Large Language Model Accelerated Maintenance Insights

Noah Getz[1] and Xiaorui Tong[1]

[1]*RTX Corporation, Enterprise Data, Brooklyn, NY, 11201, USA*

noah.getz@rtx.com
xiaorui.tong@rtx.com

ABSTRACT

Maintenance operations play a crucial role in the efficient functioning of manufacturing plants across various industries. Legacy systems that record maintenance operation data have long been utilized to facilitate day-to-day activities, conduct investigations, and maintain records. Previous studies have leveraged this data to explore insights and patterns related to interruption causes. However, the application and scalability of these studies have been impeded by issues such as data quality, text inaccuracies, and a lack of common natural language processing tools compatible with legacy systems. The rapid advancement of generative artificial intelligence, particularly large language models (LLMs), presents opportunities to address these challenges and enable quicker insights. This paper proposes a technical architecture for expediting insights into maintenance operation data through LLM-enabled data augmentation, summarization, and extraction, as well as embedding-based feature extraction and downstream clustering. A use case is presented based on maintenance data from an aerospace manufacturing plant, with user interviews conducted to evaluate the generated insights and system feedback. This work pioneers the adoption of LLMs in accelerating insights from under-utilized maintenance operation data, paving the way for an LLM-powered maintenance co-pilot.

1. INTRODUCTION

Maintenance operations are a critical aspect of day-to-day activities for manufacturing plants across various industries. A vast amount of data has been recorded and preserved, including the time of issue occurrence, problem descriptions, maintenance logs, and more. This data has long been stored in legacy databases, typically within Enterprise Resource Planning (ERP) systems. It serves not only for enterprise record-keeping purposes but also for understanding and improving maintenance operations. This improvement is achieved either by analyzing the maintenance data itself or by cross-correlating it with other data sources, the most common of which is condition-based maintenance (CBM) data.

Wolniak proposed using a combination of tools for analyzing downtimes in the production process to facilitate improvement. Specifically, the 5 WHY method and the Ishikawa diagram were employed in an automotive industry case study to demonstrate how downtime causes can be identified and categorized, generating statistics and insights from complex data. While the appropriate tools for planned usage may differ across industries, there is a lack of an automated approach to extract downtime insights universally (Wolniak, 2019).

In (Pertselakis, Lampathaki, & Petrali, 2019), the authors emphasized the importance of leveraging existing data from legacy systems, including daily operation data, maintenance events, production line data, and more. They applied a self-organizing map to a two-variable dataset (machine ID and breakdown causes) from maintenance logs to demonstrate the grouping of machines and causes that frequently led to interruptions. They noted that these patterns and correlations are otherwise unobservable by humans due to the complex manufacturing environment, diverse types of machines, and various interruption causes.

(Galar, Gusta, Tormos, & Berges, 2012) introduced a data mining methodology that combines condition monitoring data with historical maintenance management data, aiming to integrate separate systems to support maintenance decisions. (Teixeira, Teixeira, & Lopes, 2021) proposed a structure to organize machine maintenance data to enable the prioritization of failure modes of a machine for CBM purposes.

So far, most progress has been made in research. The factors contributing to the gap between research and industrial applications are:

1. Data Quality: Problem statements, maintenance notes, and descriptions are often not the most accurate due to the fast-paced production schedules in plants. Additionally, technical jargon, abbreviations, and other terms common among maintenance personnel may be unfamiliar to others.
2. Lack of Natural Language Processing (NLP) Tools: Legacy systems often lack NLP tools, and there is a general shortage of analysts with NLP skills for plants to leverage.

The advancement of large language models can tremendously decrease both development effort and time to insights for NLP tasks. Although these models are trained on general corpus data rather than domain-specific data, their strong reasoning and summarizing capabilities can still be leveraged for domain-specific purposes (Lukens & Markham, 2018). Before industrial large knowledge models (Lee & Su, 2023) become available, current LLMs, when incorporated with industrial knowledge, present opportunities for practical applications.

In this paper, a technical architecture for an LLM-enabled maintenance copilot is introduced. This system programmatically augments the raw machine maintenance history data with SME knowledge, systematically extracts problems and solutions from this data using LLMs, optimizes the number of clusters for machine downtime causes, and thus generates references for encountered maintenance problems and solutions. The generated references provide an avenue to delve into insights, including past maintenance issues and corresponding solutions. A use case has been conducted with maintenance operation data from an aerospace manufacturing plant to demonstrate the extracted insights.

These insights can greatly benefit maintenance operation optimization, personnel training and assistance, as well as data-driven decision-making. This system, therefore, creates a blueprint for a maintenance copilot that helps various personas across the maintenance operation spectrum achieve faster insights.

2. APPROACH

2.1 The Framework for LLM Accelerated Maintenance Insights

The proposed approach is presented in Figure 1. Html data exported from machine history in the Enterprise Resource Planning system is converted to a readable table format. The converted content still includes a significant amount of technical jargon, abbreviations, typos, etc., which will hinder the effective usage of downstream language models and cause potential hallucinations. Data cleaning is critical in ensuring the quality of data input. Specifically, a technical dictionary is built and incorporated to provide context for domain-specific content. The data cleaning operations involve replacing acronyms and technical terms with their full forms or descriptions, removing unnecessary characters, and expanding abbreviations.

During the feature engineering stage, prompt engineering with LLM is utilized to extract problem descriptions and solutions from the input data. The summarized data is concatenated per entry and embedded with embedding models. The embedding model can be Universal Sentence Encoder (USE) or a Sentence Transformer. Uniform Manifold Approximation and Projection (UMAP) is used for dimension reduction of the embedding features (McInnes, Healy, Melville, 2018) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) for clustering (McInnes, Healy, & Astels, 2017). Hyperparameters listed in Table X are optimized using Hyperopt (Bergstra, Yamins & Cox, 2013). According to the optimization, the optimal embedding model along with clustering results are obtained. This clusters machine maintenance history into known issue categories with labeled issue groups. The clustering results provide the following important insights:

1. A categorical view of maintenance issues, which can be utilized to understand the cost per issue group based on downtime. This information can be leveraged to better plan for operation optimization, inform maintenance management, and guide business decisions on equipment purchasing, etc.
2. Past solutions adopted for specific types of issues. These references can not only assist in accelerating the resolution of similar issues but also help train new maintenance personnel with SME knowledge in practical maintenance scenarios.

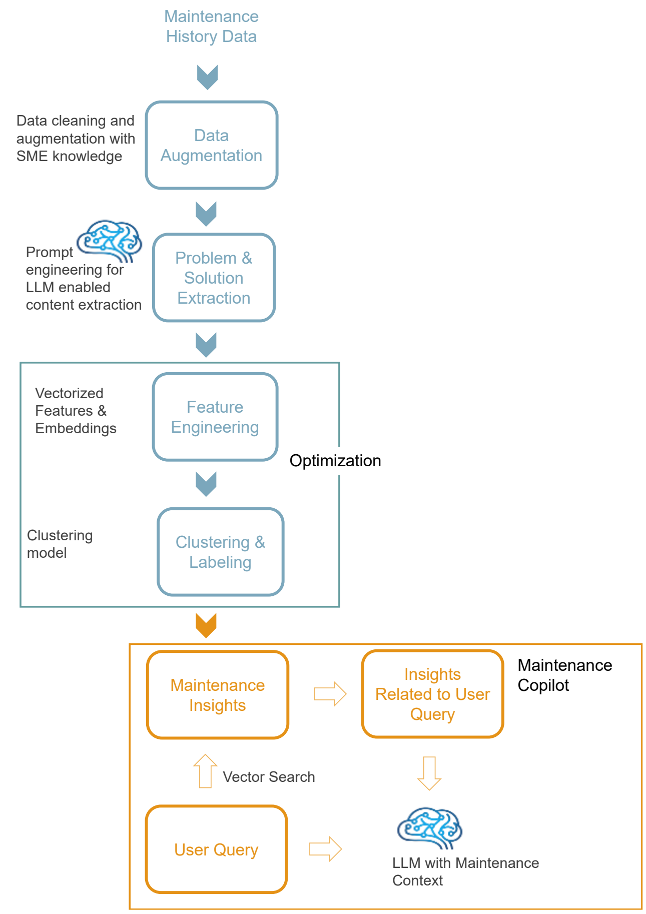

Figure 1. The framework for LLM Accelerated Maintenance Insights

2.2 End Users of the Maintenance Insights

Table 1 below provides a list of factory users and ways they can consume the generated insights. This list has been validated through an interview with the factory maintenance team.

Table 1. Factory Users and Ways of Leveraging Insights

User	Insights	Actions/Decision Enabled by Insights
Plant Managers	Top breakdown causes	Assign resources to investigate top issues. Provide information to inventory team for spare parts planning.
Maintenance Team Lead	Top breakdown causes & Solutions	Propose training to operators. Prioritize maintenance focus. Propose improvement measures.
Maintenance Team Members	Relevant issues and solutions in the past	Search relevant issues and solutions to aid in maintenance.
Predictive Maintenance Team	Top costly failure modes	Know which failure modes to install hardware and pursue monitor solutions.
Industrial Engineers	Top issues affecting availability	Propose investigations for maintenance team. Explore new solutions to address costly issues.

3. CASE STUDY

3.1 Implementation with Machine History Data

The above approach has been demonstrated using maintenance history data from an aerospace manufacturing plant. Data from the past six years was exported in HTML format from the machine history section of a machine tool stored in the Enterprise Resource Planning system. Table 2 lists the relevant fields used in this use case.

Table 2. Dictionary for Exported Machine History Data

Column Head	Content
Equipment No.	Unique equipment identifier
Problem Description	Initial description of the problem, may or may not be accurate
Completion Date	Date of maintenance completion
Down Time(H)	Down time for this entry in hours
Maintenance Log	Long text description of symptoms, investigations and actions taken.

Exported data is then processed through the proposed approach in Figure 2.

The exported data is processed using the proposed approach detailed in Figure 2., which expands upon the methodology outlined in Figure 1 by focusing specifically on data processing stages up to the Maintenance Insights stage. Figure 2.provides a more detailed version of the approach, excluding the copilot component covered in Figure 1, which remains a topic for future work. Mistral 7B is employed to extract problems and solutions from maintenance records, facilitating subsequent stages of feature engineering, clustering, and optimization. These stages involve evaluating multiple model options defined by dimension-reduced embedding features from various embedding models and optimizing hyperparameters of clustering models. The resulting optimal model reveals maintenance insights through its clustering outcomes.

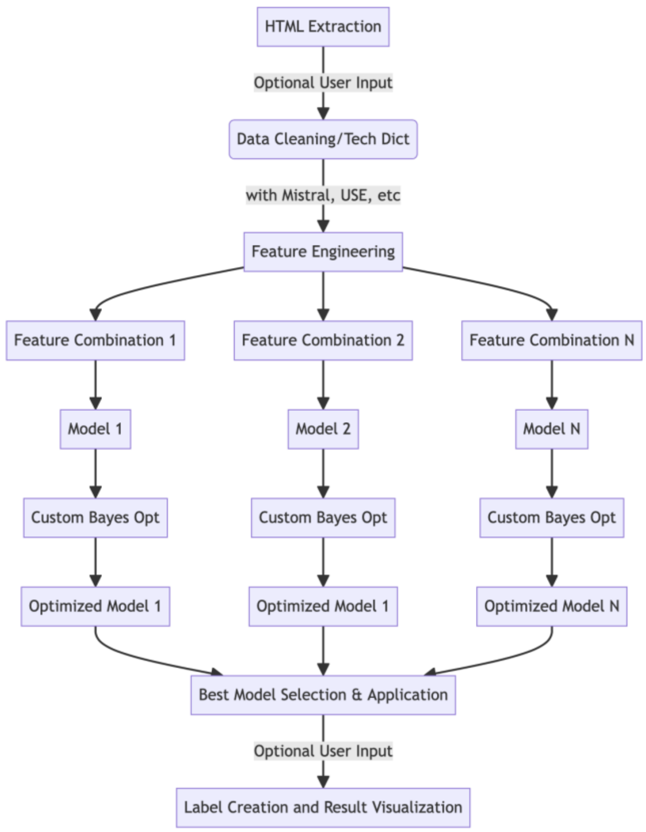

Figure 2. Case Study of LLM Enabled Maintenance Insights

.

3.2 Results and Visualizations

The presented approach has selected the model using Term Frequency - Inverse Document Frequency (TFIDF) vectorization of the "Problem Description" column and Universal Sentence Encoder embeddings of Mistral-extracted issues from the "Maintenance Log" column. Optimized hyperparameters for the HDBSCAN clustering algorithm using Hyperopt (Tong, Jung, & Banning, 2023) are listed in Table 3.

Table 3. Results of Optimized Hyper Parameters

Hyper Parameter	Optimized Value
n_components	7
n_neighbors	3
min_cluster_size	2

It returns 8 clusters with 0.01 loss. Loss is quantified as the proportion of records that have a probability of belonging to the assigned cluster lower than the threshold (0.05). This indicates that maintenance records are effectively clustered with minimal ambiguity among the clustered records.

Figure 3. Scatter Plot of 2-Dimensional t-SNE Features

Figure 3 presents a two-dimensional scatter plot achieved through t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction. Each point in the plot represents a maintenance event record, and the color of the point indicates the cluster to which the document is assigned.

In addition, a word cloud plot can visualize frequently occurring words in the summaries of maintenance records assigned to each cluster. Figure 4 displays the word cloud for the 8 clusters, where words are shown in varying sizes according to their frequency. The high-frequency words align well with the cluster labels.

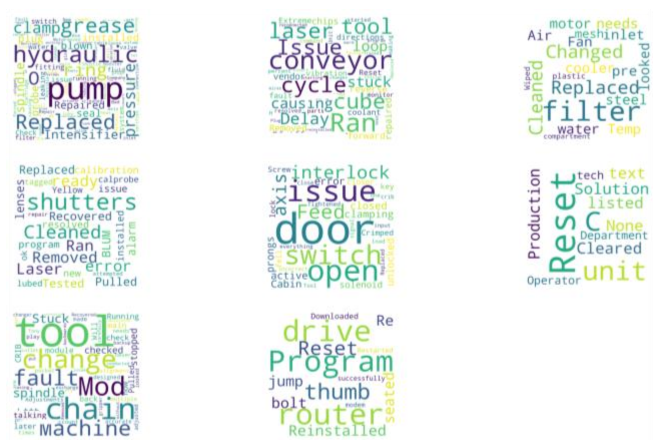

Figure 4. Word Cloud for Optimized Clusters

Lastly, for each cluster, mean and median downtime, along with counts of records, provide data-driven evidence of the operational impact associated with each type of issue. The results for the 8 clusters are displayed in Figure 5. The cluster labels have been improved using LLM interpretation based on the automatically generated labels from the clusters and validated by SME input. The cluster labeled "pressure issue with faulty pump," represents the failure mode with the highest downtime among all clusters.

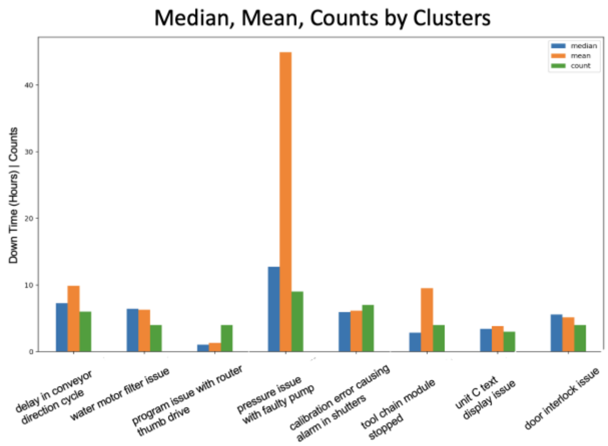

Figure 5. Downtime Summary for Labeled Clusters

4. CONCLUSIONS

This paper demonstrates leveraging LLMs to extract issues and solutions from maintenance operation data, embedding the extracted information, reducing dimensions for clustering analysis, and utilizing the clustering results to provide maintenance operation insights needed by different personnel within a factory setting. The case study optimized for 8 clusters using machine history data from a machine tool, labeling each cluster and grouping similar problems and solutions for quick maintenance insights. A use case piloted an example usage and conducted user interviews with end users.

According to the proposed framework, future work on the practical application of this system calls for additional development, including a convenient user interface to search for related contents from the clustering results. This interface would provide contents together with user queries to an LLM acting as a maintenance copilot. Integrating this proposed system with maintenance history data will enable immediate user insights into the vast amount of maintenance experience and knowledge accumulated in the factory over decades.

REFERENCES

Wolniak, R. (2019). Downtime in the Automotive Industry Production Process – Cause Analysis. Quality Innovation Prosperity, vol. 23, no. 2, pp. 101-118. doi: https://doi.org/10.12776/qip.v23i2.1259

Pertselakis, M., Lampathaki, F., Petrali, P. (2019). Predictive Maintenance in a Digital Factory Shop-Floor: Data Mining on Historical and Operational Data Coming from Manufacturers' Information Systems. In: Proper, H., Stirna, J. (eds) Advanced Information Systems Engineering Workshops. CAiSE 2019. Lecture Notes in Business Information Processing, vol 349, pp. 120-131. Springer, Cham. https://doi.org/10.1007/978-3-030-20948-3_11

GALAR D., GUSTAFSON A., TORMOS B., BERGES L. (2012). Maintenance Decision Making based on different types of data fusion. Eksploatacja i Niezawodnosc – Maintenance and Reliability, vol. 14, no. 2, pp. 135-144.

Teixeira, H., Teixeira, C., Lopes, I. (2021). Maintenance Data Management for Condition-Based Maintenance Implementation. In: Dolgui, A., Bernard, A., Lemoine, D., von Cieminski, G., Romero, D. (eds) Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems. APMS 2021. IFIP Advances in Information and Communication Technology, vol. 632. Springer, Cham. https://doi.org/10.1007/978-3-030-85906-0_64

Lukens, S., & Markham, M. (2018). Data-driven Application of PHM to Asset Strategies. Annual Conference of the PHM Society, 10(1). https://doi.org/10.36001/phmconf.2018.v10i1.245

Lee, J. & Su, H. (2023). A Unified Industrial Large Knowledge Model Framework in Smart Manufacturing. arXiv preprint arXiv:2312.14428.

McInnes, L., Healy, J., Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection. The Journal of Open Source Software, vol. 3, no. 29, pp. 861. doi: https://doi.org/10.21105/joss.00861

McInnes, L., Healy, J., Astels, S. (2017). hdbscan: Hierarchical Density-Based Clustering. The Journal of Open Source Software, vol. 2, no. 11. doi: https://doi.org/10.21105/joss.00205

Bergstra, J., Yamins, D., Cox, D. D. (2013). Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. Proceedings of the 30th International Conference on Machine Learning. PMLR, vol. 28, no. 1, pp. 115-123.

Tong, X., Jung, W. Q., & Banning, J. F. (2023). A Fine-grained Semi-supervised Anomaly Detection Framework for Predictive Maintenance of Industrial Assets. In Annual Conference of the PHM Society, vol. 15, no. 1.

BIOGRAPHIES

Noah Getz was born and raised in San Francisco, California, USA. He earned his B.S. in Industrial Engineering and Operations Research from the University of California, Berkeley, in 2024, where he also pursued minors in Data Science and Electrical Engineering & Computer Science. While at university, Noah conducted research under the Haas School of Business, where he developed a machine learning decision support system to improve digital financial services in rural emerging markets. Concurrently, he interned at RTX' Enterprise Data Services division, where he developed a machine learning model to extract actionable insights from unstructured machine maintenance logs, enhancing the effectiveness of predictive maintenance. Noah currently works as a Data Scientist at M Science in New York City, where he specializes in using artificial intelligence to identify and describe anomalies in financial data. His ongoing research focuses on leveraging artificial intelligence and advanced predictive analytics to optimize industrial and business processes.

Xiaorui Tong was born and raised in Pingliang, Gansu, China. She obtained a B.S. in Mechanical Design, Manufacturing and Automation in 2011 from Central South University in Changsha, Hunan, China, and a M.S. in Mechanical Engineering at the Center for Intelligent Maintenance Systems in 2013 from University of Cincinnati in Cincinnati, Ohio, USA with a research focus on prognostics and health management (PHM). She received her Ph.D. in Material Science and Engineering in 2018 from West Virginia University, where her research on building computational models combining machine learning and material science principles for renewable energy applications was featured as the cover article by the Journal of American Ceramic Society. She worked with companies including the Goodyear Tire & Rubber Company and Parker Hannifin Corporation on developing computational models for various industrial systems. She worked at the FORCAM Inc. on enabling early Industry 4.0 capabilities for manufacturing equipment in collaboration with multiple manufacturing facilities. She is currently an Associate Director of Data Science & Analytics in the Applied AI team within Enterprise Data at RTX. Her work focuses on leveraging domain knowledge and artificial intelligence to enable insights for industrial processes and development of enterprise AI product.