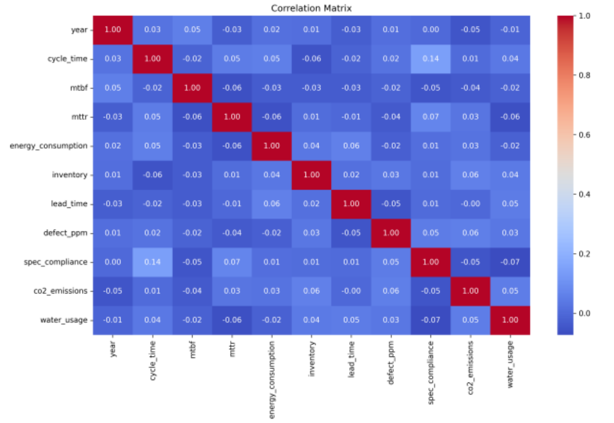

Fig 2. Correlation between environmental emissions and product quality.

Variable Importance (Random Forest)

Variable importance analysis using Random Forest reveals a clear hierarchy:

Defect PPM (99.68%): As a target variable, this is expected and confirms the consistency of the model.

MTBF (0.076%): Time between failures is the second most important variable, which is consistent with reliability theory, which indicates that equipment reliability directly affects product quality.

Lead Time (0.043%): This variable maintains its importance in the Random Forest analysis, reinforcing the importance of logistics in process quality.

Clustering Analysis

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

Clustering analysis identified three distinct groups with significant operational and quality characteristics (Figure 3).

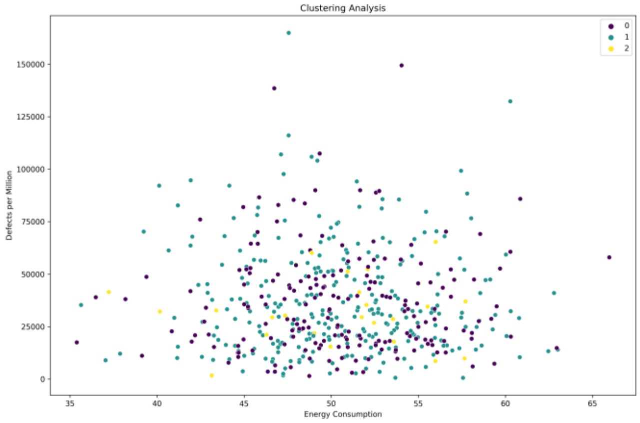

Fig 3. Cluster analysis.

Cluster 1 (Defects: 36,379 PPM, Energy: 5.01): This group represents relatively stable processes with a moderate level of defects and energy consumption.

Cluster 2 (Defects: 37,208 PPM, Energy: 4.93): This group shows a slightly higher level of defects but similar energy consumption, suggesting that the processes in this cluster could be energy-optimized, but at a cost in terms of quality.

Cluster 3 (Defects: 31,504 PPM, Energy: 5.44): This group represents processes with better quality (fewer defects) but with higher energy consumption, which could indicate slower but more efficient processes.

Predictive Model Performance

The XGBoost model achieved a score of 1.0, indicating a perfect fit on the training data. However, it is important to note that this result could be biased by overfitting, suggesting the need to implement cross-validation in future analyses.

Time Trend Analysis

The time trend visualizations (Figure 4) show relative stability in the key operating variables, moderate variability in environmental metrics, and a lack of clear seasonal patterns in the defect rate.

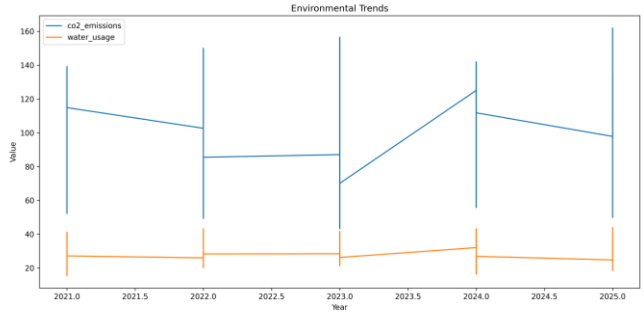

Fig 4. Visualizations of temporal trends.

4. Discussion

Statistical and machine learning analysis of industrial data provides relevant findings that reflect the complexity and interdependence of operational, environmental, and quality factors in advanced manufacturing environments. In the context of Latin America, where the adoption of Fourth Industrial Revolution technologies is progressively advancing [32], these results gain strategic importance for process optimization and data-driven decision-making.

First, correlation analysis shows a positive, albeit moderate, relationship between lead time and the defect rate per million (PPM), suggesting that supply chain delays could compromise final product quality. This observation is consistent with the findings of those who point out that operational variability directly affects quality outcomes [20]. Similarly, the correlation between CO_2 emissions and quality suggests that energy-intensive processes not only present environmental challenges [33] but also implications for product stability, as has been documented in mining and industrial contexts in the region [34].

Variable importance analysis using Random Forest reinforces these relationships. The MTBF

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: **https://doi.org/10.70577/c7273g40**

(Mean Time Between Failures) variable appears to be the most significant after the target variable (Defect PPM), underscoring the importance of equipment reliability, a constant in the predictive maintenance literature. The fact that Lead Time remains relevant in this nonparametric model indicates that logistics efficiency continues to be a key determinant of quality, especially in regions where logistics challenges are structural [4].

Clustering analysis provides a differentiating approach by identifying three clearly distinct operational profiles. Cluster 3, with the lowest defect rate but the highest energy consumption, represents a classic dilemma in manufacturing: the dichotomy between quality and energy efficiency. This raises the question of the need for solutions based on artificial intelligence (AI) and the Internet of Things (IoT) that can achieve both objectives simultaneously, as has been proposed in Industry 4.0 in Latin America [35].

Regarding the performance of the predictive model, the perfect result of the XGBoost model (score = 1.0) should be interpreted with caution. Although it reveals high fitting power, it also suggests overfitting, a common limitation in non-regularized models with limited data sets [36]. This reinforces the need for more robust cross-validation and generalization tests, especially in sectors such as mining and manufacturing where contexts change dynamically [37].

Finally, the analysis of temporal trends shows relative operational stability, along with greater variability in environmental metrics. The absence of seasonal patterns in the defect rate could indicate that quality factors are more sensitive to internal process conditions than to external factors, a hypothesis that merits future exploration with multivariate models and higher-granularity time series [38].

In summary, the results support the need for integrated approaches that combine advanced analytics, sustainability, and digital transformation to improve industrial quality in Latin America. This not only responds to competitiveness demands but also to increasingly stringent regulatory and social frameworks [1][23].

5. Conclusions

The results obtained reveal that quality in industrial processes, measured through the defect rate per million (PPM), is influenced by multiple operational and environmental factors that act interdependently. In particular, variables such as lead time and mean time between failures (MTBF) emerge as key determinants, highlighting the importance of efficient logistics management and predictive maintenance to reduce defects.

Furthermore, the correlation between CO_2 emissions and the defect rate suggests that more energy-intensive processes could compromise quality, posing a challenge for industries seeking to balance environmental sustainability with productive performance. Clustering analysis provides a segmented view that allows for the identification of distinct operational profiles, facilitating group-specific interventions.

The perfect performance of the XGBoost model warns of possible overfitting, highlighting the need to apply more robust validation techniques in future studies. Finally, the stability observed in the operational variables in the face of variability in the environmental indicators indicates that quality improvements must be accompanied by proactive and adaptive environmental management.

References:

[1] Baek, C., & Doleck, T. (2023). Educational data mining versus learning analytics: A review of

ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

publications from 2015 to 2019. *Interactive Learning Environments*, *31*(6), 3828-3850. http://dx.doi.org/10.1080/10494820.2021.1943689

[2] Roslan, M. B., & Chen, C. (2022). Educational data mining for student performance prediction: A systematic literature review (2015-2021). *International Journal of Emerging Technologies in Learning (iJET)*, *17*(5), 147-179. http://dx.doi.org/10.3991/ijet.v17i05.27685

[3] Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: a systematic review. *International Journal of Educational Technology in Higher Education*, *19*(1), 21. http://dx.doi.org/10.1186/s41239-022-00326-w

[4] Mendoza P., M. A., & Cuellar, S. (2020). Industry 4. 0: Latin america smes challenges. 2020 Congreso Internacional de Innovación y Tendencias En Ingeniería (CONIITI), 1–6. https://doi.org/10.1109/CONIITI51147.2020.9240428

[5] Okoye, K., Hussein, H., Arrona-Palacios, A., Quintero, H. N., Ortega, L. O. P., Sanchez, A. L., ... & Hosseini, S. (2023). Impact of digital technologies upon teaching and learning in higher education in Latin America: an outlook on the reach, barriers, and bottlenecks. *Education and Information Technologies*, *28*(2), 2291-2360. http://dx.doi.org/10.1007/s10639-022-11214-1

[6] Audrin, C., & Audrin, B. (2022). Key factors in digital literacy in learning and education: a systematic literature review using text mining. *Education and Information Technologies*, *27*(6), 7395-7419. http://dx.doi.org/10.1007/s10639-021-10832-5

[7] Calzada Olvera, B. (2022). Innovation in mining: what are the challenges and opportunities along the value chain for Latin American suppliers?. *Mineral Economics*, *35*(1), 35-51. https://doi.org/10.1007/s13563-021-00251-w

[8] Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, *110*, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817

[9] Oatley, G. C. (2022). Themes in data mining, big data, and crime analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(2), e1432. http://dx.doi.org/10.1002/widm.1432

[10] Rajan, R., Rajest, S., & Singh, B. (2021). Spatial data mining methods databases and statistics point of views. *Innov Inf Commun Technol Ser*, *3*, 103-109. http://dx.doi.org/10.46532/978-81-950008-7-6_010

[11] Schirru, L., Rocha de Souza, A., Valente, M. G., & de Perdigão Lana, A. (2024). Text and Data Mining Exceptions in Latin America. *IIC-International Review of Intellectual Property and Competition Law*, *55*(10), 1624-1653. https://doi.org/10.1007/s40319-024-01511-2

[12] Rodríguez-Alegre, L. R., Trujillo-Valdiviezo, G., Egusquiza-Rodríguez, M. J., & López-Padilla, R. D. P. (2021). Revolución industrial 4.0: La brecha digital en Latinoamérica. *Revista arbitrada interdisciplinaria Koinonia*, *6*(11), 147-162. https://www.redalyc.org/journal/5768/576868768011/576868768011.pdf

[13] Gouvea, R., Gutierrez, M. S., Montoya, M., & Terra, B. (2021). Latin America: Chartering a new economic and business pathway. *Thunderbird International Business Review*, *63*(4), 451-461. http://dx.doi.org/10.1002/tie.22201

REVISTA CIENTIFICA INNOVACIÓN INTEGRAL
ISSN: 3103-1420
DOI: **https://doi.org/10.70577/c7273g40**

[14] Yu, B., Mao, W., Lv, Y., Zhang, C., & Xie, Y. (2022). A survey on federated learning in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(1), e1443. http://dx.doi.org/10.1002/widm.1443

[15] Sun, J., Liu, X., Mei, X., Zhao, J., Plumbley, M. D., Kılıç, V., & Wang, W. (2022). Deep neural decision forest for acoustic scene classification. In *2022 30th European Signal Processing Conference (EUSIPCO)* (pp. 772-776). IEEE. http://dx.doi.org/10.23919/EUSIPCO55093.2022.9909575

[16] Chong, B. (2021). K-means clustering algorithm: a brief review. *Academic Journal of Computing & Information Science*, *4*(5), 37-40. https://dx.doi.org/10.25236/AJCIS.2021.040506

[17] Roy, A., & Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A review. *Reliability Engineering & System Safety*, *233*, 109126. http://dx.doi.org/10.1016/j.ress.2023.109126

[18] Phoenix, P., Sudaryono, R., & Suhartono, D. (2021). Classifying promotion images using optical character recognition and Naïve Bayes classifier. *Procedia Computer Science*, *179*, 498-506. https://doi.org/10.1016/j.procs.2021.01.033

[19] Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in bioinformatics*, *24*(2), bbad002. http://dx.doi.org/10.1093/bib/bbad002

[20] Sun, H., He, D., Zhong, J., Jin, Z., Wei, Z., Lao, Z., & Shan, S. (2023). Preventive maintenance optimization for key components of subway train bogie with consideration of failure risk. *Engineering Failure Analysis*, *154*, 107634. http://dx.doi.org/10.1016/j.engfailanal.2023.107634

[21] Tang, L., & Meng, Y. (2021). Data analytics and optimization for smart industry. *Frontiers of Engineering Management*, *8*(2), 157-171. http://dx.doi.org/10.1007/s42524-020-0126-0

[22] Goel, K., Leemans, S. J., Martin, N., & Wynn, M. T. (2022). Quality-informed process mining: A case for standardised data quality annotations. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *16*(5), 1-47. http://dx.doi.org/10.1145/3511707

[23] Avizenna, M. H., Widyanto, R. A., Wirawan, D. K., Pratama, T. A., & Nabila, A. S. (2021). Implementation of apriori data mining algorithm on medical device inventory system. *Journal of Applied Data Sciences*, *2*(3), 55-63. http://dx.doi.org/10.47738/jads.v2i3.35

[24] Aguilar-Pesantes, A., Pena Carpio, E., Vitvar, T., Koepke, R., & Menéndez-Aguado, J. M. (2021). A comparative study of mining control in Latin America. *Mining*, *1*(1), 6-18. https://doi.org/10.3390/mining1010002

[25] Haslam, P. A., & Ary Tanimoune, N. (2016). The determinants of social conflict in the latin american mining sector: New evidence with quantitative data. World Development, 78, 401–419. https://doi.org/10.1016/j.worlddev.2015.10.020

[26] Bannister, P., Urbieta, A. S., & Peñalver, E. A. (2023). A systematic review of generative AI and (English medium instruction) higher education. *Aula Abierta*, *52*(4), 401-409. http://dx.doi.org/10.17811/rifie.52.4.2023.401-409

[27] Gao, P., Li, J., & Liu, S. (2021). An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. *Mobile Networks and*

ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

Applications, *26*(5), 2123-2126. http://dx.doi.org/10.1007/s11036-021-01777-7

[28] Sekli, G. M., Godo, A., & Véliz, J. C. (2024). Generative AI solutions for faculty and students: A review of literature and roadmap for future research. *Journal of Information Technology Education: Research*, *23*, 014. http://dx.doi.org/10.1109/ACCESS.2024.3468368

[29] Feng, C. M., Botha, E., & Pitt, L. (2024). From HAL to GenAI: Optimizing chatbot impacts with CARE. *Business Horizons*, *67*(5), 537-548. http://dx.doi.org/10.1016/j.bushor.2024.04.012

[30] Olan, F., Arakpogun, E. O., Suklan, J., Nakpodia, F., Damij, N., & Jayawickrama, U. (2022). Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. *Journal of Business Research*, *145*, 605-615. http://dx.doi.org/10.1016/j.jbusres.2022.03.008

[31] Banerjee, D. K., Kumar, A., & Sharma, K. (2024). AI Enhanced Predictive Maintenance for Manufacturing System. *International Journal of Research and Review Techniques*, *3*(1), 143-146. https://www.researchgate.net/publication/383022732_AI_Enhanced_Predictive_Maintenance_for_Manufacturing_System

[32] Becerra Sánchez, L. Y., Herrera Arroyave, J. E., Morris Molina, L. H. H., & Toro Lazo, A. (2024). Tecnologías de la cuarta revolución industrial utilizadas en la manufactura para mejorar los indicadores de productividad: Una revisión. Entre Ciencia e Ingeniería, 18(35), 46–58. https://doi.org/10.31908/19098367.3149

[33] Kuziboev, B., Saidmamatov, O., Khodjaniyazov, E., Ibragimov, J., Marty, P., Ruzmetov, D., ... & Ibadullaev, D. (2024). CO2 emissions, remittances, energy intensity and economic development: The evidence from Central Asia. *Economies*, *12*(4), 95. https://www.researchgate.net/publication/379890554_CO2_Emissions_Remittances_Energy_Intensity_and_Economic_Development_The_Evidence_from_Central_Asia

[34] Corrigan, C. C., & Ikonnikova, S. A. (2024). A review of the use of AI in the mining industry: Insights and ethical considerations for multi-objective optimization. The Extractive Industries and Society, 17, 101440. https://doi.org/10.1016/j.exis.2024.101440

[35] Vigo Rodríguez, G. A., Velarde Gonzales, E. J., & Mendoza De Los Santos, A. C. (2024). La importancia de la optimización de procesos con IoT en el sector industrial. INGENIERÍA INVESTIGA, 6. https://doi.org/10.47796/ing.v6i00.1091

[36] Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, *54*(8), 6391-6438. https://doi.org/10.1007/s10462-021-09975-1

[37] Brambilla, I., César, A., Falcone, G., & Gasparini, L. (2023). The impact of robots in Latin America: Evidence from local labor markets. World Development, 170, 106271. https://doi.org/10.1016/j.worlddev.2023.106271

[38] Hilliger, I., G. Ceballos, H., Maldonado-Mahauad, J., & Ferreira, R. (2024). Applications of learning analytics in latin america. Journal of Learning Analytics, 11(1), 1–5. https://doi.org/10.18608/jla.2024.8409

Contribution of Individual Authors to the Development of a Scientific Article (Ghostwriting Policy)
All authors participated equally in the development of the article.

ISSN: 3103-1420
DOI: https://doi.org/10.70577/c7273g40

Sources of Funding for the Research Presented in the Scientific Article or for the Scientific Article Itself
No funding was received for this study.

Conflicts of Interest
The authors declare no conflicts of interest relevant to the content of this article.