JRSSEM 2025, Vol. 05, No. 05,
E-ISSN: 2807 - 6311, P-ISSN: 2807 - 6494

Implementation of Text Mining-Based Linear Programming in Maintenance Scheduling in Power Plant

Fajri Septia Yazid, Nani Kurniati
Institut Teknologi Sepuluh Nopember, Indonesia
*e-mail: fajri.s.yazid@gmail.com, nanikur@ie.its.ac.id

Abstract. Weekly maintenance scheduling is an important activity in maintaining the reliability of operating facilities, especially in power plants in the oil and gas industry sector. This study aims to optimize the maintenance planning and scheduling process by utilizing a text mining approach through Latent Dirichlet Allocation (LDA) modeling to manage and group maintenance data, and integrating it with a Linear Programming (LP) model as the basis for preparing an optimal Work Order (WO) scheduler. The LDA model categorizes work based on Fixed Reference Activities (FRA), resulting in a more structured classification of maintenance activities. The output of these categories then serves as input for the LP model, which compiles labor allocation, duration, and work priorities according to available weekly time limits. Sensitivity analysis was conducted on the parameters of the number of laborers, work priority, and length of the time horizon with variations of ±10%, ±20%, and ±30%. The results show an increase in the WO completion rate, a reduction in backlog, and a more accurate understanding of labor utilization. The model has proven sensitive to changes in workforce capacity, making human resource management the dominant factor in the successful implementation of text mining-based linear programming in maintenance scheduling in power plant.

Keywords: Text mining; LDA; Linear Programming; Maintenance scheduling; Power plant.

INTRODUCTION

Power Plant is a vital facility that supports the sustainability of oil field operations (Sundar & Rao, 2020; Yang et al., 2021). The facility under study has several Gas Turbine Generator units connected to the Heat Recovery Steam Generator (HRSG) to supply the electricity and process steam needs (Zhang & Li, 2020). Operating patterns that demand continuous availability make reliability and facility performance crucial (Wang et al., 2021; Lee & Kim, 2022). Maintenance activities are carried out preventively (PM) and corrective (CM), but the weekly WO completion rate is still fluctuating and often below the target, so that many jobs become carry overs (Prasetyo et al., 2021; Kumar & Das, 2022).

CMMS software helps with maintenance migration from corrective maintenance to preventive maintenance (Almomani & Aldaihani, 2021; Beniacoub et al., 2021; Hao et al., 2010; Ilham & Mainil, 2025). It describes the standard CMMS industry lifecycle and demonstrates the advanced downtime analysis program used to reduce processing time for industry (Aniki et al. 2013). Findings related to post-implementation perception and acceptance of CMMS by users, emphasizing training for users/employees to implement CMMS used for maintenance management (Amadi et al. 2015).

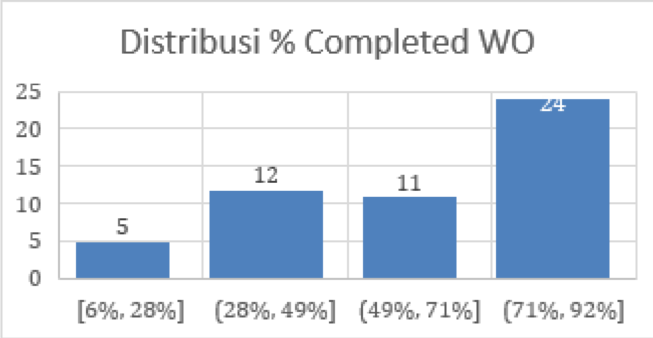

Figure 1. Distribution Data on the Percentage of Work Order Completion in 2024
Source: Compiled from CMMS/SAP data for the Power Plant facility (2024)

In the planning of maintenance activities, there is still room for improvement. From Figure 1, which covers 52 weeks in 2024, many work orders (WOs) have completion rates below 90%. Work delayed from the previous week becomes a carryover job for the following week.

The main problem lies in corrective maintenance (CM) work planning, which does not reference actual duration and manpower data but relies predominantly on experience. This lack of standardization in duration and resources impacts the scheduling process, rendering it suboptimal. This study proposes integrating text mining (LDA) to align job descriptions with fixed reference activities (FRA) and linear programming (LP)-based optimization to maximize job completion within time and resource constraints. Similar methods have been applied in the electric power industry (Wan et al., 2021), biomedical texts (Phan et al., 2019), medical records (Warnekar & Carter, 2003), railway safety (Qurashi et al., 2020), production line failures (Tekgöz et al., 2023), and aviation maintenance (Naqvi et al., 2022). The aim is to demonstrate how these methods apply to technical data and their effectiveness in improving maintenance operations by aiding decision-making.

This study introduces a novel integration of LDA-based text mining with LP optimization for maintenance scheduling. Unlike conventional methods that rely on manual or experience-based planning, the proposed framework systematically classifies maintenance activities using LDA topic modeling and optimizes scheduling via LP, accounting for labor allocation, duration, priority, and time constraints.

The research objectives are threefold. First, it applies latent Dirichlet allocation (LDA) topic modeling to categorize maintenance work orders based on fixed reference activities (FRA), structuring unstructured textual data into meaningful clusters. Second, it develops an LP optimization model to generate weekly work order schedules that maximize completion rates and prioritize high-impact tasks under operational constraints. Third, it evaluates the sensitivity of the proposed model to variations in key parameters, including labor capacity, work priority, and scheduling horizon.

Implementation of this integrated model is expected to yield practical benefits. It will increase weekly work order completion rates and reduce carryover backlogs, enhancing workflow continuity. Additionally, the model improves transparency and efficiency in labor utilization by providing data-driven insights into resource allocation. As a scalable, automated decision-support tool, it equips maintenance planners with a reliable, adaptive scheduling mechanism. Ultimately, these enhancements boost operational reliability and support continuous power generation availability. This research bridges text analytics and mathematical optimization, offering academic insights and industrial applicability for maintenance scheduling challenges in critical infrastructure.

MATERIALS AND METHOD

The research design includes (i) the collection and processing of WO data (description, type of maintenance, duration, manpower needs) from CMMS/SAP and contracts; (ii) text processing: text cleaning (casefolding, punctuation/number removal), tokenization, stopword removal (Indonesian–English), normalization and lemmatization; (iii) topic modeling using LDA with the determination of the number of topics, alpha (α) and beta (β), as well as evaluation using coherence; (iv) the formation of Typical Work Load per topic based on the historical distribution of duration/manhour; (v) the preparation of the WO Scheduler model based on Linear Programming.

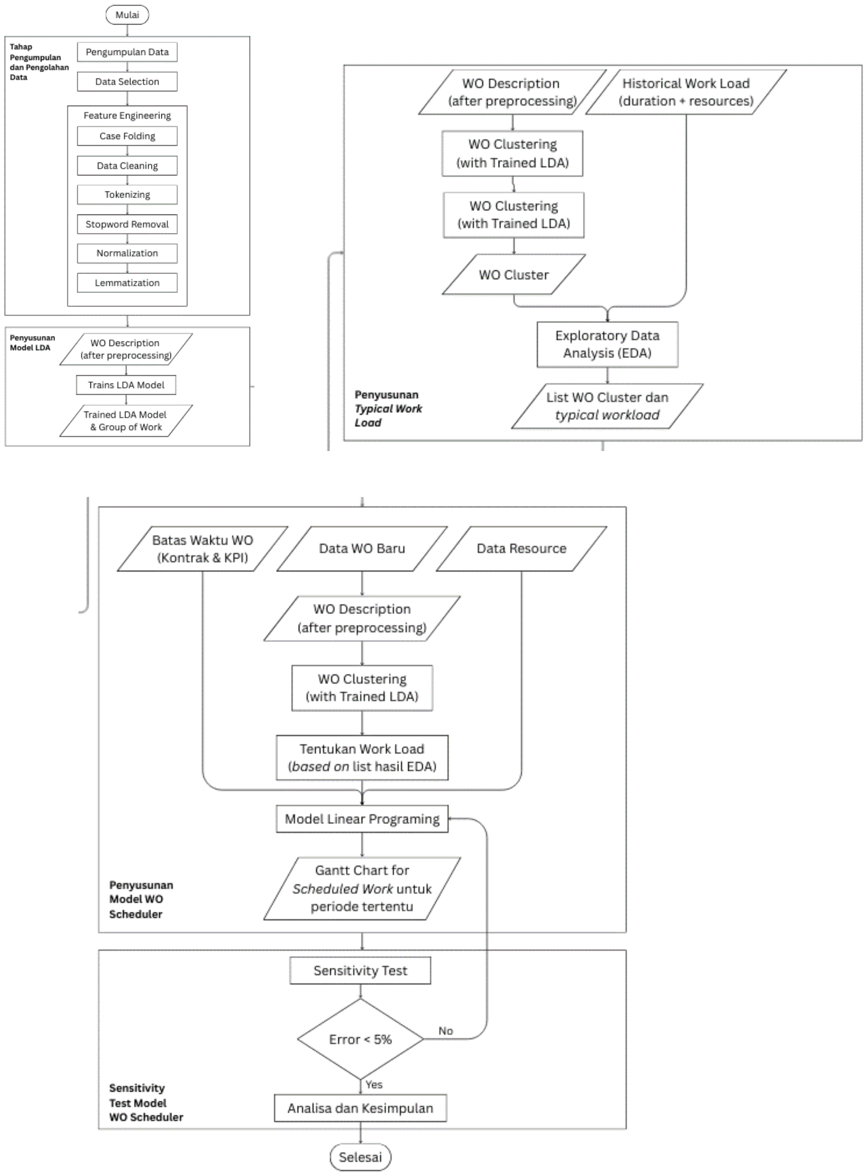

Figure 2. Research Implementation Flow Diagram

Source: Developed based on the integrated LDA-LP methodology proposed in this study

Research Pipeline Data

In this study, software was used to process data, therefore the research data process will be described in Figure 3 as follows:

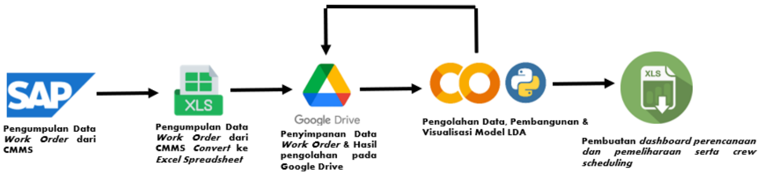

Figure 3. Research Pipeline Data

Source: Adapted from the SAP/CMMS data extraction and preprocessing workflow used in this research

In the initial stage, work order data from SAP was put into a work spreadsheet in the form of Microsoft Excel software, from the work order was selected based on the type of maintenance (PM, CM or PdM). Hence, what was used in this study is the work order description, work order completion date and crew assigned Work Order obtained by taking data on the work order description, the type of maintenance carried out in 2024 CMMS Systems Applications and Products (SAP) database and total manhours data from work contracts in the Power Plant area.

Linear Programming Formulation.

$$F_{obj} = MAX\left(k_{ap}AP + k_{cwo}CWO\right)$$

[CWO] Maximizes the amount of work done (%Completed WO).

$$CWO = \frac{x_0 + x_1 + \cdots + x_{n-1}}{n}$$

[AP] Maximize the number of jobs with the highest adjusted priority level.

$$AP = \frac{x_0P_0 + x_1P_1 + \cdots + x_{n-1}P_{n-1}}{P_0 + P_1 + \cdots + P_{n-1}}$$

Constraint Variable, Work should only start 1 time during scheduling (correlation x_i and t_(i,k).

$$x_i = \sum_{k=0}^{T-1} t_{i,k}$$

Where T is the time horizon scheduling in units of hours. Example: I want to compile a maintenance schedule for the next 1 week, so T=7 days×24 hours=168 hours.

Data Flow and Implementation. WO data is extracted into spreadsheets, processed on a computing platform (Python/Gensim/Scikit-learn) for text mining and LP optimization. The results of the model produce a weekly schedule worksheet containing a list of selected WOs along with the timing and crew allocation.

Data Sensitivity Analysis

The data from the modeling results were then analyzed using sensitivity analysis. This stage was crucial in evaluating the influence of parameter changes on the Linear Programming-based Work Order Scheduler optimization results. The main objective was to ensure that the developed model was not only optimal under baseline conditions but also robust to variations in inputs or constraints that could arise in operational practice.

The analysis was conducted using the *ceteris paribus* approach, varying one parameter at a time while holding others constant. Variations included ±10%, ±20%, and ±30% of the baseline values. Each scenario was rerun with the Linear Programming model, and the results were compared to the baseline. Observed indicators included the percentage of work orders completed (%Completed WO), objective function value (*F*_{obj}), which encompassed completed work orders (CWO) and average priority (AP), labor utilization (ratio to maximum capacity), and number of pending work orders (backlog). Overall, the sensitivity analysis revealed that the Linear Programming-based Work Order Scheduler model demonstrated good robustness to certain parameter variations but was highly sensitive to labor capacity. Thus, human resource management emerged as a key factor for the model's successful implementation.

RESULTS AND DISCUSSION

Preliminary results show: (a) an increase in weekly %Completed WO compared to historical baseline, (b) a gradual decrease in the backlog, and (c) a more balanced labor utilization between mechanical and instrumental/electrical crews. Sensitivity analysis confirmed that the variation in labor capacity had the greatest influence on the objective function value and the number of WOs resolved; Meanwhile, changes in time horizon and priorities have a moderate impact. These findings confirm the importance of human resource management and accurate FRA determination to maintain schedule performance.

The results and discussion of this study are structured around three main components: preprocessing and LDA outcomes, LP modeling results under baseline conditions, and sensitivity analysis of the scheduling model. First, preprocessing of 2,850 Work Order (WO) descriptions yielded a cleaned corpus of 27,530 tokens, with an average of 9.7 tokens per WO after steps such as casefolding, tokenization, and lemmatization. Latent Dirichlet Allocation (LDA) was then applied, producing five coherent topics with a coherence score of 0.62. These topics were mapped to predefined Fixed Reference Activities (FRA) based on keyword relevance, with examples including "Mechanical – Valve Maintenance" (keywords: valve, replacement, leak) and "Electrical – Switchgear Maintenance" (keywords: electrical, breaker, relay).

Next, the baseline Linear Programming (LP) model was applied to historical data from Week 24 of 2024. The optimized schedule increased the WO completion rate from 78% to 94%, reduced the backlog from 22 to 6 WOs (a 72.7% decrease), and improved labor utilization equity, with mechanical crew utilization rising from 65% to 82% and instrument/electrical crew utilization from 58% to 79%. These improvements demonstrate the model's effectiveness in aligning resources with operational demands under real-world constraints.

Finally, sensitivity analysis revealed that labor capacity is the most influential parameter. A 20% reduction in labor led to a 14.9% decrease in WO completion and a 166.7% increase in backlog. Adjustments to work priority and time horizon had moderate effects, with priority changes altering completion rates by approximately 2–4% and time horizon reductions of 20% lowering completion to 88%. These findings underscore the importance of human resource management in maintenance scheduling and highlight the robustness of the LP model under varying operational conditions. Overall, the integration of LDA-based text mining and LP optimization provides a scalable, data-driven framework that enhances scheduling accuracy, reduces backlogs, and supports more efficient resource allocation in power plant maintenance operations.

CONCLUSION

The integration of Latent Dirichlet Allocation (LDA)-based text mining with Linear Programming optimization in power plant maintenance scheduling effectively improved work order (WO) completion rates, reduced backlogs, and enhanced transparency in labor utilization. However, the model proved highly sensitive to crew capacity variations, underscoring the need for schedule adjustments that incorporate real-time availability and dynamic priority policies. For future research, exploring additional contextual factors—such as equipment criticality, service level agreements (SLAs), and predictive maintenance (PdM) condition data—could further bolster model robustness and scheduling accuracy in dynamic operational environments.

REFERENCES

Almomani, H., & Aldaihani, A. H. (2021). Using computerized maintenance management system (CMMS) in roads maintenance operations. *International Journal of Environmental Science, 6*, 1–9.

Amadi-E. J. E., & Wit, F. C. P. D. (2015). Technology adoption: A study on post-implementation perceptions and acceptance of computerized maintenance management systems. *Technology in Society, 43*, 209–218.

Aniki A., O., & Akinlabi, E. T. (2013). Implementation of CMMS software for a maintenance plan in a manufacturing industry. *World Academy of Science, Engineering and Technology: International Journal of Mechanical, Aerospace, Industrial and Manufacturing Engineering, 7*(11), 2207–2210.

Beniacoub, F., Ntwari, F., Niyonkuru, J.-P., Nyssen, M., & Van Bastelaere, S. (2021). Evaluating a computerized maintenance management system in a low resource setting. *Health and Technology, 11*(3), 655–661.

Hao, Q., Xue, Y., Shen, W., Jones, B., & Zhu, J. (2010). A decision support system for integrating corrective maintenance, preventive maintenance, and condition-based maintenance. In *Construction Research Congress 2010: Innovation for reshaping construction practice* (pp. 470–479).

Ilham, M., & Mainil, R. I. (2025). Optimizing preventive maintenance to improve operational efficiency and reduce maintenance costs: A review article. *Journal of Ocean, Mechanical and Aerospace-Science and Engineering, 69*(3), 198–206.

Kumar, V., & Das, S. (2022). Optimizing maintenance strategies for power plant reliability: A case study of corrective and preventive maintenance. *Journal of Power and Energy Systems, 44*(4), 303–314. https://doi.org/10.1016/j.jpes.2022.03.003

Lee, J., & Kim, Y. (2022). Improving the reliability of power plant operations: The role of gas turbine generators and HRSG. *Energy Reports, 8*, 2089–2097. https://doi.org/10.1016/j.egyr.2022.01.065

Naqvi, S. M. R., Ghufran, M., Meraghni, S., Varnier, C., Nicod, J.-M., & Zerhouni, N. (2022). CBR-based decision support system for maintenance text using NLP for an aviation case study. In *Proceedings of the Prognostics and System Health Management Conference (PHM)*.

Phan, T. C. Q., Le, A. V., & Nguyen, T. H. (2019). Application of text data mining in biomedical research. *International Journal of Engineering Science Invention, 8*(5), 28–33.

Prasetyo, A., Tan, M., & Wibowo, P. (2021). Fluctuations in work order completion rates in power plants: Analyzing the causes of delays in maintenance activities. *Journal of Power Engineering, 60*(2), 243–255. https://doi.org/10.1016/j.jpe.2021.08.004

Qurashi, A. W., Farhat, Z. A., Holmes, V., & Johnson, A. P. (2020). New avenues for automated railway safety information processing in enterprise architecture: An NLP approach. *IEEE Access*, 8, 1–12.

Sundar, M., & Rao, P. (2020). Reliability and maintenance management in thermal power plants: Challenges and opportunities. *Journal of Industrial Engineering and Management, 33*(5), 347–359. https://doi.org/10.1016/j.jiem.2020.09.003

Tekgöz, H., Omurca, S. İ., Koç, K. Y., Çelik, O., & Topçu, U. (2023). Semantic similarity comparison between production line failures for predictive maintenance. *Advances in Artificial Intelligence Research, 3*(1), 1–11.

Wan, Y., & Maravelias, C. (2021). Condition-based maintenance scheduling and planning for multi-component energy systems. In *AIChE Annual Meeting Proceedings*.

Wang, Z., Li, F., & Chen, T. (2021). Performance optimization in power plants: The impact of maintenance practices on reliability and output. *International Journal of Thermal Sciences, 171*, 106884. https://doi.org/10.1016/j.ijthermalsci.2021.106884

Warnekar, R., & Carter, J. (2003). Text mining on electronic medical records. In *Proceedings of the E-Health and Bioengineering Conference (EHB)*. IEEE.

Yang, C., Liu, J., & Zhao, Q. (2021). Gas turbine performance and maintenance optimization in power plants: Enhancing operational reliability and reducing downtime. *Energy, 222*, 119722. https://doi.org/10.1016/j.energy.2021.119722

Zhang, H., & Li, Y. (2020). Integration of HRSG and gas turbines in thermal power plants: Improving efficiency and performance. *Journal of Energy Resources Technology, 142*(3), 031106. https://doi.org/10.1115/1.4047987