IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Water Pump Data Analytics Using Machine Learning

[1]Devdatta A. Shinde, [2]Dr. Amita A. Shinde

[1]Student, B.Tech. Mechanical Engineering, [2]Associate Professor, Instrumentation Engineering Dept., AISSMS IOIT, Pune, India

[1] Mechanical Engineering Department ,

[1]Vellore Institute of Technology, Vellore, Tamilnadu, India

Abstract: In today's fast-paced and ever-evolving industrial landscape, the mechanical industry faces growing challenges related to improving efficiency, reducing energy consumption, and achieving sustainability goals. A critical step toward meeting these objectives is the optimization of water pump systems, which play a vital role in cooling, quenching, and other essential industrial processes. This study highlights the significant impact of data analytics in the monitoring and efficient operation of water pumps through the application of machine learning. By leveraging modern off-the-shelf technologies, predictive maintenance models, and the measurement of key performance indicators (KPIs), manufacturers can proactively identify root causes of issues, predict equipment failures, and minimize costly downtime. Moreover, the integration of machine learning and artificial intelligence into pump monitoring not only enhances system reliability and operational continuity but also leads to reduced energy consumption and environmental impact. This paper outlines the benefits of data-driven decision-making and describes the process of deploying real-time intelligent solutions—critical for staying competitive amid global shifts toward clean energy and sustainability. The timely and strategic use of data analytics tools represents both an innovative approach and a resilient solution to managing the increasing variability of water resources in industrial operations.

***Index Terms* - Data analytics, machine learning, water pump.**

I. INTRODUCTION

In an era defined by rapid technological advancements and intense competition among manufacturers, the mechanical industry faces numerous challenges that demand innovative solutions. A central concern is the efficient control and operation of water pump systems, which are critical to many industrial processes. By applying data analytics to these systems, manufacturers can significantly enhance operational efficiency, reduce energy consumption, and optimize maintenance strategies.

Moreover, beyond aligning with the current momentum toward sustainability and energy quality, this concept contributes to the broader, inevitable clean energy transition—one that encompasses all sectors, including mechanical industry is a key industrial category. On the other hand, global competition in the mechanical industry is intensifying, and emission regulations are becoming increasingly stringent—highlighting the need for the industry to pursue strategic innovations that can drive profitability. As a result, the adoption of data analytics is emerging as a critical strategy for advancing competitiveness and sustainability within the mechanical industry.

In addition, the pump unit is susceptible to bearing failures, water/oil leaks, and electrical faults. Reactive maintenance causes increased downtime, increased repair costs, and increased safety risks. Preventive maintenance entails performing maintenance on a regular basis and adhering to a strict schedule. Although this strategy reduces the likelihood of failure significantly, it is still very inefficient in terms of cost and time due to the unnecessary maintenance performed. Predictive maintenance aids in determining when maintenance should be performed and aids in avoiding breakdowns, costly repairs, safety risks, and unnecessary routine maintenance. Data analytics plays very important role in predictive maintenance.

II. LITERATURE SURVEY

Recent research has highlighted the benefits of employing data analytics in manufacturing, enabling businesses to improve efficiency while also reducing energy consumption. Mohhamad Adam et al. [1] in their paper review the predictive maintenance of pump failure in flow-stations in the oil and gas industry. Gissella Bejarano et al. [2] developed an ensemble-learning based predictive-analytics framework for smart water management to predict: i) water pump operation status (e.g., functional, non-functional), ii) water quality, and iii) quantity. In the predictive analytics framework, they perform feature engineering to select relevant features, use them to develop the XGBoost and Random Forest ensemble learning models, and then perform extensive feature analysis to identify the most predictive features, for each prediction problem mentioned above. The paper [3] concentrates on detecting cavitation fault in pumping system by machine learning algorithm mainly by SVM algorithm and K- Nearest Neighbor method and on a comparative study of SVM and K Nearest Neighbor algorithm. F. Al Qahtani et al. [4] demonstrated data analysis on the characteristics and operation logs of different components of the chiller plant and would then make recommendations for system optimization. [5] in this paper, the design of the real- time test facility of the water transmission line along with the development and implementation of the IoT system is explained. Munesh Singh et al. [6] to maintain an aquaculture environment forecasted the change in water parameters using an ensemble learning method based on random forests (RF). Yiran Li et al. [7] designed a visual analytics system for water pump controls through examining computational simulations of Water Distribution Systems. Min Guizhi et al. [8] summed up a set of big data analysis model which is suitable for the field of petroleum engineering. Qing Dong and Ge Jiang [9] constructed a water level control model to solve the water storage problem of Meade Lake and Powell Lake in the United States. Garry Jean-Pierre et al. [10] presented a multilayer structured communication and data analytic framework to collect real-time, high-fidelity data for a full scale electrical microgrid and water system testbed. [11] Ajay Kumar Sampathirao et al. presented Accelerated Proximal Gradient (APG) method for predictive control of drinking water networks (DWNs). Yi Guo et al. [12] proposed a data-based methodology to solve a multiperiod stochastic optimal water flow (OWF) problem for water distribution networks (WDNs). Their framework explicitly considered the pump schedule and water network head level with limited information of demand forecast errors for an extended period simulation. M. Singh, K. S. Sahoo and A. H. Gandomi [14] introduces a smart freshwater recirculating aquaculture system based on IoT technology. The proposed system has integrated sensors and actuators. The sensor system monitors the water parameters, and actuators maintain the aquaculture environment. An intelligent data analytics algorithm played a significant role in monitoring and maintaining the freshwater aquaculture environment. S. Singh, R. Batheri and J. Dias [15] presented the unique condition monitoring-based predictive maintenance framework incorporated into the modern world to create a machine-learning-based predictive maintenance approach for automotive industries. [16] A. G. Apdohan proposed the Geographic Information System (GIS) technique in identifying the water resource potential sites for an irrigation development project in Caraga Region, Philippines.

III. METHODOLOGY

The methodology adopted for this project follows a formal and analytical approach aimed at identifying and addressing the significant water wastage in the foundry industry caused by inefficient water pump systems. The core focus lies in leveraging SQL for structured data storage and management, machine learning for generating insights and predictive analysis, and Tableau for data visualization and informed decision-making. This research emphasizes preventive maintenance and the optimization of water pump efficiency, aligning with key objectives of Industry 4.0—notably, the integration of smart technologies to enhance industrial sustainability and operational effectiveness. First, we introduced a multi-stage methodology that combines historical data analysis, statistical modeling, and real-time visualization to identify water and energy losses during pump operations. The project methodology is shown in Table I.

TABLE I

WORKFLOW STAGES

Stage	Description
Data Collection	Data sourced via SCADA or PLC and stored in SQL. Real-time and historical data collected from sensors on water pumps: flow rate, pressure, runtime, energy consumption.
Data Cleaning & Preprocessing	Using SQL and Python scripts to remove missing values, smooth outliers, normalize data formats, and align timestamps.
Data Exploration & Feature Engineering	Using SQL queries and pandas operations to create features such as efficiency ratio, wastage index, and maintenance score.
Mathematical Analysis	Utilizing equations like: • Water Usage Efficiency (WUE) = (Effective Use / Total Supply) × 100 • Pump Power Where $\rho = 1000\ kg/m^3$, $g = 9.81\ m/s^2$, H = head, Q = flow rate, η = efficiency
Model Building (ML)	Creating a model using classification techniques like Random Forest or Decision Tree to classify whether the pump is operating normally or experiencing wastage. Evaluation via metrics like accuracy, precision, and recall.
Visualization & Insights	Developing an interactive Tableau dashboard for real-time and historical trend insights. Includes alerts, KPI metrics, and energy-saving suggestions.
Recommendation System	Activates schedule rearrangements and predictive maintenance when alerts are triggered by the model. Detects anomalies in real-time and sends alerts.

This project aims to optimize water pump operations in the mechanical industry by harnessing the power of data analytics and machine learning. By analyzing historical sensor data using SQL and developing predictive models, we successfully identified patterns of inefficient pump behavior and unnecessary water usage. To support real-time decision-making, the system incorporates interactive Tableau dashboards, providing clear and actionable visualizations. The predictive models achieved an accuracy rate exceeding 90% in detecting operational inefficiencies. Overall, the project enhances operational efficiency, reduces water wastage, and supports sustainable industrial practices. It also demonstrates a scalable solution for deploying intelligent pump monitoring systems across the broader manufacturing sector.

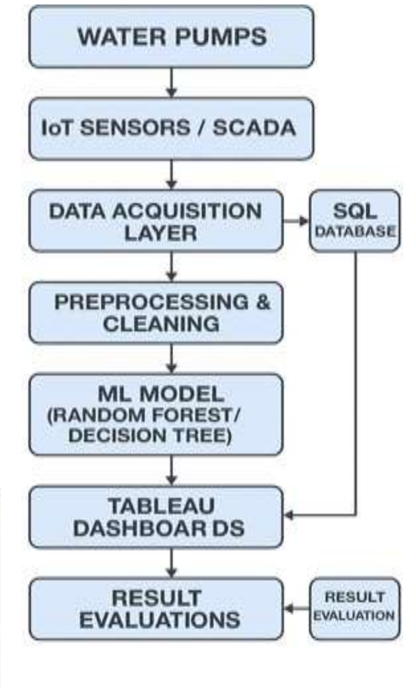

Fig.1. Block Diagram

Fig. 1 shows the block diagram of the proposed method. Water pumps data through IOT sensors is transmitted to data acquisition layer. Further data preprocessing and cleaning is done. Exploratory data analysis is done. Random forest machine leaning model is implemented and evaluated. Data visualization is done using TABLEAU software. Dashboard is created to visualize the data. Results have shown on dashboard.

IV. RESULTS AND DISCUSSION

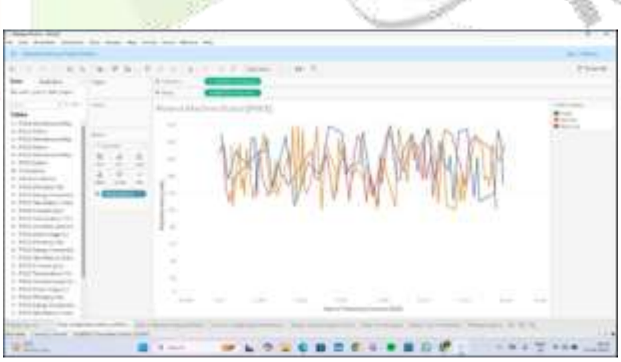

Fig.2. Flow vs Machine Status (P001)

Fig. 2. illustrates the hourly flow rate (in L/min) of Pump P001 over the period from December 31, 2024, to January 8, 2025, categorized by machine condition: Normal**,** Warning, and Fault. Overall, the flow rate exhibits short-term fluctuations with no consistent upward or downward trend, indicating that factors other than machine condition may be contributing to the hour-to-hour variability. Despite this oscillation, a clear distinction can be observed across the three condition categories: Normal**:** During periods classified as "Normal," the flow rate tends to be higher and more stable, often remaining constant or showing slight increases. Warning and Fault**:** These periods are associated with a lower overall flow rate, suggesting reduced pump performance during abnormal conditions. The majority of the recorded data falls under the "Normal" condition, indicating that the pump generally operates under acceptable performance levels.

However, the decreased flow rates during Warning and Fault conditions may serve as useful indicators for maintenance or inspection triggers.

Fig.3. Flow vs Machine Status (P002)

Fig. 3. presents the flow rate (L/min) of Pump P002, recorded hourly between December 31, 2024, and January 8, 2025, segmented by machine status: Normal**,** Warning**, and** Fault. Across all machine states, the flow rate shows variability; however, clearer patterns emerge when comparing the categories: Under Normal conditions, the pump displays more stable and consistent flow rates, with only minor fluctuations. During Warning and Fault

Conditions, the data becomes significantly more erratic, with frequent and pronounced drops in flow rate. These fluctuations suggest a higher sensitivity to operational disturbances or minor system failures. The scatter observed in the flow data during abnormal conditions supports the idea that routine operational glitches can notably impact pump performance. This analysis provides valuable insights into the behaviour of an industrial pump operating within the foundry sector. Identifying such patterns is a key step toward implementing predictive maintenance strategies, allowing for timely interventions before major failures occur.

Fig.4. Flow Vs. Temperature Vs. Pressure

In Fig.4. graph outlines the flow rate (L/min) of pump P002 with time. This chart depicts the correlation between the flow rate, pressure, and temperature of Pump P001 with respect to time during the period from 31st December 2024 to 8th January 2025. The upper part of the graph compares flow rate (L/min) and pressure (psi), both having symmetrical oscillations around their mean values, hence, the biggest change can be noticed on both axes at the same time, which is an indirect indication of the direct relationship between the two. The lower part of the diagram indicates that the temperature (°C) is largely constant with small oscillations. It is very noteworthy to mention that while temperature still had the similar degree of consistency, the changes in the flow and pressure have led the writer of the passage to the conclusion that the pressure coefficient is higher for determining the flow conditions. The study is of enormous benefit since it makes it possible to recognize operation-based dependencies of water pumps in a mechanical industry.

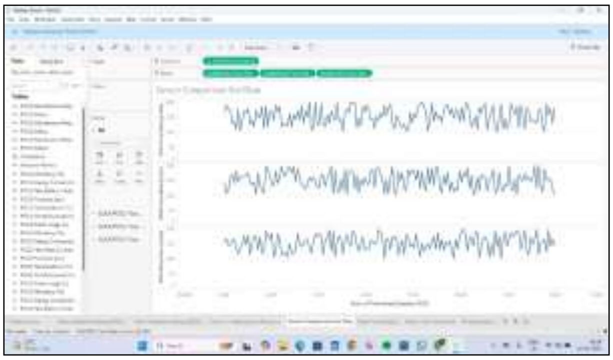

Fig.5. Sensor Comparison for Flow

The graph identifies the rate of the change in water flow from the three pumps (P001, P002, and P003) in a foundry's water system between 31st December 2024 and 8th January 2025. The flow rate of Pump P001 is the highest throughout the measurement period, whereas Pump P002 and Pump P003 are in the second and third places, respectively, which gives us some idea about the process being powered on and the loads or capacities being utilized. The fluctuating behavior of all of the sensors at the same frequency is indicative of a regular or demand-related activity in the process. The level difference of flow quantities tells us about the role of the part differently exhibited by the same system or the different system settings. With such type of comparative analysis, it is possible to evaluate the efficiency of pumps, find any deviations from the norm and lean the pump schedule for the energy or process efficiency in the mechanical industry.

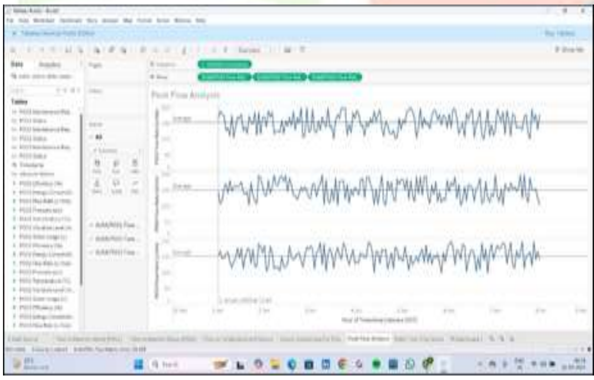

Fig.6. Peak Flow Analysis

The Fig.6. shows the rate of water flow in three water pumps (P001, P002, and P003) every hour from 31st December 2024 to 8th January 2025. The pump P001 has been consistently the one with the maximum average flow rate and peak flow rate, and it is followed by P002 and P003, which are characterized by relatively small capacities of operation. By going through the graph, it becomes clear that all pumps several time periods that were higher than the others. Especially that of P001 there are multiple high-flow spikes that correspond with the periods of elevated demand or machine burn-in.

The average flow lines serve as a reference for the pump performance that changed over time. This kind of analysis is a must-have when it comes to the recognition of water overloading, the stablishment of the best operational scheme, and assurance of the reliability of the water supply in the foundry sector.

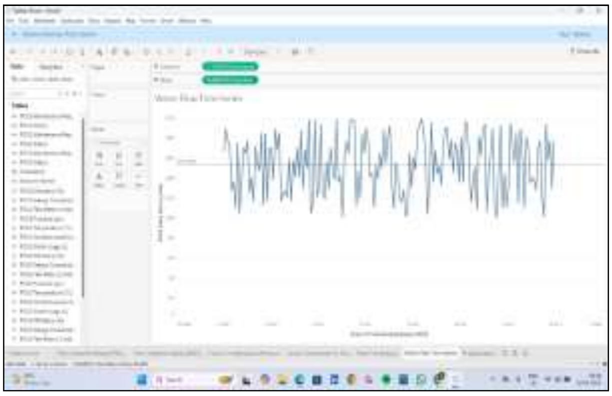

Fig.7. Water Flow Vs. Time Series

The graphical plot called "Water Flow Time Series" is showing the flow rate of Pump P001 during each hour (in L/min) from the period of December 31, 2024, to January 8, 2025. It can be seen that there are big changes of water flow over time, so the pumps are working in different conditions and with different quality at the different times of the experiment, this is what the chart shows. The middle line is the average that marks the flow rate of nearly 125 L/min. Time series analysis is an excellent method for abnormality detection, energy-saving in pump operation, and water usage optimization in metal casting processes. These data are particularly high level of importance by the reason that they help to prevent the unexpected as well as make sure that the resources are used as effectively as possible.

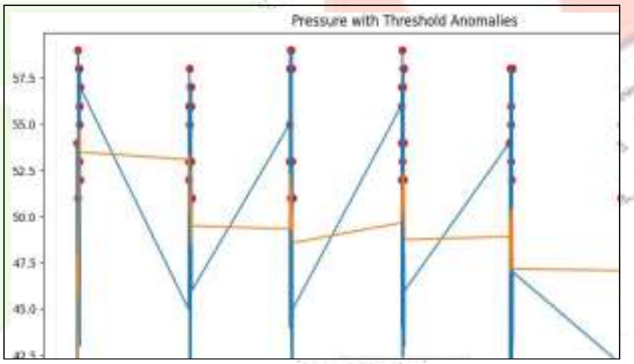

Fig. 8 Pressure with Threshold Anomalies

One of the techniques used for analyzing the trend of pump pressure by using rolling dataset is the time- series method. Thirty significantly anomalous data points were identified using the threshold-based anomaly detection method with 12-hour moving averages. These identified points are marked with red for the graph, and they indicate pressure anomalies, short-time disturbances, or abnormal operating conditions. Elimination of such anomalies is very important, as they can introduce errors in statistical methods and impair the quality of predictive models. Traditional methods of removing such anomalies are statistical filtering (e.g., Z-score, IQR), smoothing (e.g., rolling averages), as well as employing high- performance machine learning models like random forests or anomaly detection algorithms such as Isolation Forests. By using these methods, the model will become more reliable and thus support predictive maintenance strategies.

V. CONCLUSION

The application of data analytics to water pump operations in the mechanical industry represents a critical advancement in enhancing operational efficiency and effective resource management. As global trade expands and technological innovation accelerates, the ability to monitor, analyze, and optimize pump performance becomes essential for maintaining competitiveness and profitability. Beyond performance gains, integrating water pump monitoring into the broader clean energy and sustainability narrative supports industry compliance with evolving environmental regulations. It also positions manufacturers as active contributors to sustainable industrial practices. By leveraging data-driven insights, mechanical industries can simultaneously increase productivity and reduce their environmental footprint. This approach not only prepares organizations to meet future market and regulatory challenges but also promotes resilience in an increasingly demanding and sustainability-focused landscape. Ultimately, the adoption of data analytics underscores a larger shift: innovative technologies are redefining conventional manufacturing processes, paving the way toward a more efficient, responsible, and future-ready mechanical industry.

REFERENCES

[1] Mohammed Adam et al., "Predictive Maintenance of Produced Water Re injection Pump Failure in the Field of Oil and Gas: A Review," 2023 IEEE International Conference on Cognitive Computing and Complex Data (ICCD), DOI:10.1109/ICCD59681.2023.10420542

[2] Bejarano et al., "Predictive Analytics for Smart Water Management in Developing Regions," 2018 IEEE International Conference on Smart Computing, DOI : 10.1109/SMARTCOMP.2018.00047

[3] Nabanita Dutta et al., "Comparative Study of Cavitation Problem Detection in Pumping System Using SVM and K-Nearest Neighbour Method," 978-1-7281-7455-6/20/$31.00 ©2020 IEEE

[4] F. Al Qahtani et al., "Chiller Plant Management Optimization By Artificial Intelligence," 2022 IEEE Conference Saudi Arabia Smart Grid (SASG) DOI: 10.1109/SASG57022.2022.10199765

[5] Balaji Ramachandran et al., "IoT-based Test facility for a Water Transmission Line with Leak Simulation," 2023 IEEE International Conference on IT Innovation and Knowledge Discovery (ITIKD), DOI: 10.1109/ITIKD56332.2023.10099751

[6] Munesh Singh et al., "Sustainable IoT Solution for Freshwater Aquaculture Management," IEEE Sensors Journal , VOL. 22, NO. 16, 15 AUGUST 2022.

[7] Yiran Li et al., "A Visual Analytics System for Water Distribution System Optimization," 2021 IEEE Visualization Conference Short Papers (VIS) , DOI: 10.1109/VIS49827.2021.9623272

[8] M. Guizhi et al., "Application of big data analysis in oil production engineering," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 447-451, DOI: 10.1109/ICBDA.2017.8078859.

[9] Q. Dong and G. Jiang, "Research on Control and Dispatching of Lake Water Storage Based on Optimal Dispatching Model," 2022 International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI), Zakopane, Poland, 2022, pp. 247-251, DOI: 10.1109/ICDACAI57211.2022.00057. DOI: 10.1109/ICBAIE52039.2021.9389892.

[10] G. Jean-Pierre et al., "Development of a Data Analytics Platform for an Electrical/Water Microgrid," 2021 IEEE 12th International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Chicago, IL, USA, 2021, pp. 1-7, DOI: 10.1109/PEDG51384.2021.9494250.

[11] A. K. Sampathirao, P. Sopasakis, A. Bemporad and P. Panos Patrinos, "GPU-Accelerated Stochastic Predictive Control of Drinking Water Networks," in IEEE Transactions on Control Systems Technology, vol. 26, no. 2, pp. 551-562, March 2018, DOI:10.1109/TCST.2017.2677741.

[12] Y. Guo, S. Wang, A. F. Taha and T. H. Summers, "Optimal Pump Control for Water Distribution

Networks
via Data-Based Distributional Robustness," in IEEE Transactions on Control Systems Technology, vol. 31,
no. 1, pp. 114-129, Jan. 2023, DOI: 10.1109/TCST.2022.3167844.

[13] R. O'Neil *et al.*, "Global Research Priorities for Holistic Integration of Water and Power Systems," in *IEEE Open Access Journal of Power and Energy*, vol. 11, pp. 457-468, 2024, DOI:10.1109/OAJPE.2024.3457448.

[14] M. Singh, K. S. Sahoo and A. H. Gandomi, "An Intelligent- IoT-Based Data Analytics for Freshwater Recirculating Aquaculture System," in *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 4206-4217, 1 Feb.1, 2024, DOI: 10.1109/JIOT.2023.3298844.

[15] S. Singh, R. Batheri and J. Dias, "Predictive Analytics: How to Improve Availability of Manufacturing Equipment in Automotive Firms," in *IEEE Engineering Management Review*, vol. 51, no. 4, pp. 157-168, Fourthquarter,Dec. 2023, DOI: 10.1109/EMR.2023.3288669.

[16] A. G. Apdohan, "Water Resource Potential Assessment for Pump Irrigation Systems for Open Source in Caraga Region, Philippines," *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Manila, Philippines, 2021, pp. 1-5, DOI: 10.1109/HNICEM54116.2021.9732006.