Enhancing Predictive Accuracy in Aircraft Engine MRO using Clustering and Similarity Methods

Leonardo Mendonça*, Flávia Pires*, José Barbosa*, Miguel Duarte †, Paulo Leitao*

* Research Centre in Digitalization and Intelligent Robotics (CeDRI),
Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC)
Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal
Email: {leonardomendonca, fpires, jbarbosa, pleitao}@ipb.pt
† OGMA-Indústria Aeronáutica de Portugal, Parque Aeronáutico de Alverca, 2615-173 Alverca, Portugal
Email: miguel.duarte@ogma.pt

***Abstract*—In the aircraft engine Maintenance, Repair, and Overhaul (MRO) process, effective task planning relies heavily on the expertise of lead engineers. However, when predictive models are used to assist decision-making, issues with incomplete, unbalanced, and inconsistent data can lead to errors in the planning task. Therefore, reliable predictions are crucial for optimising the operational efficiency. This paper proposes a methodology to improve the prediction of maintenance times for an aircraft engine MRO process by integrating K-means clustering with Cosine and Jaccard similarity methods to define a reliable prediction interval. This methodology was compared with the predictions of the Simple Linear Regression method, which resulted in the prediction interval approach significantly reducing prediction errors, increasing prediction accuracy, while optimising process management and maintenance task planning throughout the MRO process.**

***Index Terms*—Similarity Methods, Prediction, MRO, Aircraft Engines.**

I. Introduction

The accuracy of machine learning (ML) models is fundamentally influenced by both the quality and quantity of the data used during the training process [1]. The building of reliable models comprises several steps, such as data collection [2], data cleaning [3], and feature treatment [4]. These steps help to balance the dataset and mitigate biases that could affect the model performance [2]. However, even with meticulous data preparation, there is no guarantee that a ML model will be robust [1], as the evaluation metrics alone are often insufficient to ensure that the model accurately reflects the complexities of operations and processes [5].

When predictions support the decision-making in the maintenance planning process, errors may emerge due to a lack of data, variations, or inconsistencies, compromising the reliability of predictions and leading to operational inefficiencies [6], highlighting the need for the continuous monitoring and refinement of ML models. In the Maintenance, Repair, and Overhaul (MRO) process of the aircraft engine sector, the maintenance task planning highly dependent on the expertise of the lead engineer [7], which can introduce potential risks, limitations, or inaccuracies in the planning process, particularly when critical data is unavailable, e.g., when new motors arrive for repair without any historical data. Thus, the digitalisation of the MRO process data analysis can reduce costs by optimising the labour and machinery usage, enhancing the operational efficiency and minimising errors [8].

One approach to improve the accuracy and reliability of ML predictions is the application of similarity methods. Similarity refers to the degree of resemblance between two or more elements based on shared characteristics [9]. The majority of approaches that apply similarity methods primarily focus on Recommendation Systems (RS) using Collaborative Filtering (CF) [10], and Natural Language Processing (NLP) [11], as well as on improving ML techniques. Within ML, significant progress has been made in areas such as classification, clustering [12], [13], and feature selection [14]. However, the use of similarity methods to enhance the final predictions of ML models remains relatively unexplored, representing an area with considerable potential for further research.

This paper is proposing to use similarity methods to develop a robust methodology to determine a prediction interval for maintenance time prediction, referring to a range in which a future maintenance time is expected to fall, capturing both model uncertainty and task variability. This methodology enhances the prediction of maintenance times for the aircraft engine MRO process by ensuring the optimal reliability and accuracy while reducing planning errors. For this purpose, the K-means algorithm is employed to define clusters that categorise maintenance tasks into different classes based on the complexity of the maintenance process. At the same time, the Cosine and Jaccard similarity measures are used to identify relationships between maintenance processes. The preliminary results show a reduction of prediction errors in the case study related to the aircraft engine MRO process.

The paper is structured as follows: Section II provides an overview of related work on existing similarity methods and their applications in the maintenance sector. Section III describes the case study and introduces the proposed methodology. Section IV presents the main results obtained from the application of the methodology, including analyses and quantitative metrics. Finally, Section V summarises the conclusions and points out future work.

II. Related Work

The field of maintenance management has changed significantly with the emergence of data-driven approaches, which enable more effective decision-making and optimisation of maintenance planning. Accurate predictions are essential for the successful planning and execution of maintenance tasks, and similarity methods have proven valuable in analysing maintenance data. These methods help to identify patterns and correlations, ultimately improving the operational efficiency. Additionally, prediction intervals contribute to maintenance planning by quantifying the variability and uncertainty, thereby improving the reliability of maintenance predictions.

The studies conducted by [9], [15] and [16] present the application of similarity measures, as summarised in Table I, which provides an overview of the four common similarity measures. Cosine and Pearson methods focus on vector-based comparisons; while the Cosine measures the angle between vectors, Pearson captures the linear relationships. The Jaccard method is most commonly used for binary comparisons, assessing the overlap between categorical data, while the Euclidean Distance method measures the geometric similarity based on absolute differences. Each method has strengths and limitations, making its selection dependent on the data's specific characteristics.

Several studies have applied similarity measures to enhance decision-making in the maintenance sector. The authors of [17] introduced a similarity-based framework for monitoring rotating machinery using vibration signals, applying Cosine Similarity and Euclidean Distance, and achieving high accuracy with minimal labelled data. Similarly, [18] used Cosine and Pearson similarity to analyse failure descriptions and estimate downtime in Predictive Maintenance (PdM), demonstrating effective failure prediction but highlighting the need for further machine data analysis to understand downtime factors better. In preventive maintenance, [19] applied weighted similarity coefficients to cluster machines based on predicted failure severity, aiming to reduce costs and optimise maintenance efforts. However, the approach relied heavily on historical data and lacked validation across diverse manufacturing contexts. Additionally, [20] employed sequence matching and Cosine Similarity to improve the accuracy of Bayesian maintainability models, though challenges in the task representation and the computational complexity of the remaining method.

In the context of prediction interval, the authors of [21] combined Finite Element (FE) simulation data with ML to improve short-term pavement rutting prediction accuracy from 90.3% to 94.2% (R^2). The model quantified uncertainty by running simulations multiple times to create reliable prediction intervals, helping to improve the decision-making on road maintenance. However, the multiple simulations could amplify the effects of unreliable data, leading to inaccurate uncertainty intervals. Similarly, [22] introduced a Confidence-Weighted Extreme Learning Machine (CW-ELM) method that calculates input-dependent prediction intervals to enhance the prediction reliability. This method improves the prediction interval coverage and efficiency, handling large datasets like skin colour detection with 3.4 million samples in 11 minutes. However, this approach may not be accurate if the data does not follow a normal distribution or consistent variability.

Despite the significant advancements in similarity methods for maintenance analysis and the application of prediction intervals, challenges remain in accurately predicting maintenance tasks times and improving the overall reliability. While previous studies have demonstrated the potential of similarity measures, there is limited work that specifically addresses the accurate prediction of maintenance times. This study aims to address these challenges by introducing a methodology that applies similarity methods to calculate a prediction interval and enhance the final maintenance prediction time, focusing specifically on MRO aircraft engines.

III. Integration of Clustering and Similarity for Maintenance Planning

This section presents the case study's specifications and the methodology used to calculate an enhanced prediction accuracy using similarity methods and clustering algorithms.

A. Description of the Case Study

The case study focuses on MRO services in an aircraft engine maintenance line, using a dataset that includes maintenance tasks data from only a single engine model. The maintenance process is organised into a structured task hierarchy that contributes to the overall workflow. An engine maintenance visit begins when this arrives for servicing, with the maintenance process divided into distinct phases, such as *Evaluation*, *Disassembly*, *Cleaning*, and *Repair*. Each phase is managed by specialised departments and divided into smaller operations that are critical for the successful completion of the maintenance.

This case study was explored in [23], where a ML model based on the Simple Linear Regression (SLR) method was implemented within a graphical interface to predict maintenance times at different levels (e.g., general, task, or operation). The SLR method was tested among several models, including Lasso, Ridge, Random Forest, Support Vector Regression, Decision Tree, Gradient Boosting, and K-Nearest Neighbours. All models were trained using five key features related to the maintenance context: *Phase, Department, Operation, Engine and Estimated hours*. The SLR model achieved an R^2 of 0.81, outperforming the evaluated models. However, due to limitations in data incompleteness and imbalance for training the ML model, e.g., lack of maintenance visits, limited engine models, and new engines without historical data, the predictions showed a percentage deviation of 22.1% compared to the actual maintenance time. Although this solution facilitates maintenance planning and decision-making, it has slightly reduced the reliability of the predicted maintenance times. With this in mind, the objective of this paper is to propose mitigation measures for this problem by calculating a prediction interval

TABLE I: Overview of similarity methods and their characteristics.

Similarity Method	Description and Range	Characteristics
Cosine [15], [16]	Measures the similarity between vectors through the cosine of the angle between them. The range is from 0 to 1, where 0 indicates no similarity and 1 indicates identical direction.	Focuses on the orientation of vectors, ignoring magnitude. However, it is not sensitive to magnitude, which may be a limitation when the magnitudes of vectors differ significantly.
Pearson [15], [16]	Measures the linear correlation between two vectors. Ranges from -1 to 1, where -1 indicates an inverse correlation, 1 indicates a direct correlation, and 0 indicates no correlation.	Effective at detecting linear relationships, but it assumes linearity and is sensitive to outliers. This can make it less reliable if the data deviates from a linear trend or contains extreme values.
Jaccard [9], [16]	Considers the number of common occurrences between vectors, divided by the number of occurrences that appear in at least one of the vectors. Ranges from 0 to 1, where 0 means not similar and 1 means identical sets.	Particularly useful for binary data and easy to apply. However, it does not consider the magnitude of values, limiting its use for continuous or non-binary data.
Euclidean Distance [9], [16]	Measures the distance between vectors through the line length between them. The range is from 0 to ∞, where 0 indicates identical vectors.	Easy to understand and effective for measuring absolute differences between vectors. However, it is sensitive to scale and does not perform well with vectors with different magnitudes.

to enhance the prediction accuracy of the model by integrating clustering and similarity methods.

B. Methodology

The proposed methodology, illustrated in Fig. 1, consists of three main parts: *ML Prediction*, *Clustering*, and *Similarity*. The main objective of the proposed model is to generate more reliable and accurate predictions, and consequently promote a more informed decision-making by the planning engineers.

The ML Prediction part of the methodology consists of applying the ML prediction model developed in [23], which was trained with the historical data to predict the time required for the maintenance process. This prediction is based on the selection of specific tasks to be performed, which are specified by the engineer based on the client's reported problems.

The second part, Clustering, of the methodology involves the application of a clustering algorithm to classify the historical maintenance data into different types of maintenance. The type of maintenance can differ depending on several factors that can significantly affect the complexity and duration of the process, such as the problems identified throughout the maintenance process, the characteristics of the engine, and its current physical condition. The results from the clustering will be the basis for the application of the similarity methods.

Finally, the third part of the methodology, Similarity, focuses on the application of similarity methods to improve the accuracy of the predictions associated with the maintenance types identified in the clustering phase. These methods assess the degree of similarity between different maintenance processes, allowing for more accurate predictions by considering patterns and relationships in the data. After this, the probability of that type of maintenance occurring is calculated, and based on that, the prediction interval is established and added to the prediction time.

C. Prediction Interval Calculation

The main objective of this methodology is the enhancement of the prediction accuracy by calculating the prediction interval. This is based on the standard application of clustering, in this case, the K-means algorithm, and the similarity methods. The application of the similarity methods requires a more detailed explanation.

Thus, for two maintenance processes to be considered similar, they must satisfy two premises: i) they must involve the same tasks, and ii) have the same processing time for each task. However, since maintenance processes may differ in the number of tasks they include, the similarity between the maintenance tasks is calculated using the Jaccard method, as shown in Eq. (1).

$$J(A,B) = \frac{|R_A| \cap |R_B|}{|R_A| \cup |R_B|} \tag{1}$$

where *A* and *B* denote the two vectors being compared, while *R* represents the elements associated with each vector. This method can measure the proportion of identical elements in relation to the total number of unique elements. Thus, it quantifies the degree of similarity between two maintenance processes based on the overlap of their corresponding tasks.

Considering the similarity between the times of maintenance tasks, the Cosine was used, represented by the Eq. (2).

$$\text{COS}(A,B) = \frac{\sum_{i \in I_{A,B}} r_{Ai} \cdot r_{Bi}}{\sqrt{\sum_{A \in I_A} r_{Ai}^2} \cdot \sqrt{\sum_{B \in I_B} r_{Bi}^2}} \tag{2}$$

where *I* refers to a set of indices used to indicate the positions of non-zero values in a vector. The term *r* represents the value of an element at a specific position within a vector, where r_{Ai} indicates the value at position *i* in vector *A*, and r_{Bi} indicates the value at position *i* in vector *B*. This equation compares the vectors *A* and *B*.

Therefore, by individually addressing the specific characteristics of the case study, the integration of the Jaccard and Cosine methods enhances both the accuracy and effectiveness of measuring similarity between maintenance tasks. Combined, these methods are not sensitive to vector magnitude, i.e., even

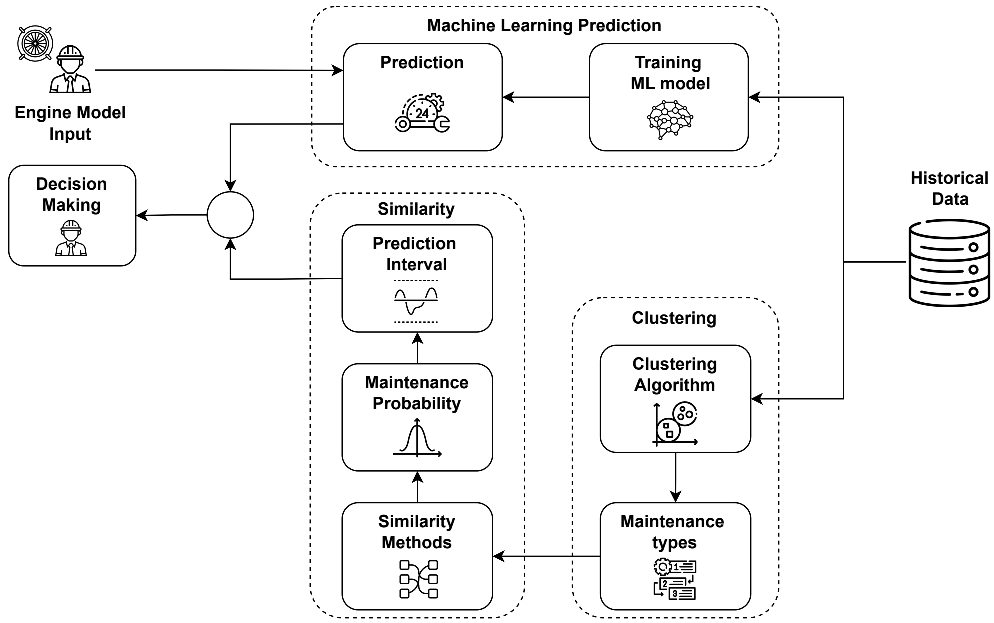


Fig. 1: Methodology for enhancing the prediction accuracy combining clustering and similarity methods.

when one maintenance task has a significantly longer total time than another, they can still reveal similar patterns across similar tasks.

After the similarities are calculated, the probability of each maintenance process occurring again in the future is calculated, i.e., identifying which process shares the highest degree of similarity with the other. This probability is calculated using the Eq. (3).

$$Prob_i = \frac{\sum Similarity_{>\theta}}{n_{visits}} \tag{3}$$

This equation calculates for each visit i the proportion between the sum of similarities that exceeds θ (a minimum similarity value between 0 and 1) relative to the total number of visits in the historical data. The value of θ chosen was 60%, an acceptable threshold when considering the complexity of maintenance tasks performed, ensuring sufficient similarity for a reliable comparison and considering the variability inherent in maintenance processes. Based on these probabilities, the prediction interval is defined according to Eq. 4

$$P_{Interval} = \left| P_{max} - \bar{P}_{cluster} \right| \tag{4}$$

In this equation, the prediction interval ($P_{Interval}$) is determined by the absolute difference between the predicted maintenance time with the highest probability (P_{max}) and the average maintenance time within the corresponding cluster ($\bar{P}_{cluster}$). This interval represents the expected range of deviation in future maintenance time estimations, capturing both the similarity between visits and the specific characteristics of each maintenance class. By incorporating this interval into the case study, the prediction process becomes more robust and accurate, supporting more effective management of time in the maintenance process.

IV. Experimental Validation and Results

This section describes the application of the proposed methodology to the case study for predicting MRO maintenance times. The achieved results were discussed, particularly the integration of the K-means clustering algorithm with the Jaccard and Cosine similarity methods to improve the prediction accuracy.

A. Clustering Algorithm

The K-means algorithm was applied to perform clustering and divide the historical data into classes representing the different types of maintenance applied to MRO. Moreover, the elbow method was applied, as illustrated in Fig. 2, to identify the optimum number of clusters (k) to be used, considering the historical data of the maintenance processes.

This method calculates the inertia, which is the sum of the squared distances between each data point and the centroid of its assigned cluster, where increasing the number of clusters generally decreases the inertia, as the data points are more tightly grouped within each cluster. The result showed that the optimal number of clusters is between 3 and 4. This result was confirmed by the maintenance company, as the maintenance processes are divided into three classes, light, medium, and heavy, according to their complexity. Fig. 3 illustrates the dispersion of maintenance processes across the clusters obtained using the K-means algorithm, with the number of clusters (k) determined by the elbow method and the default hyperparameters applied.

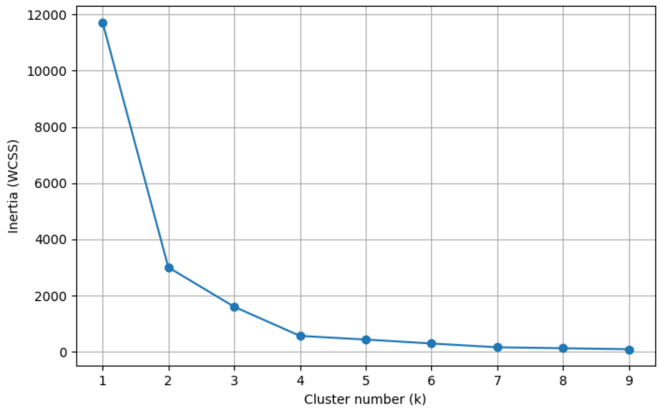


Fig. 2: Result of the Elbow method.

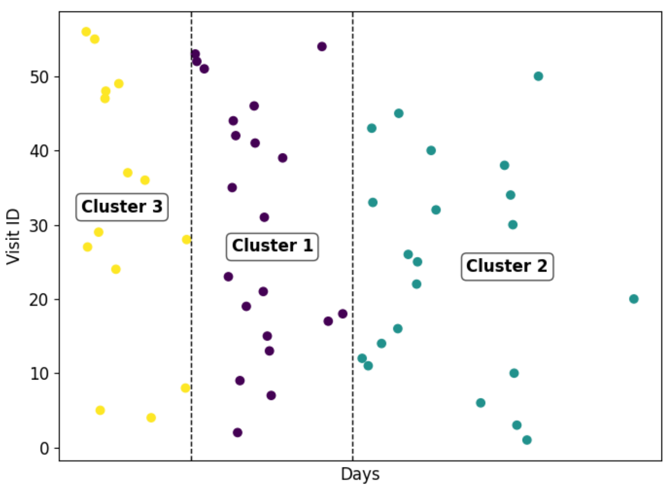


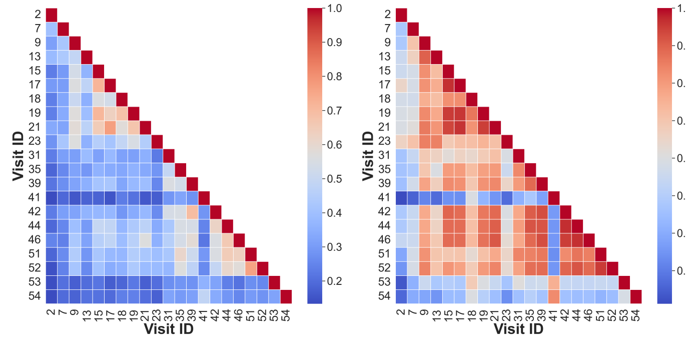
Fig. 3: Clusters definition based on historical data.[1]

B. Analysis of Similarity

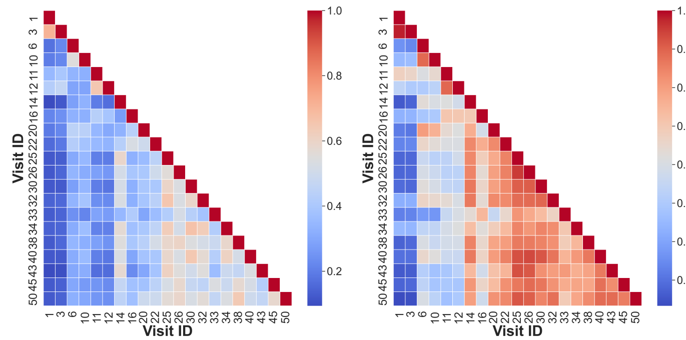
The similarity analysis was performed by comparing the historical maintenance data in each cluster separately, thereby improving the reliability and accuracy of the results by minimising the influence of unrelated maintenance processes. For this purpose, the Jaccard and Cosine similarity measures were applied to the possible cluster maintenance combinations, generating the results shown in Fig. 4.

The results show that warmer colours (i.e., stronger reds and values closer to 1) are associated with a higher correlation between maintenance visits, while colder colours (i.e., stronger blues and values closer to 0) are associated with a weaker correlation. However, the individual analysis of the results does not fully represent the true relationship of similarity between the maintenance visits, as the performed maintenance tasks significantly influence the similarity measure more than the maintenance time. To address this, weights of 0.7 and 0.3 were assigned to the Jaccard and Cosine results, respectively, to generate a mixed similarity measure. These values were chosen to emphasise the greater importance of the task similarity captured by the Jaccard method, while still incorporating the aspects captured by the Cosine similarity. The resulting mixed similarity provides a more balanced and representative measure, as illustrated in Fig. 5.

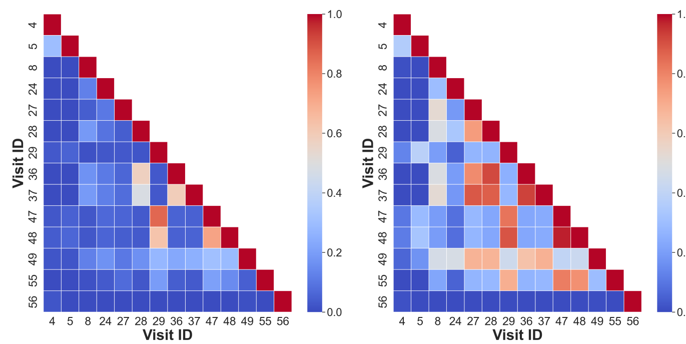
Using the probability equation (Eq. (3)), the maintenance visits with the highest probability of being repeated in the future had values of 40%, 55% and 16% for clusters 1, 2 and 3, respectively. These probabilities indicate that maintenance processes in cluster 2 are the most probable to happen again in the future. In contrast, the probability for cluster 3 exhibits



(a) Cluster 1



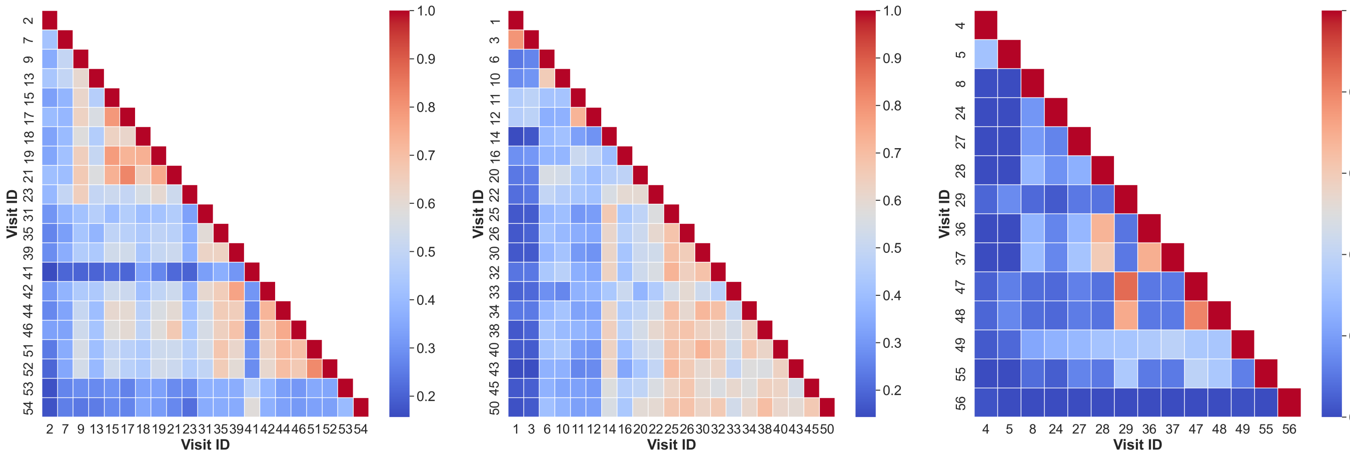
(b) Cluster 2



(c) Cluster 3

Fig. 4: Results of similarity analysis using the Jaccard method (left side) and Cosine method (right side).

[1]The days presented in this figure have been omitted to protect the confidentiality of the company's sensitive information.



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3

Fig. 5: Mixed Similarity by combining Jaccard and Cosine results.

greater uncertainty, suggesting that processes in this cluster may be more variable and more complex to predict. Thus, to enhance the prediction and reliability of the model used by [23], the maintenance with the highest probability in each cluster was selected to calculate the prediction interval. The identified prediction intervals are for cluster 1 of 3.31 days, cluster 2 of 4.17 days, and cluster 3 of 2.57 days.

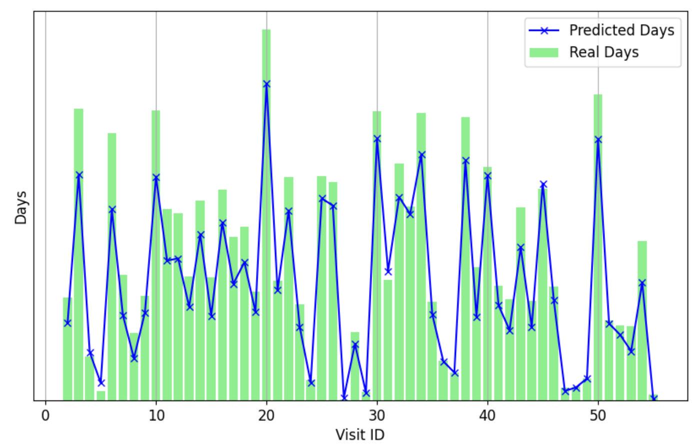
C. Evaluation of Prediction Interval

To assess the efficiency of the developed methodology, a comparison was made between the actual maintenance time and the predicted maintenance time with and without the adjustment using the calculated prediction interval. These predictions were generated using the Simple Linear Regression model trained in [23], based on historical maintenance data. The results of this comparison are illustrated in Fig. 6.

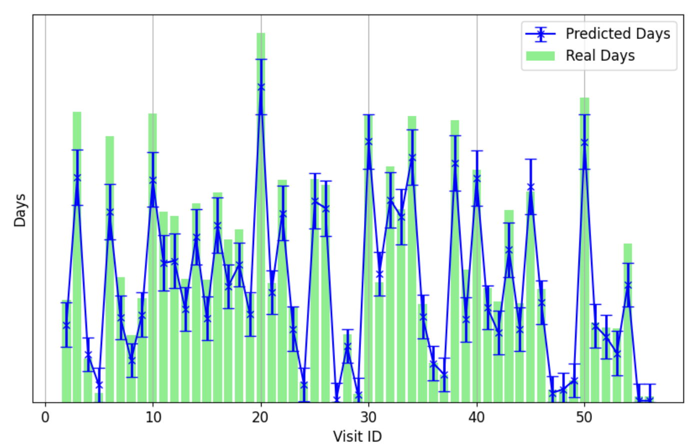
The first graph displays green bars representing the actual maintenance time, overlaid with a blue line indicating the predicted time using the ML model. The second graph presents a similar structure, but the blue line is now accompanied by error bars representing the prediction intervals. A qualitative analysis suggests that including the prediction intervals allows the predictions to reflect the actual maintenance times more accurately. Complementing this, a quantitative analysis was performed and the results are presented in the Table II, which reports the absolute errors, calculated as the average difference in time units between the predicted and actual maintenance times. It also presents the corresponding percentage deviations, which express the error relative to the actual time and provide a standardised way to compare prediction accuracy.

Therefore, the table shows that, when comparing the engineer estimation with the prediction using the interval, the absolute error reduces by 46.2%. At the same time, the percentage deviation demonstrates an even larger improvement, decreasing by 72.2%. This indicates that even a small

[2]Values for Engineer Estimation and ML Prediction are sourced from [23].

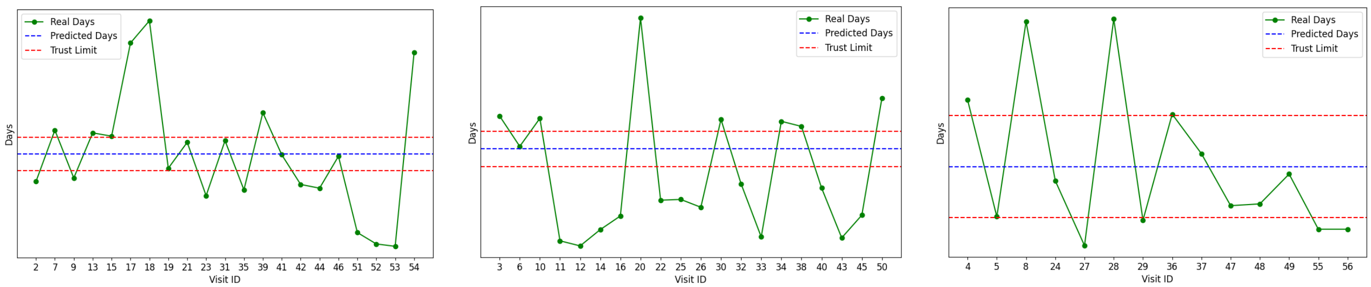


(a) Prediction of maintenance time using the ML model.[1]



(b) Prediction of maintenance time using the prediction interval.[1]

Fig. 6: Comparison between the actual maintenance time and the predictions.



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3

Fig. 7: Evaluation of the generic prediction integrated with the prediction interval.[1]

TABLE II: Impact of prediction interval on prediction accuracy.[2]

Approach	Absolute Error (Days)	Percentage Deviation (%)
Engineer Estimation	2.4	17.4
ML Prediction	3.8	22.1
ML Prediction with Interval	1.8	6.8

change in absolute error can result in a relatively significant reduction in the percentage deviation, making the improvement in predictive accuracy more significant. Furthermore, the use of the prediction interval for estimating maintenance time proves to be more efficient and reliable when applied to the MRO maintenance process.

In addition, to provide a more general prediction of the maintenance time for each cluster, which can be used as an initial estimate for a new maintenance process, the average prediction of the clusters was combined with the defined prediction intervals to assess how they compare with the actual maintenance time, as shown in Fig. 7.

Thus, with the introduction of the prediction interval, it is possible to include the times of the maintenance process efficiently, even in a more generic prediction. To evaluate these measures, the percentage deviations of the actual maintenance times in relation to the prediction intervals were calculated, resulting in values of 14.4%, 12.3%, and 12.5%, respectively. These results show that even in a general prediction, introducing the prediction intervals allows for greater accuracy compared to the predictions realised by the engineers.

D. *Feedback from company*

According to the company holding the case study, the prediction approach improves the reliability of the Turnaround Time (TAT) for the aircraft engine maintenance. Since TAT is an important Key Performance Indicator (KPI) in the MRO sector, accurate and realistic predictions of the maintenance times are essential for ensuring customer satisfaction and avoiding contractual penalties.

By leveraging historical maintenance data, the ML analyses past interventions on similar engines, considering the expected work scope to calculate a more realistic TAT. This reduces unexpected schedule deviations and enables a more reliable commitment to the customer, preventing unforeseen delays in the planned shop floor tasks.

The main advantage of this approach is the mitigation of risks associated with the unpredictability of maintenance execution by identifying patterns in similar engines, allowing the MRO shop to predict:

- Components with a higher probability of premature wear;
- Probability of encountering unexpected damages (additional works from the initial work scope);
- Critical areas requiring detailed inspection before maintenance begins;

With these insights, the MRO shop can prepare for potential issues in advance by ensuring spare parts availability, optimising resource allocation, and proactively notifying customers of potential scope adjustments.

V. Conclusion

The task planning of aircraft engine MRO processes heavily relies on the expertise of lead engineers. Digitalising MRO data analysis is crucial for reducing costs and minimising planning errors. While similarity methods have shown strong potential in various applications, including ML predictions, their use in improving maintenance time estimation remains limited.

This paper presents a methodology that integrates clustering and similarity methods to define a prediction interval to enhance MRO time prediction accuracy, addressing key challenges such as data incompleteness and imbalance in training the ML model. Maintenance tasks were classified using the K-means clustering algorithm, while the Jaccard and Cosine similarity measures served as the foundation for defining the prediction interval.

The achieved experimental results demonstrate that implementing similarity methods to assess maintenance process similarity effectively defines a prediction interval that complements the predictions generated by the ML model. The quantitative analysis shows that this approach reduces the absute error by 46.2% and the percentage deviation by 72.2% significantly improving the accuracy and reliability of MRO prediction time, while minimising planning errors. Additionally, the general prediction analysis showed lower percentage deviations compared to the estimations made by engineers, confirming the effectiveness of prediction intervals even in less specific scenarios.

In future studies, methodological improvements will be considered, including the validation of the methodology with data from different aircraft engines, ensuring a broader applicability. Alternative clustering algorithms will be evaluated, as they may better capture the complex patterns present in maintenance data. To support this, different techniques for determining the optimal number of clusters will be explored to improve clustering accuracy. Furthermore, the current empirical selection of similarity weights will be refined through a principled validation strategy or sensitivity analysis. Finally, incorporating additional features like cluster ID may enhance the model's predictive performance.

Moreover, a further critical challenge that will be addressed in future work is the prediction of MRO operation times for new engines for which historical data is not available. In such cases, similarity methods may be employed to identify comparable engines based on their physical characteristics, thereby enabling the proposed methodology to estimate the MRO task times for these new engines.

ACKNOWLEDGMENTS

This work was supported by national funds: UID/05757 - Research Centre in Digitalization and Intelligent Robotics (CeDRI); and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020). Additionally, this work is co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project Produtech_R3, with reference 60.

REFERENCES

[1] S. E. Whang, Y. Roh, H. Song, and J. G. Lee, "Data collection and quality challenges in deep learning: a data-centric ai perspective," *VLDB Journal*, vol. 32, pp. 791–813, 7 2023.

[2] R. Pradhan, "Valuation-based data acquisition for machine learning fairness," *Proceedings of the VLDB Endowment*, vol. 2150, p. 8097, 2024.

[3] I. F. Ilyas and T. Rekatsinas, "Machine learning and data cleaning: Which serves the other?" *J. Data and Information Quality*, vol. 14, no. 3, Jul. 2022.

[4] T. Rawat and V. Khemchandani, "Feature engineering (fe) tools and techniques for better classification performance," *International Journal of Innovations in Engineering and Technology*, 05 2019.

[5] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics*, vol. 10, no. 5, 2021.

[6] A. Bousdekis, K. Lepenioti, D. Apostolou, and G. Mentzas, "A review of data-driven decision-making methods for industry 4.0 maintenance applications," *Electronics*, vol. 10, no. 7, 2021.

[7] A. Chatzi, W. Martin, P. Bates, and P. Murray, "The unexplored link between communication and trust in aviation maintenance practice," *Aerospace*, vol. 6, 06 2019.

[8] A. Apostolidis, M. Pelt, and K. P. Stamoulis, "Aviation data analytics in mro operations: Prospects and pitfalls," in *Proceedings of Annual Reliability and Maintainability Symposium (RAMS)*, 2020, pp. 1–7.

[9] F. Fkih, "Similarity measures for collaborative filtering-based recommender systems: Review and experimental comparison," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 7645–7669, 2022.

[10] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: Recommendation models, techniques, and application fields," *Electronics*, vol. 11, no. 1, 2022.

[11] D. Chandrasekaran and V. Mago, "Evolution of semantic similarity—a survey," *ACM Comput. Surv.*, vol. 54, no. 2, Feb. 2021.

[12] J. Irani, N. Pise, and M. Phatak, "Clustering techniques and the similarity measures used in clustering: A survey," *International Journal of Computer Applications*, vol. 134, pp. 9–14, 01 2016.

[13] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, 2023.

[14] H. Alazzam, A. Sharieh, and K. E. Sabri, "A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer," *Expert Systems with Applications*, vol. 148, p. 113249, 2020.

[15] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web*, ser. WWW '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 285–295.

[16] J. Joy and R. V. G, "Comparison of generic similarity measures in e-learning content recommender system in cold-start condition," in *Proceedings of IEEE Bombay Section Signature Conference (IBSSC)*, 2020, pp. 175–179.

[17] S. Aburakhia, T. Tayeh, R. Myers, and A. Shami, "Similarity-based predictive maintenance framework for rotating machinery," in *Proceedings of 5th International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2022, pp. 1–6.

[18] H. Tekgöz, S. I. Omurca, K. Koc, U. Topçu, and O. Çelik, "Semantic similarity comparison between production line failures for predictive maintenance," *Advances in Artificial Intelligence Research*, 2022.

[19] A. Abdelhadi and T. Khreis, "Preventive maintenance operations based on weighted similarity coefficient," in *Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2016, pp. 1431–1435.

[20] Z. Wu and J. Hao, "A maintenance task similarity-based prior elicitation method for bayesian maintainability demonstration," *Mathematical Problems in Engineering*, vol. 2020, no. 1, p. 2730691, 2020.

[21] A. Akusok, A. Gritsenko, Y. Miche, K.-M. Björk, R. Nian, P. Lauren, and A. Lendasse, "Adding reliability to elm forecasts by confidence intervals," *Neurocomputing*, vol. 219, pp. 232–241, 2017.

[22] K. Chen, M. E. Torbaghan, N. Thom, A. Garcia-Hernández, A. Faramarzi, and D. Chapman, "A machine learning based approach to predict road rutting considering uncertainty," *Case Studies in Construction Materials*, vol. 20, p. e03186, 2024.

[23] L. Mendonça, F. Pires, M. Duarte, J. Barbosa, and P. Leitão, "Monitoring and prediction of maintenance operations for aircraft engine repair," *Accepted for presentation at the 15th IFAC Workshop on Intelligent Manufacturing Systems (IMS)*, 2025.