IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Journal Européen des Systèmes Automatisés
Vol. 58, No. 1, January, 2025, pp. 21-29
Journal homepage: http://iieta.org/journals/jesa

# A Data-Driven Predictive Maintenance Approach for Industry 4.0 Using LSTM with Cross-Validation and the IDAIC Framework

Hicham Raffak[1*], Hicham Ghatous[2], Mohamed Mansouri[2], Abdallah Lakhouilii[1]

[1] Faculty of Sciences and Techniques, Hassan First University of Settat, Settat 26000, Morocco
[2] National School of Applied Sciences, Hassan First University of Settat, Berrechid 26100, Morocco

Corresponding Author Email: h.raffak@uhp.ac.ma



https://doi.org/10.18280/jesa.580103

**Received:** 10 December 2024
**Revised:** 14 January 2025
**Accepted:** 21 January 2025
**Available online:** 31 January 2025

***Keywords:***
*maintenance 4.0, predictive maintenance, IDAIC, LSTM, cross-validation*

**ABSTRACT**

In the current context of intense competition, industrial maintenance plays a crucial role in ensuring the performance and resilience of companies. It ensures the continuous availability of equipment, which is essential to avoid unplanned downtime that can lead to significant economic losses. Moreover, maintenance improves production quality by reducing failures and manufacturing defects, and by optimizing the costs associated with maintenance interventions. Predictive maintenance, which is a fundamental part of Industry 4.0, allows for anticipating failures before they occur by leveraging real-time data to predict malfunctions and plan the necessary actions. This not only reduces unplanned downtime but also lowers the overall cost of repairs and equipment replacements. However, data acquisition and processing present major challenges for data science project managers, as they require appropriate frameworks and approaches tailored to each problem and context. This study proposes an innovative solution with a predictive maintenance model developed using the industrial data analysis improvement cycle (IDAIC) approach, specifically designed for industrial maintenance projects. By using a deep learning algorithm, long short-term memory (LSTM), and techniques such as early stopping, the model was applied to the data of a plastic injection molding machine and achieved impressive results. With an $R^2$ of 96% and an MSE of 99%, it presents itself as a powerful decision-support tool for industrial maintenance.

## 1. INTRODUCTION

Manufacturing companies and equipment manufacturers are facing two major transformations that are redefining their activities: digital transformation and sustainable development. Digital transformation, often associated with the Fourth Industrial Revolution, known as "Industry 4.0," has introduced intelligent production systems capable of monitoring physical processes and making optimized real-time decisions through the interconnection of humans, machines, and sensors [1-3].

Currently, industrial maintenance primarily relies on reactive or preventive strategies, while the adoption of predictive approaches remains limited and is generally reserved for critical situations. These traditional approaches fail to fully leverage the vast volumes of data generated on production sites or emerging technologies such as the Internet of Things (IoT), cloud computing, or augmented reality. However, the paradigm is shifting, positioning maintenance as a critical strategic factor for improving productivity and optimizing the performance of industrial systems [4, 5].

This evolution has led to the development of new maintenance approaches, notably Prognostics and Health Management (PHM) and Condition-Based Maintenance (CBM), which leverage operational data to detect anomalies in asset behavior [6]. These approaches enable a shift from reactive maintenance to proactive management, reducing unexpected downtime and improving equipment reliability. With the increasing complexity of machinery and their growing criticality in terms of reliability and availability, it has become imperative to mitigate the risks and consequences of unexpected interruptions in an increasingly digitized production environment [7].

According to reference [8], smart maintenance is based on proactive and adaptive management, focusing on data collection, analysis, and visualization, as well as the continuous improvement of decision-making processes [9]. With the emergence of Industry 4.0, new tools and methods are required to meet the demands of smart factories (Figure 1) [10].


**Figure 1.** Components of Industry 4.0 technologies

In this context, this work proposes the development of a predictive maintenance model following the IDAIC methodology and based on LSTM networks, with cross-validation to ensure the robustness and reliability of predictions. This model leverages advanced data analysis to continuously monitor systems, detect disturbances early, and optimize maintenance interventions. By integrating machine learning technologies, the model aims to provide intelligent decision support, facilitating interventions and enhancing asset management in alignment with Industry 4.0 principles.

## 2. EVOLUTION OF INDUSTRIAL MAINTENANCE

### 2.1 Predictive maintenance

Industry 4.0 drives the creation of fully integrated manufacturing ecosystems powered by real-time data analytics. Within this framework, maintenance practices are undergoing a paradigm shift: outdated, expense-heavy breakdown-based interventions are being phased out in favor of proactive, algorithm-guided strategies. These advanced methods, often termed cognitive maintenance systems, prioritize forecasting failures before they occur.

Throughout successive industrial revolutions as shown in Figure 2, maintenance approaches have gradually evolved and are now a continuous process [11]. Indeed, for years, decision-makers have been shifting from corrective maintenance towards preventive maintenance, defined as a series of actions intended to prevent and reducing the risks of failures, as well as the duration and number of shutdowns, thus optimizing the overall maintenance costs [12].

Predictive maintenance is the most innovative form of maintenance. It ensures extended lifespan and high reliability of equipment, while providing more eco-friendly and cost-effective solutions [13]. Proactive maintenance, which involves solving problems by tracing their origin, has recently become more popular and is a highly effective complementary method when combined with predictive maintenance [14]. Advanced calculation and Visualization tools, built on the latest technological innovations, have become essential components of digital transformation within Industry 4.0 [15].

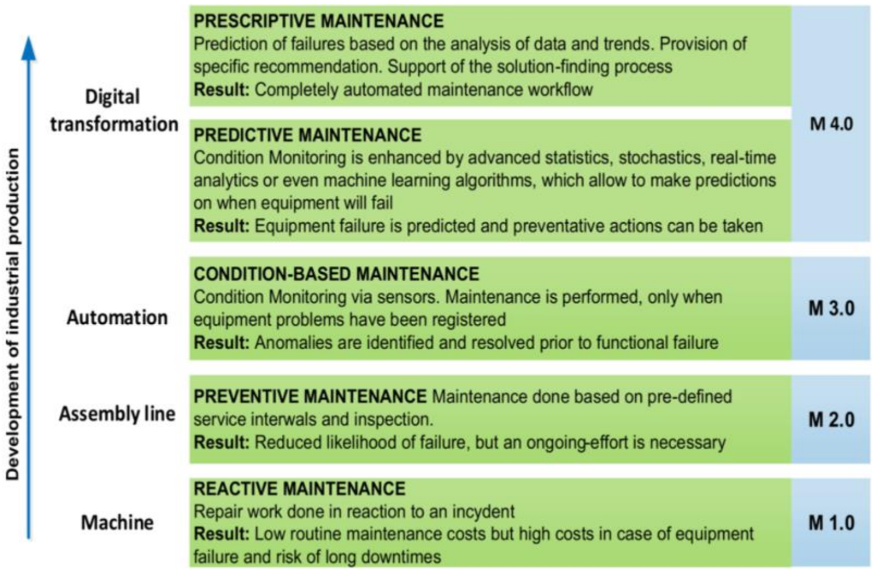


**Figure 2.** Evolution of maintenance

### 2.2 Predictive maintenance tools

Several preventive maintenance tools and techniques have been developed in the literature, aiming to provide frameworks for the successful implementation of this strategy. The most recent is the CPS (Cyber-Physical Systems) presented in five stages (Figure 3) [16].

Another framework is the one called IIoT (Industrial Internet of Things). It forms an interconnected framework that unifies all cyber-physical systems, thus facilitating the automatic collection and retrieval of the large volume of data flow. By monitoring this data in real-time and analyzing it through artificial intelligence or machine learning models, or even through statistical models, this constitutes the concept of Big Data [17].

### 2.3 Predictive maintenance models

Two main preventive maintenance models are found in the literature. The first model is CBM (Condition-Based Maintenance) (Figure 4), in which maintenance decisions are made based on the current or future condition of the equipment operates through three fundamental phases: gathering data, processing that information, and making informed decisions [18]. The goal is to monitor indicators related to the condition of the equipment, which trigger an alarm once a deterioration level is reached [19].

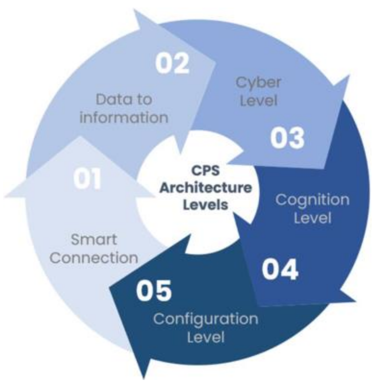


**Figure 3.** CPS steps

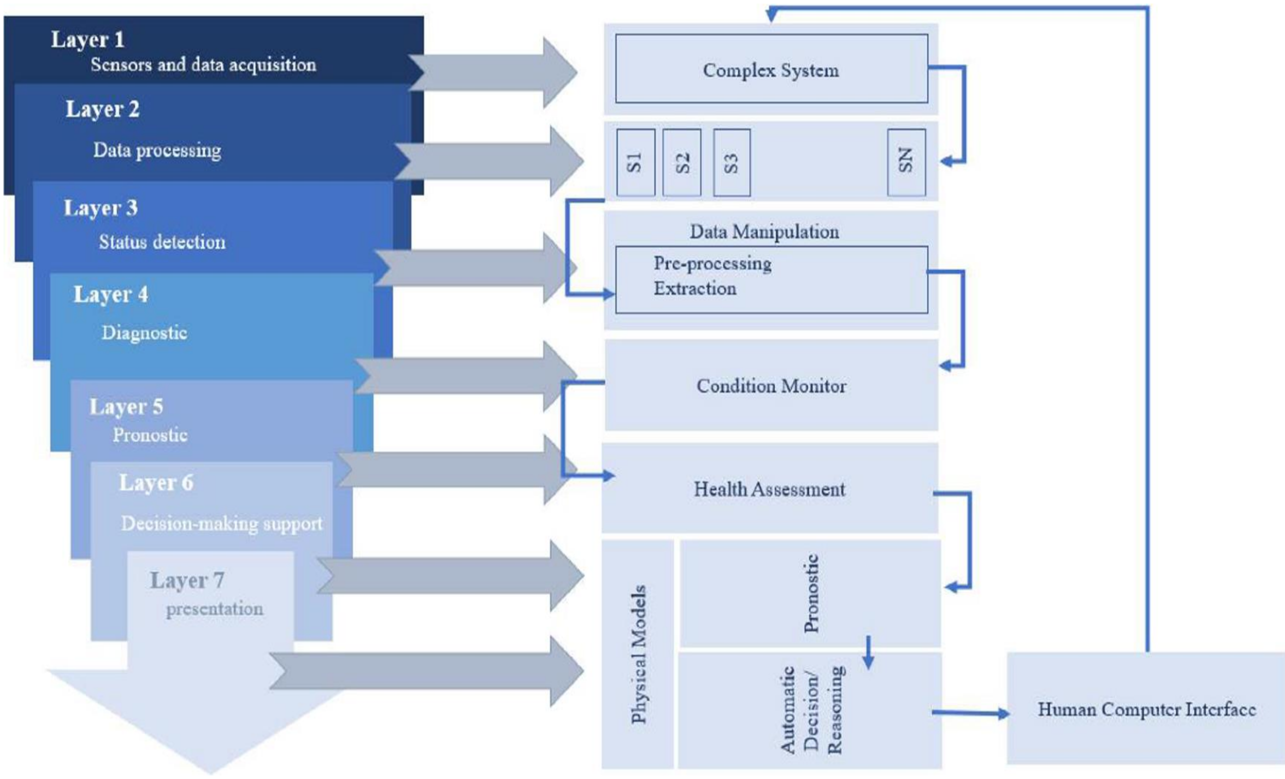


**Figure 4.** CBM levels

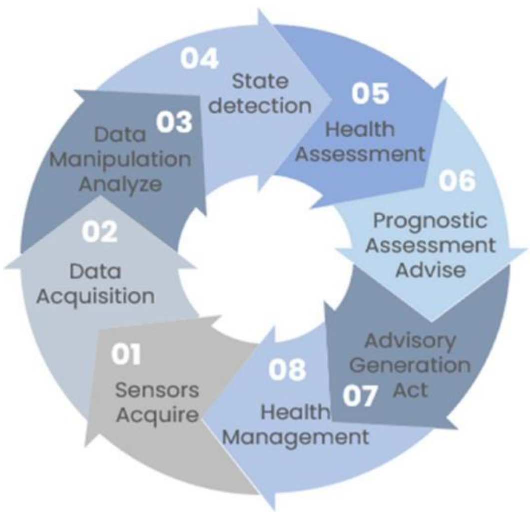


**Figure 5.** PHM steps

Another common model is the Prognostics and Health Management (PHM), which originated in the early 1990s. The objective is the dynamic monitoring of a system's state, with the process being considered key. It evaluates future degradation, offering the advantage of describing the different maintenance scenarios, whether preventive or predictive [20]. Figure 5 presents the 8 stages of PHM [21].

PHM relies on techniques that fall into three main categories: data-driven methods, model-based methods, and hybrid methods that leverage the strengths of both approaches.

## 3. PREDICTIVE MAINTENANCE WORKFLOW

To successfully implement a predictive maintenance project, it is essential to follow key steps based on the 4U framework: Understanding, Utilizing, Upgrading, and Updating [22].

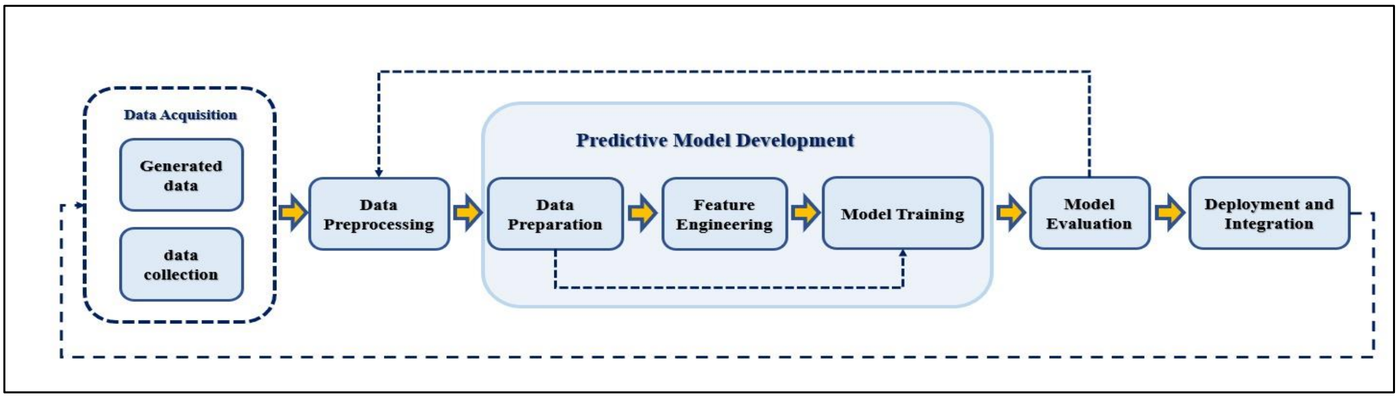


**Figure 6.** Adapted maintenance workflow

The process begins with a comprehensive Understanding of the project's needs, including the analysis of systems, the definition of critical parameters, and the identification of potential failures. Next, the strategic Utilizing of data collected through IoT technologies such as sensors, detectors, and collection systems is crucial. This includes the comprehension, mastery, and selection of relevant data, data cleaning, handling missing or outlier values, and extracting new features through data engineering. The newly formed dataset must then be optimized and improved (Upgrading) by structuring the data into a format suitable for modeling, ensuring its quality and relevance. Finally, the regular Updating of data and models ensures that feedback is integrated, enabling increasingly effective interventions aligned with maintenance objectives. This process (Figure 6) also allows for the continuous feeding of the database, thereby enhancing the understanding and performance of intelligent algorithms.

## 4. EVOLUTION OF MAINTENANCE DATA ACROSS THE ERAS OF MAINTENANCE

Data plays a crucial role in the implementation of scientific practices in maintenance. Its management and utilization have evolved in parallel with technological advancements across the different eras of maintenance, as described in Table 1 [23]. This evolution demonstrates how digitization and artificial intelligence are revolutionizing maintenance data management, enabling more precise analyses, optimized decision-making, and proactive maintenance in the industry 4.0 environment.

**Table 1.** Comparison of maintenance data in different maintenance ages

| Aspect | Maintenance 1.0 | Maintenance 2.0 | Maintenance 3.0 | Maintenance 4.0 |
|---|---|---|---|---|
| Data source | Operators' experience | Maintainers and machines | Operators, maintainers, systems | Operators, maintainers, systems, OEM |
| Data collection | Manual | Manual | Semi-automated | Automated via sensors and IoT |
| Data storage | Operators' memory | Written documents | Databases | Cloud services |
| Data analysis | Arbitrary | Reliability theory | Classical algorithms | Advanced algorithms (AI, ML, etc.) |
| Data transfer | Verbal communication | Written documents | Digital files | Digital files |
| Data management | N/A | Human operators | Information systems | Cloud and artificial intelligence |

## 5. PROPOSED APPROACH

There are different approaches to conducting a data science project, with the most popular being CRISP-DM (Cross-Industry Standard Process for Data Mining) (Figure 7) [24].

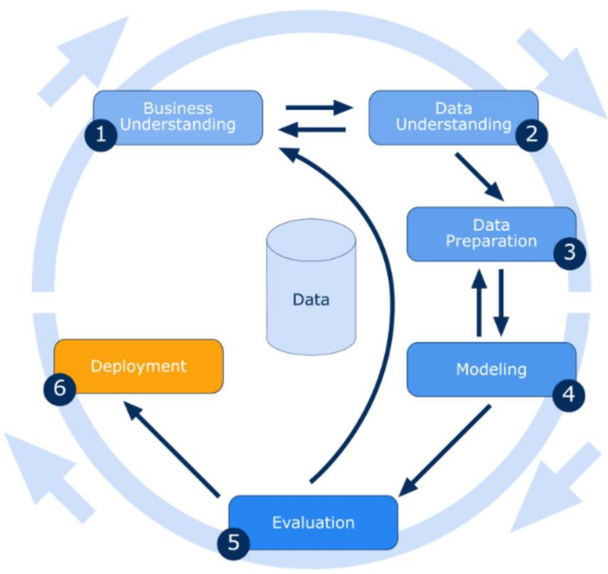


**Figure 7.** CRISP-DM approach

While it is effective for solving visible and goal-oriented problems, it lacks the necessary adaptations to address invisible problems that require exploratory approaches. The approach adopted in our study for developing a predictive maintenance model is the IDAIC (Figure 8), a framework developed and adapted based on the CRISP-DM methodology, specifically oriented towards proactive maintenance projects.

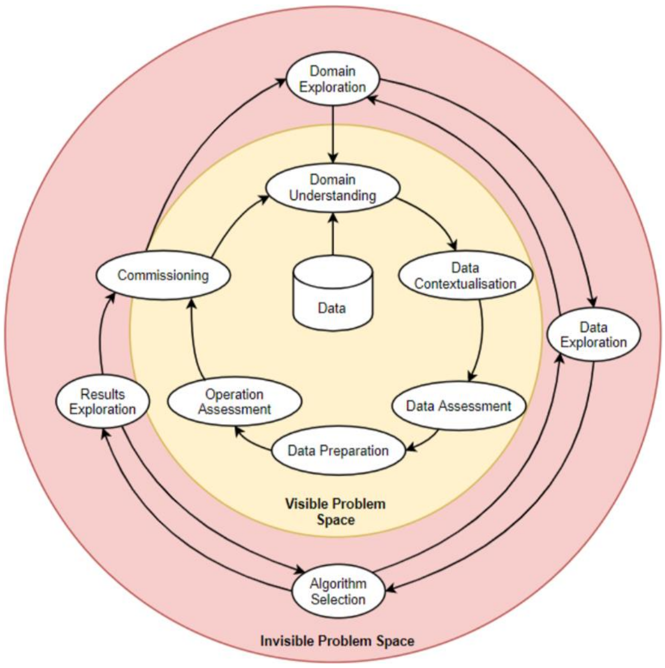


**Figure 8.** IDAIC

The IDAIC framework consists of key phases, including:

• Domain understanding, which involves collecting information on information systems, people, and equipment;

• Data contextualization, to evaluate the alignment of the data with project objectives;

• Data evaluation, focusing on assessing its usability, completeness, and objectivity;

• Data preparation, which involves cleaning and structuring the data for effective analysis.

• Operational assessment addresses visible problems, while the commissioning phase aligns systems with specifications to prepare for the transition to predictive and proactive maintenance.

• The exploration of invisible problems leverages advanced techniques such as exploratory data analysis and machine learning to uncover hidden patterns and facilitate the proactive planning of maintenance activities [25].

## 6. RESULTS AND DISCUSSION

We applied this approach to develop a predictive maintenance process for an injection molding machine dedicated to manufacturing parts for the automotive industry. We will subsequently detail the various results obtained in relation to the IDAIC model.

### 6.1 Data collection and preprocessing

The data acquisition process for predictive maintenance is divided into several steps, ranging from the initial collection of raw data from machines to its preparation for analysis. The key steps in the process are described in Figure 9.

The communication interfaces and parameters required for data exchange between injection molding machines and IT systems are defined in our case by the Euromap 63 standard. This standard is based on the OPC-UA (OLE for Process Control – Unified Architecture) communication protocol, which is widely used in industrial environments [26]. It facilitates the integration of equipment into the production environment, as well as process monitoring and manufacturing optimization, enabling injection molding machines to transmit real-time production data to a server or management system.

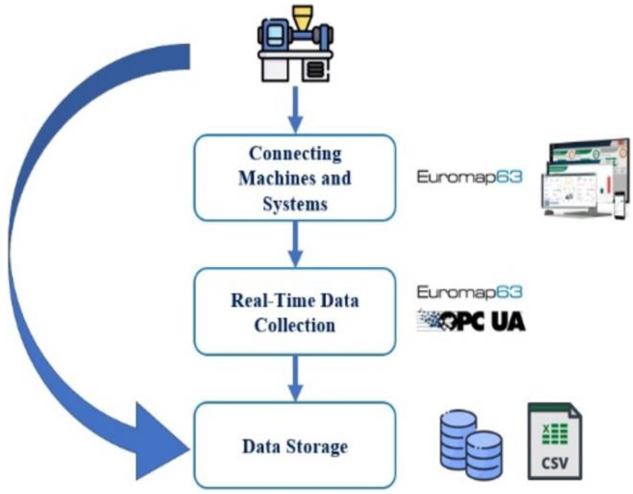


**Figure 9.** Data collection

Once the data is stored, it will be loaded into the Jupyter Notebook data analysis environment, where the data preprocessing phase will begin (Figure 10).

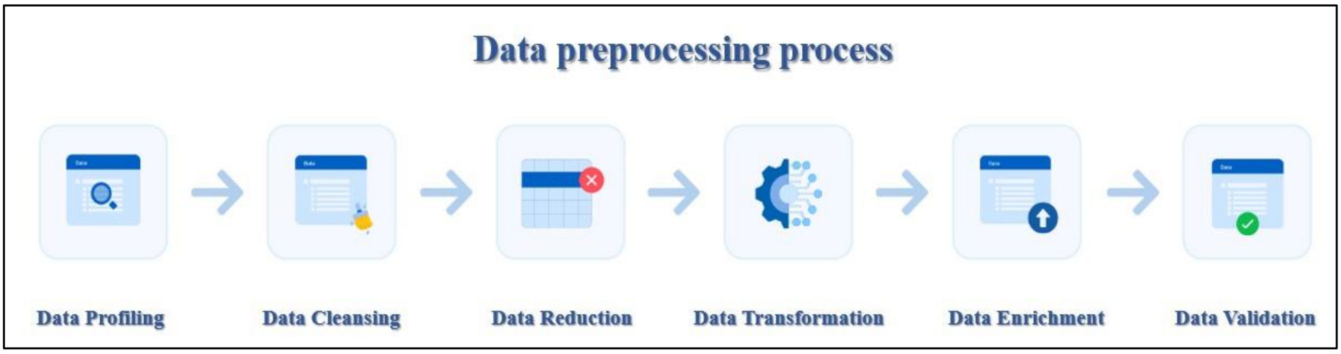


**Figure 10.** Preprocessing steps

In this phase, the data is profiled, cleaned, transformed, and enriched by removing duplicates and handling missing values to ensure data integrity. For example, the column "ActStsMach" encodes status information as strings. This information was split into multiple sub-columns, and a data mapping to numerical values was performed to ensure proper interpretation by machine learning algorithms.

### 6.2 Creation of the "Failure" column

The "Failure" column was created after merging two databases: the failure log database and the sensor data storage database. A 5-minute tolerance was applied to accurately associate failure events with sensor measurements.

To better understand the incidents, we also visualized the frequency of failures and types of stoppages using bar charts (Figure 11). The count of values in the "Failure" variable indicates that a significant number of failures were recorded during the observed period.

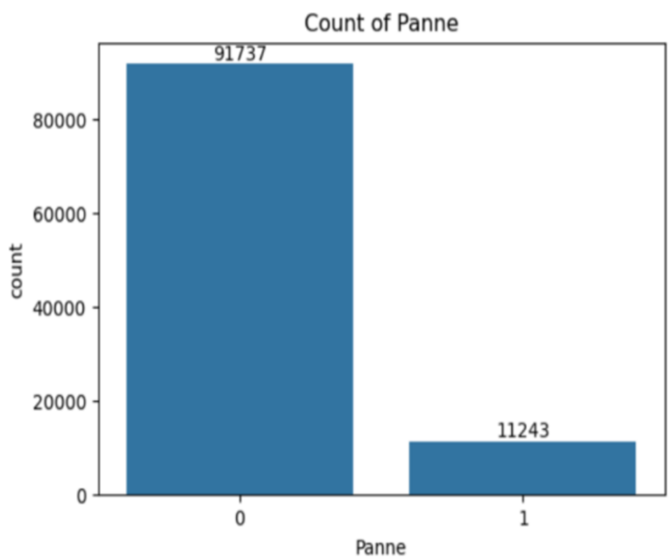


**Figure 11.** Counting of failures

The visualization of failure types from the "Description" column (Figure 12) allowed us to identify the most frequent types of failures, which may indicate recurring issues.

The correlation matrix in (Figure 13) was generated to evaluate the relationships between the different quantitative variables in our database.

In our analysis, the correlations between parameters are generally weak, suggesting that there are no significant linear relationships between most variables in our database.

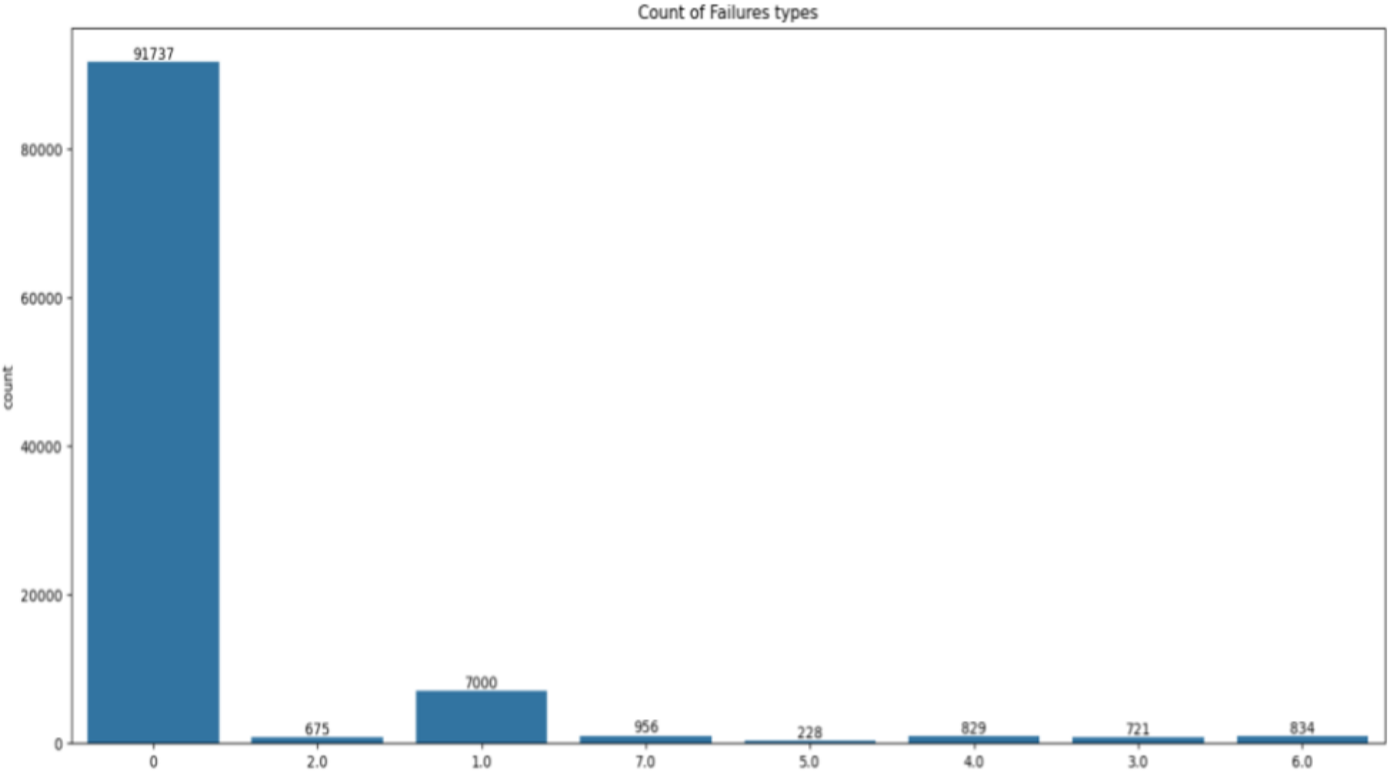


**Figure 12.** Counting of failure types

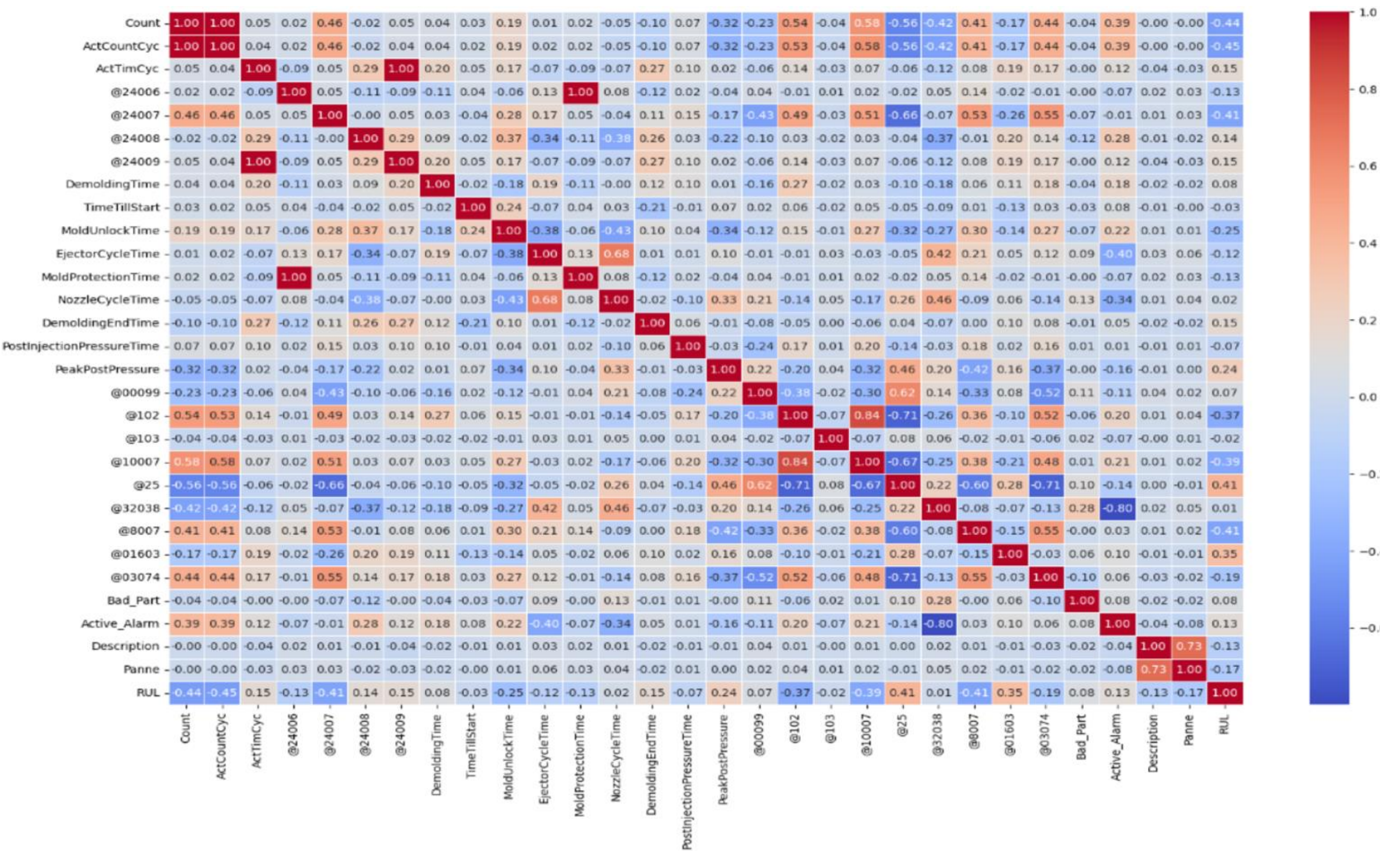


**Figure 13.** The correlation matrix

### 6.3 LSTM model with cross-validation for failure prediction

As part of this study, we implemented a failure prediction model on industrial data using a LSTM neural network (Figure 14). The model was evaluated in terms of recall, precision and F1 score for each class. The MSE and $R^2$ score were calculated to demonstrate the model's performance in terms of errors and data variance. The confusion matrix and the ROC curve will then be visualized to confirm the excellent performance of the model. The "early stopping" technique was used in our model to prevent overfitting. It stops the training if the validation loss stops decreasing after a certain number of epochs, ensuring the model's generalization.

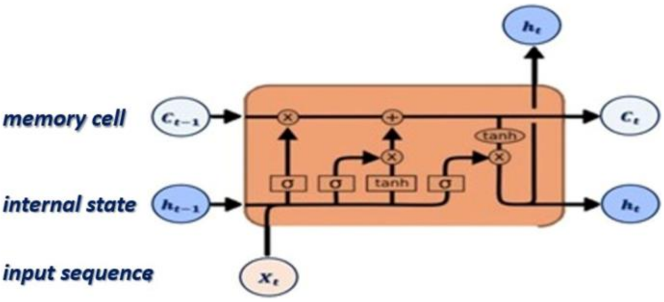


**Figure 14.** Compact form of an LSTM model

To evaluate the performance of the LSTM model, a 5-fold stratified cross-validation was used. It is often favored in the context of model evaluation in machine learning, as it provides an optimal compromise between bias and variance. It allows for a robust evaluation of the model while limiting the computational costs associated with a higher number of folds, such as 10. This number of folds is supported by common practices and empirical recommendations in the scientific literature, being considered a good compromise in most application cases. The objective was to ensure that the model generalizes well to unseen data and to avoid potential bias due to imbalanced class distribution in the training and validation sets. In each fold, the model was trained and validated, and the average performance was calculated across all folds.

The model demonstrated outstanding performance, achieving an overall accuracy of 99%. The $R^2$ score of 96% shows that the model effectively explains the variance in the data, with predictions strongly correlated to the actual labels. The mean squared error (MSE) is remarkably low at 0.0099, highlighting minimal discrepancies between the predicted and actual values.

```
Classification Report :
              precision    recall  f1-score   support

           0       0.99      0.99      0.99    100276
           1       0.99      0.99      0.99    100276

    accuracy                           0.99    200552
   macro avg       0.99      0.99      0.99    200552
weighted avg       0.99      0.99      0.99    200552

Accuracy: 0.9900424827476165
r2_score: 0.9601699309904663
L'erreur quadratique moyenne (MSE) est : 0.009957517252383422
```

**Figure 15.** Evaluation of the LSTM model

These results (Figure 15) demonstrate excellent model performance, with scores near perfection across all metrics. This indicates that the model is highly capable of accurately distinguishing between classes in terms of both precision and recall.

This confusion matrix (Figure 16) confirms that the LSTM model is highly efficient and reliable for failure prediction in industrial systems, achieving near-perfect accuracy. These results are particularly valuable for anticipating failures and enhancing predictive maintenance.

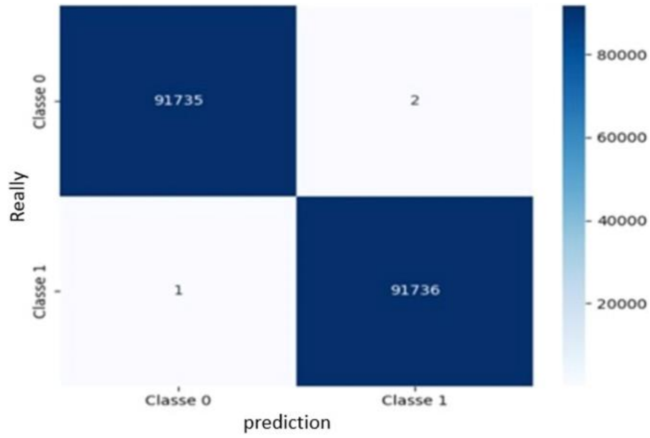


**Figure 16.** Confusion matrix of the LSTM model

The ROC curve (Figure 17) shows that the LSTM model is extremely effective in failure prediction, with an excellent detection rate and a very low false alarm rate. These results are crucial for enhancing predictive maintenance in industrial environments.

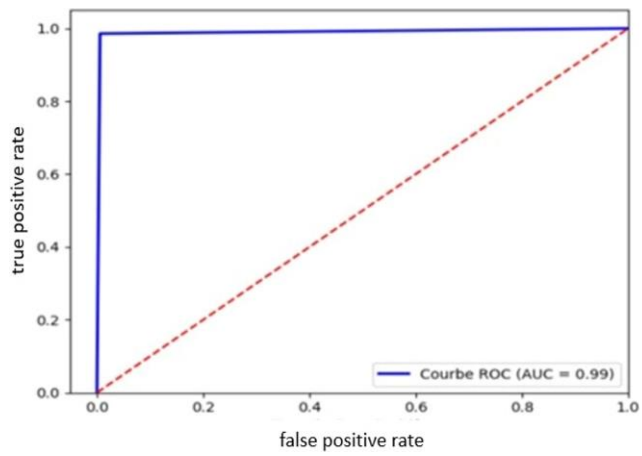


**Figure 17.** The ROC curve of the LSTM model

## 7. CONCLUSION

In a highly competitive and demanding environment, maintenance is now considered an essential component of the manufacturing process. It plays a crucial role in improving product quality and operational performance by ensuring equipment availability and meeting delivery schedules. The growing importance of maintenance has sparked significant interest in designing and adopting effective maintenance strategies aimed at enhancing system reliability, anticipating failures, and optimizing associated costs.

In this context, predictive maintenance policies and strategies, combined with big data and data science approaches, can serve as essential tools for the digitalization of maintenance processes in the industry 4.0 era. In our study, we proposed a predictive maintenance model following the IDAIC framework, which is an adapted and enhanced version of the general CRISP-DM approach. The steps of this methodology led us to develop a deep learning algorithm based on neural networks to predict future failures of a plastic injection molding machine.

The implemented algorithm is an LSTM with cross-validation applied to data collected from sensors and various systems using the Euromap 63 standard. The model was evaluated using various metrics, such as the confusion matrix, the ROC curve, and error measures, demonstrating high predictive performance.

The integration of this model into the industrial management systems of SMEs will strengthen the decision-making framework and develop proactive action plans to minimize unexpected downtimes, ensuring equipment availability and optimizing the adherence to delivery deadlines.

Future work will focus on developing classification models for failure types, estimating Remaining Useful Life (RUL), and improving the machine's cycle time within a comprehensive quality approach.

## REFERENCES


[1] Büchi, G., Cugno, M., Castagnoli, R. (2020). Smart factory performance and Industry 4.0. Technological Forecasting and Social Change, 150: 119790. https://doi.org/10.1016/j.techfore.2019.119790

[2] Leyh, C., Martin, S., Schäffer, T. (2017). Industry 4.0 and lean production—A matching relationship? An analysis of selected Industry 4.0 models. In 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, Czech Republic, pp. 989-993. https://doi.org/10.15439/2017F365

[3] Wang, S., Wan, J., Zhang, D., Li, D., Zhang, C. (2016). Towards smart factory for industry 4.0: A self-organized multi-agent system with big data based feedback and coordination. Computer Networks, 101: 158-168. https://doi.org/10.1016/j.comnet.2015.12.017

[4] Faccio, M., Persona, A., Sgarbossa, F., Zanin, G. (2014). Industrial maintenance policy development: A quantitative framework. International Journal of Production Economics, 147: 85-93. https://doi.org/10.1016/j.ijpe.2012.08.018

[5] Sharma, A., Yadava, G.S., Deshmukh, S.G. (2011). A literature review and future perspectives on maintenance optimization. Journal of Quality in Maintenance Engineering, 17(1): 5-25. https://doi.org/10.1108/13552511111116222

[6] Garg, A., Deshmukh, S. G. (2006). Maintenance management: Literature review and directions. Journal of Quality in Maintenance Engineering, 12(3): 205-238. https://doi.org/10.1108/13552510610685075

[7] Jasiulewicz-Kaczmarek, M., Gola, A. (2019). Maintenance 4.0 technologies for sustainable manufacturing—An overview. IFAC-PapersOnLine, 52(10): 91-96.

[8] Kinz, A., Bernerstaetter, R., Biedermann, H. (2016). Lean smart maintenance—Efficient and effective asset management for smart factories. In Proceedings of the 8th International Scientific Conference Management of Technology–Step to Sustainable Production, Porec, Croatia, pp. 1-3.

[9] Kumar, U., Galar, D. (2018). Maintenance in the era of industry 4.0: Issues and challenges. In Quality, IT and Business Operations: Modeling and Optimization, pp. 231-250. https://doi.org/10.1007/978-981-10-5577-5_19

[10] Boy, G.A. (2017). The Handbook of Human-Machine Interaction: A Human-Centered Design Approach. CRC Press.

[11] Lee, J., Ni, J., Singh, J., Jiang, B., Azamfar, M., Feng, J. (2020). Intelligent maintenance systems and predictive manufacturing. Journal of Manufacturing Science and Engineering, 142(11): 110805. https://doi.org/10.1115/1.4047856

[12] de Faria Jr, H., Costa, J.G.S., Olivas, J.L.M. (2015). A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. Renewable and Sustainable Energy Reviews, 46: 201-209. https://doi.org/10.1016/j.rser.2015.02.052

[13] Cakir, M., Guvenc, M.A., Mistikoglu, S. (2021). The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system. Computers & Industrial Engineering, 151: 106948. https://doi.org/10.1016/j.cie.2020.106948

[14] Poór, P., Basl, J., Zenisek, D. (2019). Predictive Maintenance 4.0 as next evolution step in industrial maintenance development. In 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, pp. 245-253. https://doi.org/10.23919/SCSE.2019.8842659

[15] Jimenez-Cortadi, A., Irigoien, I., Boto, F., Sierra, B., Rodriguez, G. (2019). Predictive maintenance on the machining process and machine tool. Applied Sciences, 10(1): 224. https://doi.org/10.3390/app10010224

[16] Lee, J., Ardakani, H.D., Yang, S., Bagheri, B. (2015). Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. Procedia Cirp, 38: 3-7. https://doi.org/10.1016/j.procir.2015.08.026

[17] Silvestri, L., Forcina, A., Introna, V., Santolamazza, A., Cesarotti, V. (2020). Maintenance transformation through Industry 4.0 technologies: A systematic literature review. Computers in Industry, 123: 103335. https://doi.org/10.1016/j.compind.2020.103335

[18] Prabhakar, D., Raj, V.J. (2014). CBM, TPM, RCM and A-RCM—A qualitative comparison of maintenance management strategies. International Journal of Business & Management Studies, 4(3): 49-56.

[19] Peng, Y., Dong, M., Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: A review. The International Journal of Advanced Manufacturing Technology, 50: 297-313. https://doi.org/10.1007/s00170-009-2482-0

[20] Silva Sanchez, R.E. (2016). Contribution au pronostic de durée de vie des systèmes pile à combustible de type PEMFC. Doctoral dissertation, Université du Québec à Trois-Rivières.

[21] Diamond, S., Marfatia, A. (2013). Predictive Maintenance for Dummies. John Wiley & Sons, Inc.

[22] Jasiulewicz-Kaczmarek, M., Legutko, S., Kluk, P. (2020). Maintenance 4.0 technologies—New opportunities for sustainability driven maintenance.

Management and Production Engineering Review, 11(2): 74-87. https://doi.org/10.24425/mper.2020.133730
[23] Passath, T., Mertens, K. (2019). Decision making in lean smart maintenance: Criticality analysis as a support tool. IFAC-PapersOnLine, 52(10): 364-369. https://doi.org/10.1016/j.ifacol.2019.10.058
[24] Shimaoka, A.M., Ferreira, R.C., Goldman, A. (2024). The evolution of CRISP-DM for data science: Methods, processes and frameworks. SBC Reviews on Computer Science, 4(1): 28-43. https://doi.org/10.5753/reviews.2024.3757
[25] Ahern, M., O'Sullivan, D.T., Bruton, K. (2022). Development of a framework to aid the transition from reactive to proactive maintenance approaches to enable energy reduction. Applied Sciences, 12(13): 6704. https://doi.org/10.3390/app12136704
[26] Mikhailov, A.V., Perrone, L. (2014). A method for f o F2 short-term (1-24 h) forecast using both historical and real-time f o F2 observations over European stations: EUROMAP model. Radio Science, 49(4): 253-270. https://doi.org/10.1002/2014RS005373