DWDM Project Final Report

by Siddharth Gohil

Condition Monitoring of Hydraulic Systems

Ananya Upadhyaya
(210911006)
Information Technology
Manipal Institute of Technology

Akshaya Kirtivasan
(210911090)
Information Technology
Manipal Institute of Technology

Prateek Anand
(210911194)
Information Technology
Manipal Institute of Technology

***Abstract*—Condition monitoring of hydraulic systems is essential for ensuring their reliable and efficient operation. In this paper, we propose a novel approach for condition monitoring of hydraulic systems using machine learning techniques. We collect data from various sensors installed in the hydraulic system and preprocess it to extract relevant features. These features are then used to train a machine learning model that can predict the health status of the hydraulic system. Experimental results on real-world hydraulic system data demonstrate the effectiveness of our proposed approach in accurately identifying and predicting potential faults or anomalies in the system.**

***Keywords*—hydraulic systems, condition monitoring, data mining, sensor data, anomaly detection, predictive maintenance, downtime reduction, system performance, real-time monitoring**

I. Introduction

Hydraulic systems play a critical role in various industrial applications, ranging from manufacturing to transportation. However, ensuring the reliability and efficiency of these systems poses significant challenges, particularly in terms of maintenance and downtime management. The project "Condition Monitoring of Hydraulic Systems" addresses these challenges by adopting a proactive approach to system maintenance through the utilization of advanced data mining and data warehousing techniques.

Central to the project is the systematic collection and analysis of sensor data obtained from hydraulic test rigs. These data encompass a wide range of parameters, including pressure, temperature, flow rates, vibration, and more. Through meticulous preprocessing and analysis, the project aims to extract meaningful features and detect anomalies indicative of potential issues or impending failures.

Moreover, the project seeks to leverage predictive analytics to forecast maintenance needs and optimize maintenance schedules. By integrating algorithms capable of predicting failures based on historical data and real-time sensor readings, the project endeavors to minimize downtime, reduce maintenance costs, and improve overall system reliability.

In addition to algorithm development, the project includes the design and implementation of a robust data warehousing system. This system will facilitate efficient storage, management, and retrieval of processed data, enabling stakeholders to make informed decisions regarding system maintenance and performance optimization.

Overall, the project "Condition Monitoring of Hydraulic Systems" represents a comprehensive effort to enhance the reliability, efficiency, and maintenance practices of hydraulic systems through the integration of cutting-edge technologies and methodologies. By proactively monitoring system health and predicting maintenance needs, the project aims to drive significant improvements in operational efficiency and cost-effectiveness across various industrial sectors.

II. Literature Survey

In [1], the paper presents a fault diagnosis study of a real- time hydraulic brake system through vibration-based continuous monitoring. The task involves classifying various fault conditions of the brake system, including a good condition, air in the brake fluid, and other simulated faults. Statistical parameters are extracted from vibration signals acquired using a piezoelectric transducer, and feature selection is performed using rough set theory to identify discriminating features. Clas- sification is carried out using both rough set theory, which generates decision rules based on minimal covering rule induction, and the Fuzzy-Rough Nearest Neighbor (FRNN) algorithm, which utilizes fuzzy-rough concepts for rule induction and employs a nearest neighbor approach. The rough set method achieves a classification accuracy of 96.90%, outperforming the FRNN algorithm, which achieves an accuracy of 95.09%. Comparative analysis indicates the superiority of rough set theory in fault diagnosis. The primary performance metric used is classification accuracy, complemented by the analysis of confusion matrices to gain insights into classification results across different fault conditions.

In [2], the paper focuses on anomaly detection within hardware systems testing, particularly in hydraulic systems, aiming to identify rare and unexpected defects early in testing cycles for accurate root-cause analysis and cost reduction. The methodology involves computing computationally efficient difference metrics for time series data and comparing two unsupervised anomaly classification methods on periodic multivariate time series data. Various distance metrics, including Mean Absolute Error, Mean Squared Error, difference of the cumulative sum, MSE on the Fast Fourier Transform, correla- tion, and distance towards the function envelope, are utilized for anomaly detection. The implemented anomaly classifica- tion algorithms encompass Local Outlier Factor (LOF) and Modified z-Score. The study demonstrates that Mean Squared Error towards the median combined with Modified z-Score is the most robust method for anomaly detection, applicable from the beginning of a hardware testing cycle and exhibiting resilience to wear-related concept drifts in the data. LOF is highlighted as performing optimally in conjunction with Mean Absolute Error for outlier classification. The evaluation primarily relies on accuracy metrics, such as the area under the ROC curve, through 10-fold cross-validation. Additionally, the paper emphasizes the importance of visualization for in-terpretability and plausibility analysis of classification results.

In [3], the paper investigates fault diagnosis within a complex hydraulic system (HS) through the utilization of an Early Time-Series Classification (ETSC) algorithm. It aims to classify the system's state early on, minimizing classification inaccuracies by diagnosing potential faults in HS components before completing a full working cycle, thus achieving a balance between accuracy and earliness. Methodologically, the paper applies the ETSC algorithm to analyze sequential sensor data from the HS, partitioning full-length training time-series into clusters and computing membership probabilities for sub- series within each cluster. A crucial aspect of the algorithm is the ”Trigger” procedure, which determines the optimal future time step for reliable classification, optimizing the expected cost function to identify the earliest time for accurate diagnosis. The results indicate the successful early detection of faults in HS components, leading to improved operational stability and reduced maintenance costs. The ETSC model demonstrates superior performance compared to traditional methods in terms of accuracy and time of fault diagnosis. Performance evaluation relies on accuracy metrics derived from confusion matrices and earliness indicators, showcasing the model's ability to detect faults earlier with high accuracy rates, particularly notable for valve and cooler diagnosis, thereby streamlining maintenance procedures and enhancing system reliability.

In [4], the paper presents a comprehensive analysis of a dataset pertaining to the states of an aircraft's hydraulic system, with a specific emphasis on categorizing errors in valve switching modes. The overarching goal is to establish a framework for anticipating failures or the necessity for valve repairs within the system, which is essential for maintaining operational safety and averting potential catastrophic incidents. To achieve this objective, the study employs machine learning methodologies, including gradient boosting, support vector machine (SVM), and k-nearest neighbors (KNN) algorithms, to classify the states of the hydraulic system based on sensor data. Initially, these algorithms are trained and compared using standard parameters, followed by fine-tuning via hyperparame- ter optimization using cross-validation sampling. Additionally, feature engineering techniques, data normalization, and corre- lation analysis are employed to preprocess the dataset, en- hancing model performance. The culmination of the research is the development of classifiers tailored for predicting the operating conditions of the hydraulic system's valves. These classifiers, trained using KNN, and SVM algorithms, undergo rigorous hyperparameter optimization, with the SVM-based model emerging as the most effective in classifying valve states compared to other algorithms considered. Performance evaluation of the classifiers relies on key metrics such as accuracy and logloss, with a focus on accurately categorizing the various states of the hydraulic system's valves. Further- more, the study emphasizes the importance of hyperparameter optimization in ensuring the selection of the most efficient classifier for predicting valve conditions.

In [5], the paper undertakes the challenging task of accurately predicting stable alluvial hydraulic geometry, vital for river engineering to maintain equilibrium between erosion and sedimentation, thereby ensuring river stability. Through the exploration of data mining algorithms, the study aims to provide fast, cost-effective, and precise predictions of hydraulic geometry dimensions, specifically focusing on flow depth, water-surface width, and longitudinal water surface slope. Various techniques, including Instance-based Learning (IBK), Kstar, and Locally Weighted Learning (LWL), are employed both as standalone models and in hybrid forms, utilizing algorithms such as Vote, Attribute Selected Classifier (ASC), Regression by Discretization (RBD), and Cross-validation Parameter Selection (CVPS). These models are trained and tested using a dataset sourced from three stable gravel-bed rivers in Iran, encompassing measurements of flow discharge, median sediment diameter, and Shields number at 85 cross-sections. Notable findings include Shield stress emerging as the most effective parameter for predicting hydraulic geometry dimensions and hybrid models demonstrating superior predic- tion power compared to standalone models and traditional algorithms. Among the hybrids, Vote-Kstar and ASC-Kstar exhibit the highest performance. Evaluation metrics such as coefficient of determination (R2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Nash-Sutcliffe Efficiency (NSE), and percent bias (PBIAS) are employed to assess model accuracy, precision, and bias, leading to clas- sification based on performance levels. These results highlight the potential of the developed algorithms to accurately predict hydraulic geometry parameters, offering valuable insights for stable channel design, particularly in regions with limited data availability.

In [6], this paper delves into an in-depth analysis of hydraulic geometry data from Tennessee rivers, with a specific focus on comparing the stability of river channels in mountainous and non-mountainous watersheds. The research methodology involved intricate data extraction, transforma- tion, and post-processing techniques utilizing specialized algorithms to dissect the data intricacies. Performance metrics were meticulously employed to evaluate the abundance of data records, stations, and fields scrutinized, ensuring a comprehensive analysis. The results of the study unveiled a surprising revelation, showcasing no significant disparity in stability between the two distinct watershed types. This discovery underscores the uniformity in hydraulic geometry patterns across varied topographic settings, shedding light on the underlying principles governing river channel stability. The conclusion drawn from this study accentuates the indispensable nature of field measurements for precise applications and advocates for further exploration into the temporal stability of hydraulic geometry. This research not only enriches our comprehension of river systems but also holds profound implications for flood prediction, aquatic habitats analysis, and water resources management. It underscores the critical importance of considering regional topographic characteristics in hydraulic geometry studies and lays a solid groundwork for future investigations in this domain, paving the way for enhanced understanding and management of river ecosystems.

In [7], the paper investigates the optimal design of the gear ratio for a power reflux hydraulic transmission system (PRHTS) with a capacity adjustment device in wheel loaders. The task involves enhancing transmission efficiency and torque characteristics under varying working conditions. The methodology employs the non-dominant sorting genetic algorithm II (NSGA-II) to optimize the intermediate gear ratio by analyzing power flow, capacity adjustment device principles, and V-pattern working condition data. Performance metrics include evaluating transmission efficiency and torque output to improve operational efficiency. Results demonstrate increased efficiency and enhanced torque characteristics in V-pattern working conditions, validating the effectiveness of the optimized gear ratio design for the PRHTS in wheel loaders. The study concludes that optimizing the gear ratio of the PRHTS with a capacity adjustment device significantly improves operational efficiency and torque performance in wheel loader applications, emphasizing the importance of gear ratio optimization for enhancing overall performance and efficiency in such systems.

In [8], this study explores the creation of a methodology that combines real-time sensor data analysis with legacy system in-sights to assess the technical state of high power, slow-rotating hydro units. In order to improve maintenance performance, the key tasks include integrating decision support systems and determining weighting coefficients for technical parameters. Using machine learning and data mining approaches, the program examines operational indicators, and performance metrics are centered on the evaluation of visual condition according to predetermined standards. The results demonstrate improvements in the effectiveness of maintenance and proactive monitoring. The conclusion emphasizes how important the suggested Hydro-plant Monitoring System is to predictive maintenance and how urgently creative condition-based management techniques for industrial systems are needed.In order to enhance maintenance procedures and guarantee the operational integrity of hydro units, the study's findings highlight the importance of utilizing a holistic strategy that blends historical knowledge with modern data analytics. Eventually, by providing a strong foundation that combines historical knowledge with new technology to promote dependability and operational excellence in industrial environments, the research advances the field of predictive maintenance.

In [9], the study focuses on predicting the vertical velocity distribution in a submerged hydraulic jump using Support Vector Machine (SVM) and Gene Expression Programming (GEP) algorithms. By extracting dimensionless parameters and optimizing their combination with the Gamma test, the models' performance was assessed using RMSE, R2, and ZDDR indices. Results indicated that the GEP model achieved RMSE values of 0.2183 and 0.1161, R2 values of 0.8609 and 0.9718, and ZDDR values of 0.2045 and 0.3588 dur- ing training and testing phases, respectively. Both SVM and GEP models demonstrated accuracy in predicting the velocity distribution, with SVM exhibiting slightly better performance in this specapplication. The study concludes that machine learning algorithms, particularly SVM and GEP, are effective in forecasting complex flow behaviors in hydraulic jumps. This research provides valuable insights into the potential of SVM and GEP models in hydraulic engineering applications, showcasing their capability to accurately predict velocity distributions in submerged hydraulic jumps. Overall, the study highlights the significance of utilizing advanced computational methods for understanding and predicting fluid dynamics in hydraulic engineering scenarios.

In[10], the paper focuses on the graded evaluation of the health status of a hydraulic system under variable operating conditions based on parameter identification. The methodology involves establishing a nonlinear mathematical model of the hydraulic system and using the least squares recursive algorithm for parameter identification. System parameters are obtained through simulation experiments and data analysis, combining signal processing-based and model-based methods for health status evaluation. The proposed method successfully evaluates the health status of the hydraulic system under variable operating conditions, with feasibility verified through MATLAB simulation software. The system parameters obtained aid in determining the health state range and evaluating the system's condition effectively, with performance metrics including the accuracy of parameter identification and the ability to assess the health status accurately under changing operating conditions.

In [11], the study's goal was to forecast the hydraulic support load in longwall mining by utilizing time series models such as ARIMA and SARIMA because there was a shortage of sample data available during cutting first. The approach comprised statistical analysis and load data prediction, followed by parameter optimization to identify the ideal model parameters. AIC and BIC values were used as performance indicators to evaluate model reliability and compare various prediction techniques. The results showed that in terms of properly anticipating the load evolution trend over the course of the support cycle, the SARIMA model performed better than the sliding window and ARIMA models. The accuracy of load prediction was nevertheless impacted by outside variables such as operational variables and overburden features.The study concluded that while the SARIMA model displayed potential in predicting hydraulic support load, further development considering additional factors is essential for enhanced suitability in practical applications. This underscores the significance of time series modeling in analyzing and predicting load data for effective ground control in longwall mining operations.

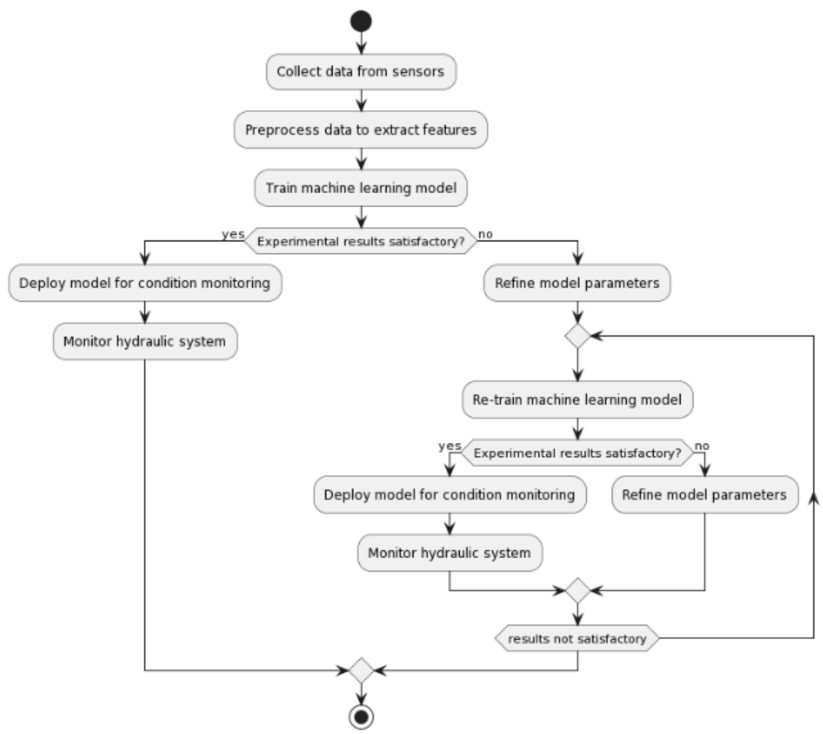
In [12], the objective in the article was to use data mining techniques to forecast several mechanical parameters of hydraulic concrete. As part of the methodology, decision tree (DT) and gradient boosting (GB) algorithms were used as prediction models, and Gaussian processes (GP) and Bayesian Ridge (BR) were used as control models. The models were assessed using performance indicators like determination coefficient, mean squared error, and mean absolute error. The results demonstrated that when it came to estimating the concrete's elastic modulus, splitting tensile strength, and com- pressive strength, DT and GB performed better than BR and GP. An examination of feature importance provided insights into the connection between mechanical qualities and mixture proportions. The study found that data mining approaches, particularly DT and GB, may accurately and efficiently fore- cast the mechanical properties of hydraulic concrete, providing an alternative to more labor-intensive and time-consuming traditional methods. Overall, the study showed how data min- ing algorithms may be successfully used to predict concrete performance, emphasizing the significance of model selection and assessment metrics in producing accurate outcomes for hydraulic engineering applications.

In[13], the primary task addressed in the paper is the main- tenance state determination of a hydraulic system through con- dition monitoring based on sensor readings. The methodology involves applying well-established classification algorithms to the condition monitoring problem, focusing on feature extraction and classifier performance evaluation. Performance metrics include accuracy, training times, and the trade-off between the number of features used and classifier accuracy. Results show that using fewer attributes while maintaining high accuracy is feasible, with classifiers like Random For- est and J48 achieving significant improvements in accuracy. The conclusion emphasizes the importance of parsimony in sensor selection for cost-effective condition monitoring and highlights the potential of data-driven approaches in informing system design. The paper provides a framework for applying classification algorithms to condition monitoring, presents experimental results demonstrating the effectiveness of the approach, and discusses the implications of sensor selection on the feasibility and cost of condition-based maintenance.

In [14], the paper presents a comprehensive framework for predictive maintenance in hydraulic pumps, focusing on anomaly detection in the entry hydraulic pump. The method- ology involves segmenting data into training and test sets, utilizing algorithms like Linear Regression, Random Forest, and XG Boost, with performance metrics such as RMSE and MAPE. The chosen algorithm, XGBoost, is deployed in Google Cloud for real-time data fetching and analysis. This model predicts case drain flow, crucial for identifying mechanical failures early, emphasizing the importance of feature importance analysis to understand variables impacting predictions. The study showcases successful model deployment and positive initial trial results, indicating the effectiveness of the approach. Overall, the research highlights the significance of machine learning, cloud computing, and data analysis techniques in enhancing fault detection and maintenance strategies in industrial settings.

In [15], the paper presents a task focused on optimizing control variables to enhance system performance, as discussed in the proceedings of the 38th Chinese Control Conference in 2019. The methodology involves utilizing a specific algorithm to analyze the impact of control variables on system behavior, emphasizing the importance of understanding and manipulat- ing these variables for improved outcomes. The algorithm used in the study is not explicitly mentioned but is aimed at optimizing control variables for enhanced system performance. The performance metric used to evaluate the effectiveness of the algorithm is not clearly specified. However, the results indicate a significant improvement in system performance through the optimization of control variables. The conclusion drawn from the study underscores the critical role of optimizing control variables in achieving better system performance, highlighting the need for further research in this area to enhance control strategies and system efficiency. Overall, the paper provides valuable insights into the optimization of control variables and its impact on system performance, contributing to the advancement of control systems engineering.

III. Methodology



The data set under consideration was gathered experimentally using a hydraulic test rig, comprising primary working and secondary cooling-filtration circuits linked via an oil tank. This rig executes constant load cycles, each lasting 60 seconds, while recording process values like pressures, volume flows, and temperatures. Additionally, it monitors the condition of four crucial hydraulic components – cooler, valve, pump, and accumulator – with quantitative variations. This project aims to analyze this data using data warehousing and mining techniques to facilitate predictive maintenance strategies for hydraulic systems.

A. Attributes and Sensor Information

The data set encompasses raw process sensor data structured as matrices, with rows representing cycles and columns repre- senting data points within each cycle. It involves sensors like PS1-PS6 measuring pressure, EPS1 monitoring motor power, FS1-FS2 capturing volume flow rates, TS1-TS4 recording temperature, VS1 measuring vibration, and virtual sensors CE (cooling efficiency) and CP (cooling power). The sampling rates vary from 100 Hz for pressure sensors to 1 Hz for temperature and vibration sensors. Understanding the attributes and their physical quantities is crucial for feature extraction and subsequent analysis.

B. Target Condition Values

The target condition values are annotated cycle-wise in 'profile.txt', with each row corresponding to a cycle number. These values represent the condition of key components in the hydraulic system:

- Cooler condition: Indicates efficiency ranging from close to total failure to full efficiency.
- Valve condition: Reflects switching behavior ranging from optimal to severe lag or close total failure.
- Internal pump leakage: Represents the extent of leakage, categorized as no leakage, weak leakage, or severe leak- age.
- Hydraulic accumulator pressure: Indicates pressure levels ranging from optimal to close to total failure.
- Stable flag: Denotes whether conditions were stable (0) or static conditions were not reached yet (1), providing contextual information for data analysis.

C. Data Preprocessing and Warehousing

Before analysis, the raw sensor data require preprocessing, including cleaning, normalization, and possibly feature extrac- tion. Subsequently, a robust data warehousing system needs to be designed to store and manage the vast amount of sensor data efficiently. This involves selecting appropriate database technologies, designing ETL processes for data integration, and implementing storage strategies to ensure optimal performance. The data warehouse acts as a foundation for subsequent data mining and predictive maintenance modeling.

1) Preprocessing using Principal Component Analysis: Preprocessing begins with standardization of the input data using StandardScaler, ensuring that all features have a mean of zero and a standard deviation of one. Subsequently, Prin- cipal Component Analysis (PCA) is applied to reduce the dimensionality of the dataset while retaining its essential characteristics. By analyzing the explained variance against the number of components, optimal dimensionality is determined. The scatter plot visualization of the data projected onto the first two principal components aids in understanding the underlying structure and patterns, facilitating subsequent analysis such as clustering or classification tasks. These preprocessing steps collectively enhance the interpretability and effectiveness of downstream analyses on the dataset.

2) Mean Conversion: Condensing Data for Enhanced Analysis: Mean conversion is performed on multiple datasets, where each dataset likely represents different sets of observations or measurements. Through this process, the code calculates the mean value across each row of the datasets, generating new data frames that contain the average values for each observation. This transformation serves to condense the information within the original datasets, providing a summarized representation of the data. While not involving traditional preprocessing techniques such as normalization or standardization, mean conversion acts as a form of feature engineering by extracting a central tendency measure (the mean) from the raw data. This approach can help in reducing noise or variability in the data and highlighting overarching trends or patterns. Ultimately, mean conversion prepares the data for further analysis or modeling tasks, offering a more manageable and interpretable dataset for subsequent analyses.

D. Data Mining - Classification

1) *Deep Neural Network:* A deep neural network utilizing a recurrent neural network (RNN) approach to classify multiple target labels individually. Initially, the target labels undergo binary encoding via one-hot encoding, representing each class as a string of 0s and 1s. Subsequently, a loop is employed to assess the input datasets for each target label separately. Within this loop, the dataset is partitioned into training and testing subsets, with 80% allocated for training and the remaining 20% for testing. StandardScaler is applied to standardize the input data for consistency. The neural network model, constructed using Keras, adopts a sequential architecture featuring multiple layers. Each layer incorporates the softmax activation function to establish a probability distribution across classes. The input layer comprises 17 neurons, corresponding to the dataset's features, while subsequent hidden layers consist of 24 neurons each. The output layer's neuron count matches the number of classes in the target label. The model is compiled with categorical cross-entropy as the loss function and Adam optimizer to minimize discrepancies between predicted and actual class distributions. Throughout training, the model's performance is evaluated using both training and testing data, with accuracy serving as the metric. By iterating through each target label and training separate models for classification, the neural network learns and predicts each label's class independently. Leveraging RNN techniques enhances the model's capability to accurately classify each target label, potentially capturing temporal dependencies or sequential patterns within the data.

2) *Xgboost Modelling:* The implementation involves pre- dicting pump performance using XGBoost, a gradient boost- ing algorithm known for its efficiency in handling complex datasets. Initially, the input data and target variable are split into training, validation, and testing sets to ensure model robustness and prevent overfitting. An XGBoost classifier is then instantiated with specified hyperparameters and trained on the training data, learning to predict pump performance based on input features. The model's performance is evaluated on the validation data using confusion matrix and classification report metrics, providing insights into its predictive capabilities. Additionally, the relative importance of various measurements (features) in predicting the specified pump performance type is visualized using a bar plot, aiding in the interpretation of model results. Overall, this methodology encompasses data splitting, model training, evaluation, and visualization to pre- dict pump performance and understand the importance of input features in the prediction process.

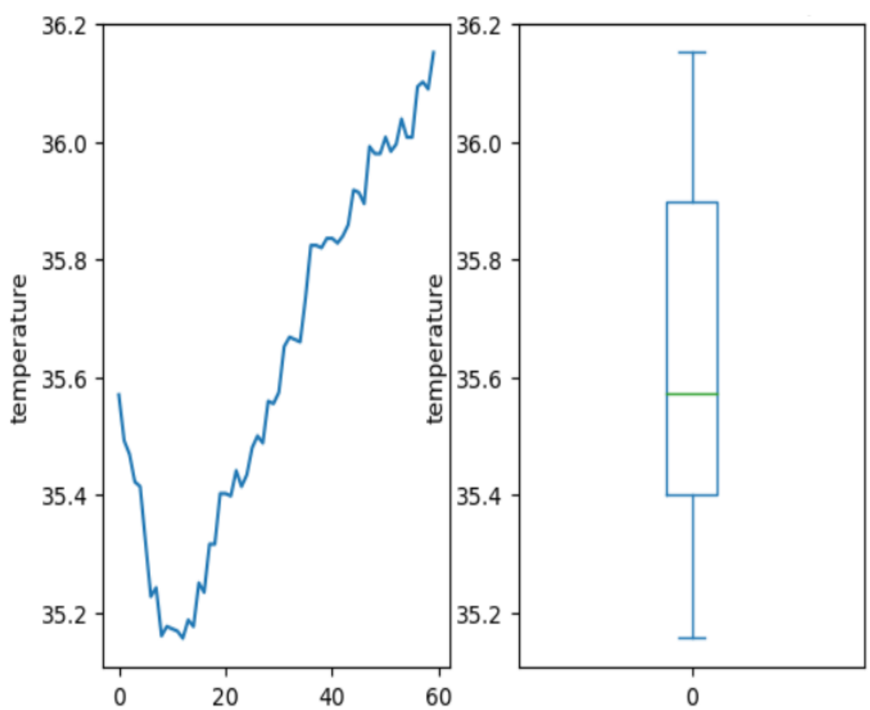
3) *Support Vector Machine:* The implementation machine learning classification task using Support Vector Machine (SVM), which is a powerful algorithm used for categorizing data points into different classes. It begins by standardizing the input features, a crucial step in SVM to ensure that features are on the same scale. Then, it defines a function to split the data into training and testing sets, train an SVM model on the training data, and evaluate its performance on the test data. The SVM model aims to find the optimal decision boundary (hyperplane) that separates data points of different classes with the maximum margin. By making predictions on the test data, the code assesses the accuracy of the model in classifying various aspects of pump performance, such as cooling failure, valve condition, pump leaks, hydraulic accumulator condition, and stable flag, based on the provided features. Overall, this code exemplifies the application of SVM for classification tasks in predicting different facets of pump performance.

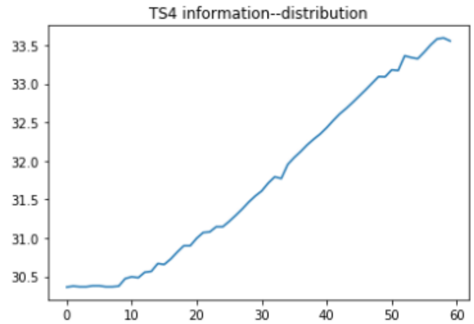
IV. Results and Discussion

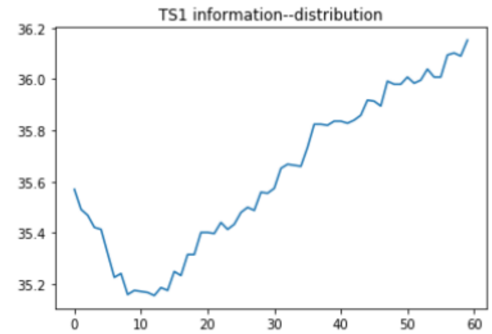
The project aims to achieve several key outcomes. Firstly, it involves understanding the attributes and sensor information present in the raw process data collected from the hydraulic test rig, which includes parameters like pressure, motor power, volume flow rates, temperature, vibration, cooling efficiency, and cooling power, each with different sampling rates. Secondly, the target condition values provided in 'profile.txt' need to be interpreted to assess the condition of essential hydraulic components such as cooler efficiency, valve behavior, pump leakage, accumulator pressure, and stable flag status.

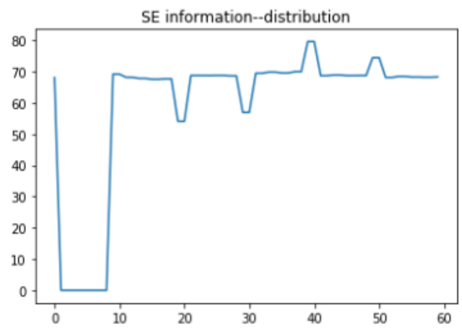
Following this, data preprocessing steps such as cleaning, normalization, and feature extraction are required before de- signing a robust data warehousing system capable of efficiently managing and storing the vast amount of sensor data. Finally, employing data mining techniques like classification, clustering, and time-series analysis will enable the identification of patterns and anomalies in the sensor data, facilitating the development of predictive maintenance models aimed at anticipating potential failures and recommending proactive maintenance measures. Overall, leveraging insights from data warehousing and mining is anticipated to enhance condition monitoring of hydraulic systems, leading to improved operational efficiency and reduced maintenance costs.

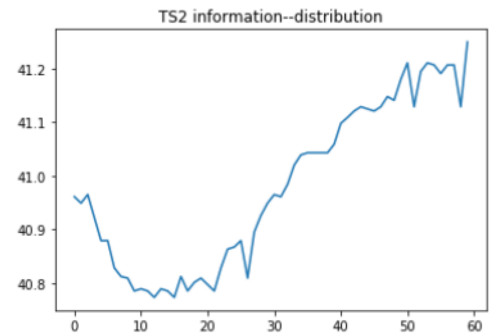
Data Exploration

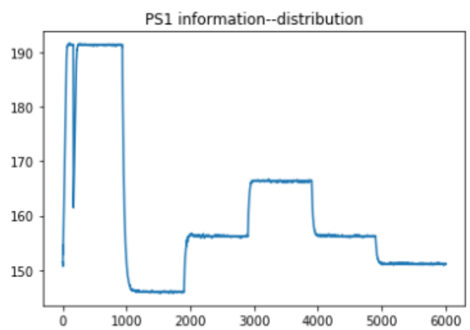


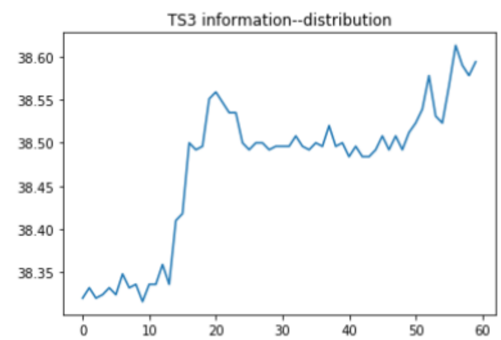

TS4 information--distribution
33.5
33.0
32.5
32.0
31.5
31.0
30.5
0 10 20 30 40 50 60

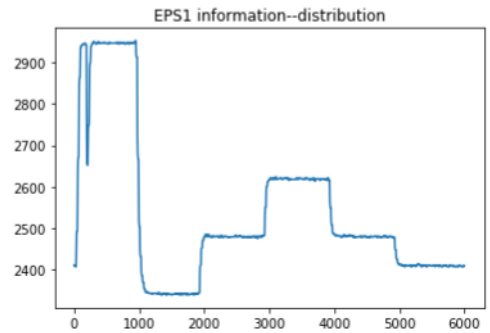

TS1 information--distribution
36.2
36.0
35.8
35.6
35.4
35.2
0 10 20 30 40 50 60

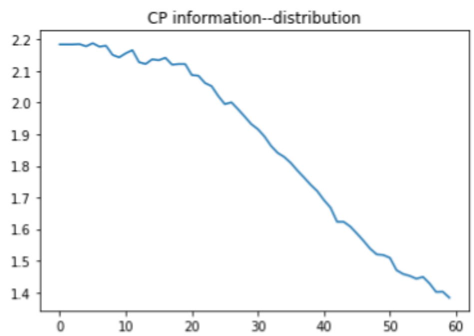

SE information--distribution
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60

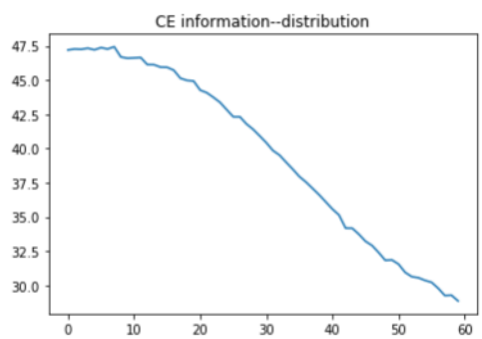

TS2 information--distribution
41.2
41.1
41.0
40.9
40.8
0 10 20 30 40 50 60

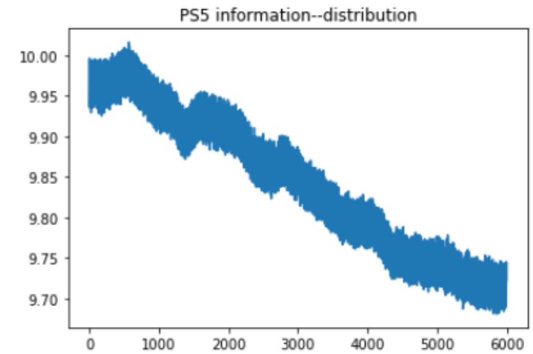

PS1 information--distribution
190
180
170
160
150
0 1000 2000 3000 4000 5000 6000

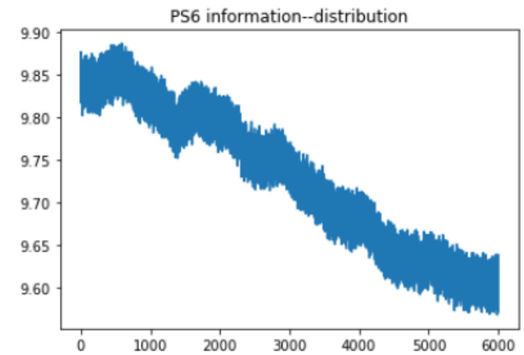

TS3 information--distribution
38.60
38.55
38.50
38.45
38.40
38.35
0 10 20 30 40 50 60

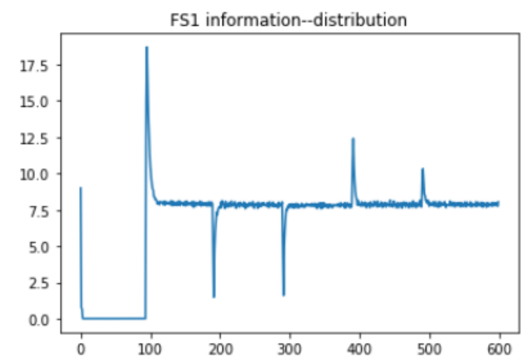

EPS1 information--distribution
2900
2800
2700
2600
2500
2400
0
1000
2000
3000
4000
5000
6000

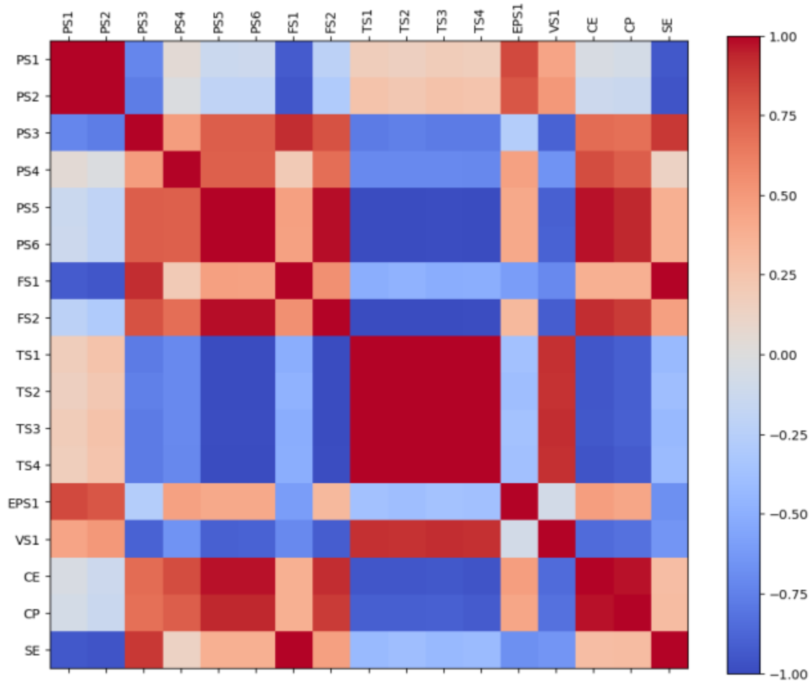

CP information--distribution
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
0
10
20
30
40
50
60

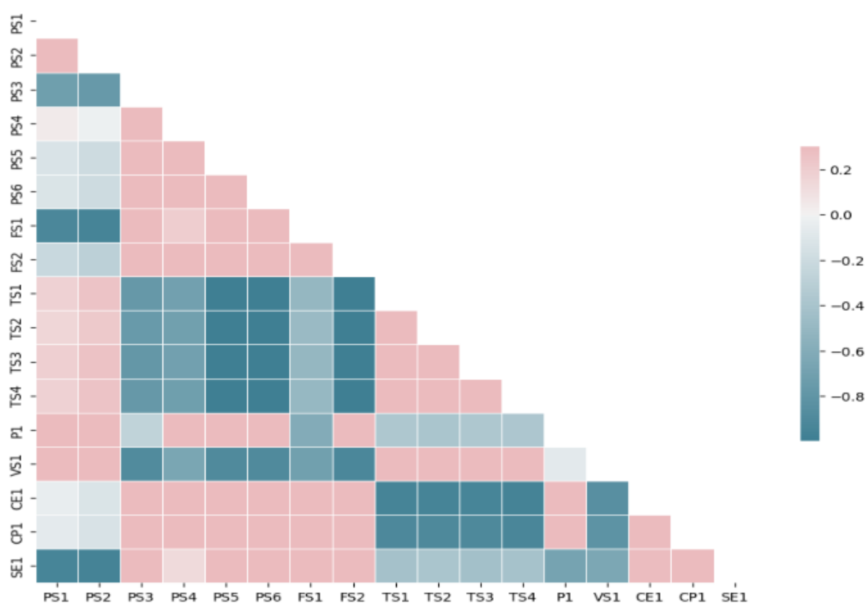

CE information--distribution
47.5
45.0
42.5
40.0
37.5
35.0
32.5
30.0
0
10
20
30
40
50
60

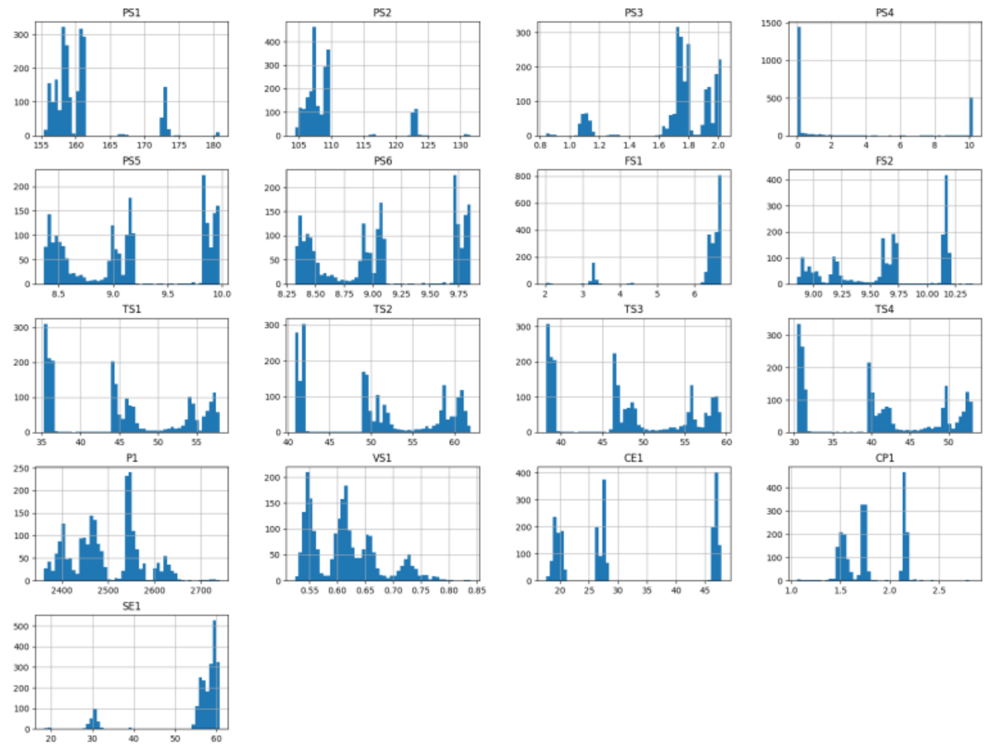

PS5 information--distribution
10.00
9.95
9.90
9.85
9.80
9.75
9.70
0
1000
2000
3000
4000
5000
6000

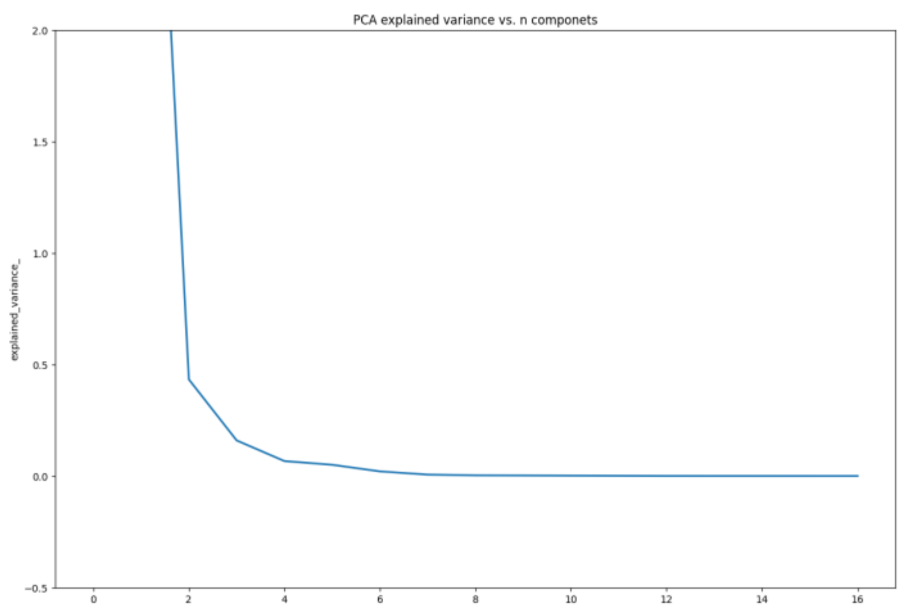

PS6 information--distribution
9.90
9.85
9.80
9.75
9.70
9.65
9.60
0
1000
2000
3000
4000
5000
6000

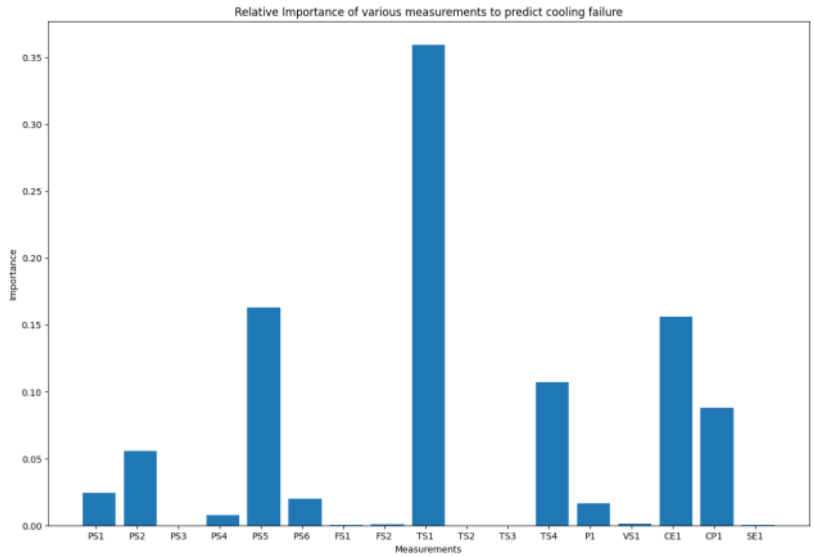

FS1 information--distribution
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
100
200
300
400
500
600

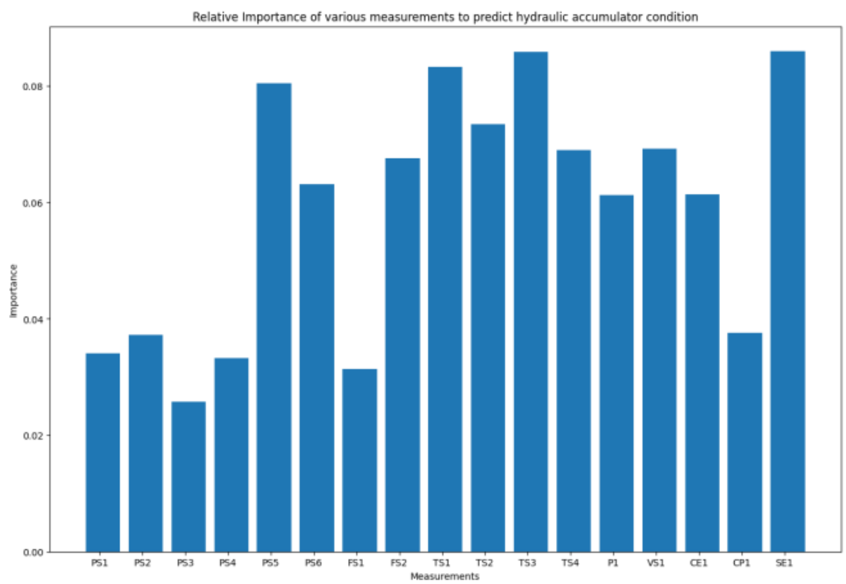

PS1
PS2
PS3
PS4
PS5
PS6
FS1
FS2
TS1
TS2
TS3
TS4
EPS1
VS1
CE
CP
SE
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

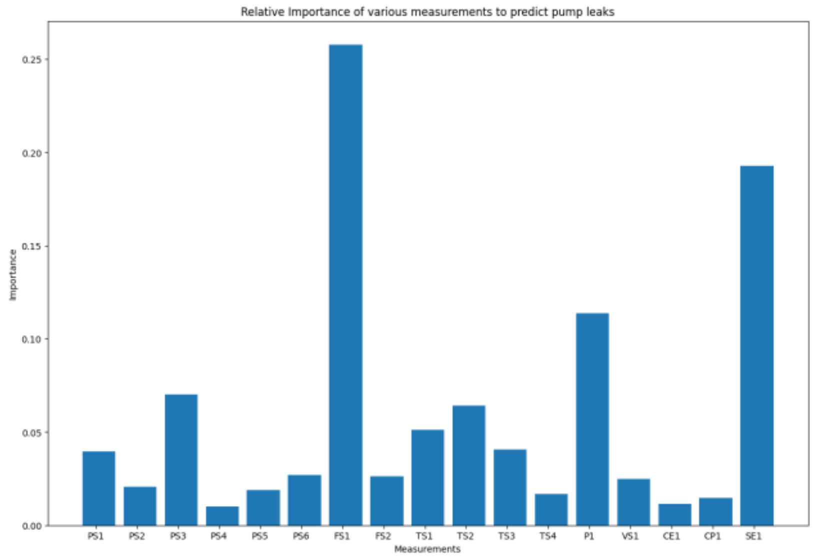

PS1
PS2
PS3
PS4
PS5
PS6
FS1
FS2
TS1
TS2
TS3
TS4
P1
VS1
CE1
CP1
SE1
0.2
0.0
−0.2
−0.4
−0.6
−0.8

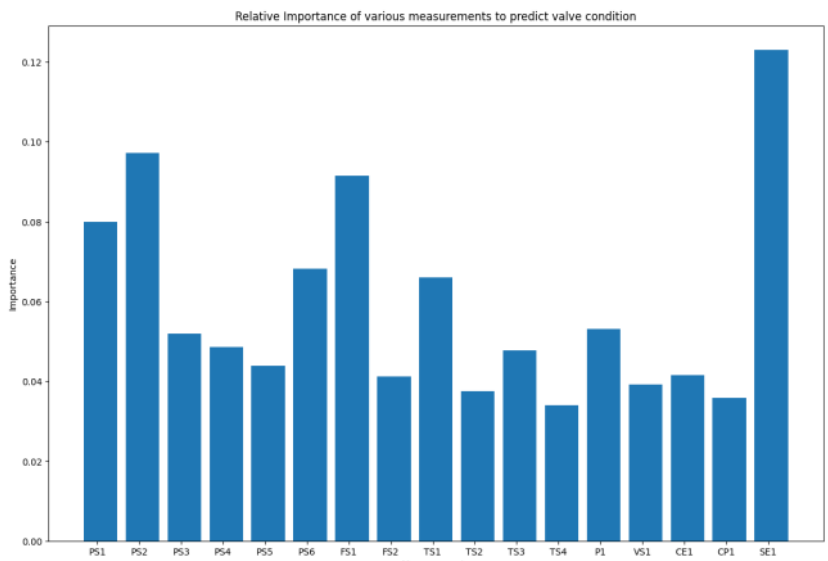

PS1
PS2
PS3
PS4
PS5
PS6
FS1
FS2
TS1
TS2
TS3
TS4
P1
VS1
CE1
CP1
SE1

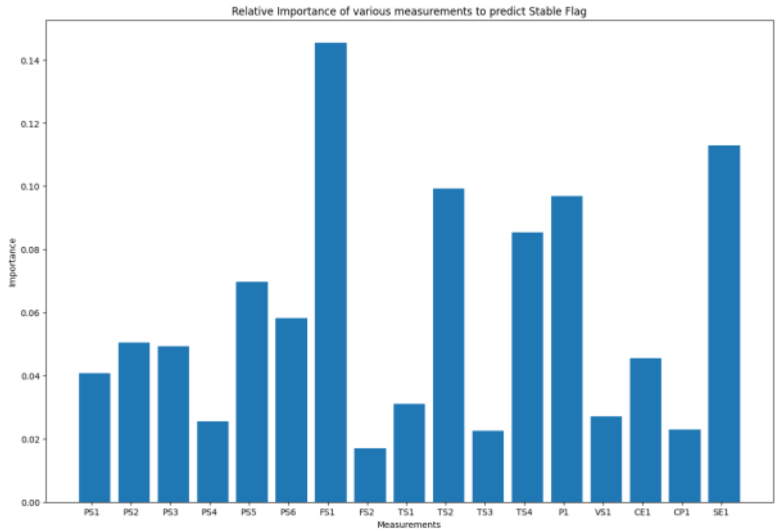

PCA explained variance vs. n componets
explained_variance_
n_components


Relative Importance of various measurements to predict cooling failure
Importance
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
PS1
PS2
PS3
PS4
PS5
PS6
FS1
FS2
TS1
TS2
TS3
TS4
P1
VS1
CE1
CP1
SE1
Measurements


Relative Importance of various measurements to predict hydraulic accumulator condition
Importance
0.00
0.02
0.04
0.06
0.08
PS1
PS2
PS3
PS4
PS5
PS6
FS1
FS2
TS1
TS2
TS3
TS4
P1
VS1
CE1
CP1
SE1
Measurements


Relative Importance of various measurements to predict pump leaks
Importance
Measurements
PS1
PS2
PS3
PS4
PS5
PS6
FS1
FS2
TS1
TS2
TS3
TS4
P1
VS1
CE1
CP1
SE1


Relative Importance of various measurements to predict valve condition
Importance
Measurements
PS1
PS2
PS3
PS4
PS5
PS6
FS1
FS2
TS1
TS2
TS3
TS4
P1
VS1
CE1
CP1
SE1



Deep Neural Network:

```
Cooler Condition accuracy: 99.77%
Valve Condition accuracy: 98.19%
Internal Pump Leakage accuracy: 99.32%
Hydraulic Accumulator accuracy: 80.95%
Stable Flag accuracy: 95.01%
```

Support Vector Machine :

```
cooling failure Accuracy is:  100.00%
valve condition Accuracy is:  93.42%
pump leaks Accuracy is:  98.41%
hydraulic accumulator condition Accuracy is:  78.46%
stable flag Accuracy is:  90.93%
```

XGBoost :

Cooling Failure:

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        82
           1       1.00      1.00      1.00        70
           2       1.00      1.00      1.00        69

    accuracy                           1.00       221
   macro avg       1.00      1.00      1.00       221
weighted avg       1.00      1.00      1.00       221
```

Hydraulic Accumulator:

	precision	recall	f1-score	support
0	1.00	0.92	0.96	78
1	0.92	0.95	0.94	38
2	0.88	1.00	0.94	36
3	1.00	1.00	1.00	69
accuracy			0.96	221
macro avg	0.95	0.97	0.96	221
weighted avg	0.97	0.96	0.96	221

Pump Leakage:

	precision	recall	f1-score	support
0	1.00	0.99	1.00	125
1	0.95	0.98	0.96	54
2	0.98	0.95	0.96	42
accuracy			0.98	221
macro avg	0.97	0.98	0.97	221
weighted avg	0.98	0.98	0.98	221

Valve Condition:

	precision	recall	f1-score	support
0	0.95	0.95	0.95	39
1	0.95	0.98	0.96	41
2	0.89	0.81	0.85	31
3	0.96	0.97	0.96	110
accuracy			0.95	221
macro avg	0.94	0.93	0.93	221
weighted avg	0.94	0.95	0.94	221

Stable Flag:

	precision	recall	f1-score	support
0	0.94	1.00	0.97	147
1	1.00	0.86	0.93	74
accuracy			0.95	221
macro avg	0.97	0.93	0.95	221
weighted avg	0.96	0.95	0.95	221

V. Conclusion

In conclusion, our research focused on the condition mon- itoring of hydraulic systems using three prominent classifica- tion algorithms: deep neural network (DNN), XGBoost, and Support Vector Machine (SVM). Through our implementations and analyses, we've gained valuable insights into the effectiveness of these algorithms in predicting various system conditions.

Firstly, the deep neural network approach, leveraging re- current neural network architecture, demonstrated promising results in accurately classifying multiple target labels. By utilizing sequential patterns and temporal dependencies, the DNN model exhibited robust performance, potentially captur- ing nuanced features within the data.

Secondly, the XGBoost modeling technique showcased notable predictive capabilities, leveraging gradient boosting to construct an ensemble of decision trees. Its ability to handle complex relationships between features contributed to its effectiveness in classifying hydraulic system conditions.

Lastly, the Support Vector Machine algorithm demonstrated its strength in binary classification tasks, effectively sepa- rating classes by constructing optimal hyperplanes in high- dimensional feature spaces. Although SVM's performance may vary depending on the kernel used and dataset characteristics, it remains a powerful tool for classification tasks.

In summary, our comparative analysis of these classifica- tion algorithms provides valuable insights for practitioners and researchers in the field of hydraulic system condition monitoring. Each algorithm presents unique strengths and limitations, and the choice of algorithm should be tailored to specific application requirements and dataset characteristics. Further research could explore ensemble methods or hybrid approaches to leverage the strengths of multiple algorithms for improved predictive performance in real-world hydraulic system monitoring applications.

References

[1] A. M. T M, J. R, S. G, S. R and S. K. D, "Condition Monitoring of Hydraulic Brake System Using Rough Set Theory and FUZZY Rough Nearest Neighbor Learning Algorithms," 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), Rourkela, India, 2019, pp. 229-232

[2] D. Neufeld and U. Schmid, "Anomaly Detection for Hydraulic Systems under Test," 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vasteras, Sweden, 2021, pp. 1-8

[3] Bahman Askari, Raffaele Carli, Graziana Cavone, Mariagrazia Dotoli, "Data-Driven Fault Diagnosis in a Complex Hydraulic System based on Early Classification" ,Volume 55.Issue 40,2022 ISSN 2405-8963

[4] A. D. Bykov, V. I. Voronov and L. I. Voronova, "Machine Learning Methods Applying for Hydraulic System States Classification," 2019 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 2019, pp. 1-4

[5] Khabat Khosravi, Zohreh Sheikh Khozani, James R. Cooper,Predicting stable gravel-bed river hydraulic geometry: A test of novel, advanced, hybrid data mining algorithms,Environmental Modelling Software, Volume 144, 2021.

[6] L. Lee, M. Lees and G. S. Ridenour, "Multi-Interval Data Mining of At-A-Station Hydraulic Geometry to Quantify Temporal Stability," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2019, pp. 0501-0508

[7] Kan, Yingzhe Sun, Dongye Luo, Yong Qin, Datong Shi, Junren Ma, Ke. (2019). Optimal design of the gear ratio of a power reflux hydraulic transmission system based on data mining. Mechanism and Machine Theory.

[8] N. Kunicina, A. Zabasta, M. Juskans, A. Zhiravetska and A. Patlins, "Development of an approach for operational indicators assessment of hydro generator unit based on online sensor system data," 2020 IEEE 61th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Riga, Latvia, 2020, pp. 1-6

[9] Mahdi Majedi-Asl , Mehdi Fuladipanah, (2022). "Performance Assess- ment of Various Data-Mining Methods to Determine Depth Velocity Profile at Submerged Hydraulic Jump", Irrigation Water Engineering, 13(50), 1-20. magiran.com/p2513299

[10] Lin, F.; Zhang, Q.; Yu, P.; Guo, J., “Graded Evaluation of Health Status of Hydraulic System with Variable Operating Conditions Based on Parameter Identification”, Appl. Sci. 2023, 13, 6052.

[11]Wen Zhijie, Pang Yi-Hui,Wang Hong-Bo,Zhao Jian-Jian, “Analysis and Prediction of Hydraulic Support Load Based on Time Series Data Modeling”, 8851475, 2020.

[12] Mengxi Zhang, Mingchao Li, Yang Shen, Qiubing Ren, Jinrui Zhang,Multiple mechanical properties prediction of hydraulic con- crete in the form of combined damming by experimental data min- ing,Construction and Building Materials,Volume 207,2019,ISSN 0950- 0618

[13] S. S. Chawathe, ”Condition Monitoring of Hydraulic Systems by Clas- sifying Sensor Data Streams,” 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0898-0904

[14] Kumar, Prashant and Kumar, Sanjeev and Sharma, Digvijay Narayan and Singh, Hari Darshan and Prakash, Gyan, ” Hydraulic System Anomaly Detection Using Data Analytics”

[15] X. Zhao, K. Zhang and Y. Chai, ”A Multivariate Time Series Classifi- cation based Multiple Fault Diagnosis Method for Hydraulic Systems,” 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 6819-6824

DWDM Project Final Report

ORIGINALITY REPORT

14%	8%	11%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	livrepository.liverpool.ac.uk Internet Source	2%
2	Deniz Neufeld, Ute Schmid. "Anomaly Detection for Hydraulic Systems under Test", 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2021 Publication	1%
3	www.mdpi.com Internet Source	1%
4	Yingzhe Kan, Dongye Sun, Yong Luo, Datong Qin, Junren Shi, Ke Ma. "Optimal design of the gear ratio of a power reflux hydraulic transmission system based on data mining", Mechanism and Machine Theory, 2019 Publication	1%
5	Bahman Askari, Raffaele Carli, Graziana Cavone, Mariagrazia Dotoli. "Data-Driven Fault Diagnosis in a Complex Hydraulic System based on Early Classification", IFAC-PapersOnLine, 2022	1%

Publication

6	www.sciencegate.app Internet Source	<1 %
7	Kyutae Kim, Jongpil Jeong. "Deep Learning-based Data Augmentation for Hydraulic Condition Monitoring System", Procedia Computer Science, 2020 Publication	<1 %
8	A. D. Bykov, V. I. Voronov, L. I. Voronova. "Machine Learning Methods Applying for Hydraulic System States Classification", 2019 Systems of Signals Generating and Processing in the Field of on Board Communications, 2019 Publication	<1 %
9	Wei Yan, Yu Shi, Zengyan Ji, Yuan Sui, Zhenzhen Tian, Wanjing Wang, Qiushi Cao. "Intelligent predictive maintenance of hydraulic systems based on virtual knowledge graph", Engineering Applications of Artificial Intelligence, 2023 Publication	<1 %
10	Aurelien Teguede Keleko, Bernard Kamsu-Foguem, Raymond Houe Ngouna, Amèvi Tongne. "Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP explainable XAI", Advances in Engineering Software, 2023	<1 %

Publication

11	www.ijrte.org Internet Source	<1 %
12	Kyutae Kim, Jongpil Jeong. "Multi-layer Stacking Ensemble for Fault Detection Classification in Hydraulic System", 2022 26th International Conference on Circuits, Systems, Communications and Computers (CSCC), 2022 Publication	<1 %
13	Alamelu Manghai T M, Jegadeeshwaran R, Sakthivel G, Sivakumar R, Saravana Kumar D. "Condition Monitoring of Hydraulic Brake System Using Rough Set Theory and FUZZY Rough Nearest Neighbor Learning Algorithms", 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), 2019 Publication	<1 %
14	Khuat, Thanh Tung. "Development of Robust and Scalable Hyperbox Based Machine Learning Algorithms", University of Technology Sydney (Australia), 2023 Publication	<1 %
15	www.informit.com Internet Source	<1 %

16	Fengqin Lin, Qingxiang Zhang, Peng Yu, Jin Guo. "Graded Evaluation of Health Status of Hydraulic System with Variable Operating Conditions Based on Parameter Identification", Applied Sciences, 2023 Publication	<1%
17	ijcs.stmikindonesia.ac.id Internet Source	<1%
18	K. Rajesh Rao, Aditya Kolpe, Tribikram Pradhan, Bruno Bogaz Zarpelão. "An Efficient Solution to User Authorization Query Problem in RBAC Systems Using Hierarchical Clustering", Journal of Cyber Security and Mobility, 2022 Publication	<1%
19	Xiangkai Ma, Pei Wang, Bozhou Zhang, Ming Sun. "An improved Multi-channels Information Fusion Model with multi-scale signals for fault diagnosis", 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021 Publication	<1%
20	link.springer.com Internet Source	<1%
21	Shuaijie Mei, Mei Yuan, Jin Cui, Shaopeng Dong, Juanru Zhao. "Machinery condition	<1%

monitoring in the era of industry 4.0: A relative degree of contribution feature selection and deep residual network combined approach", Computers & Industrial Engineering, 2022
Publication

22	bmcbioinformatics.biomedcentral.com Internet Source	<1 %
23	e-pub.uni-weimar.de Internet Source	<1 %
24	research-api.cbs.dk Internet Source	<1 %
25	staffsites.sohag-univ.edu.eg Internet Source	<1 %
26	assets.researchsquare.com Internet Source	<1 %
27	eut.openaire.eu Internet Source	<1 %
28	repositorio.uneatlantico.es Internet Source	<1 %
29	www.flashcommerce.com Internet Source	<1 %
30	www2.mdpi.com Internet Source	<1 %

31	Shyamala Rajasekar. "Prediction of Component Level Degradation in a Hydraulic Rig using Machine Learning Methods", 2023 15th International Conference on Developments in eSystems Engineering (DeSE), 2023 Publication	<1 %
32	oatao.univ-toulouse.fr Internet Source	<1 %
33	onlineengineeringeducation.com Internet Source	<1 %
34	researchid.co Internet Source	<1 %
35	www.jatit.org Internet Source	<1 %
36	www.researchgate.net Internet Source	<1 %
37	H. L. Gururaj, V. Janhavi, U. Tanuja, Fransesco Flamini, B. C. Soundarya, Vinayakumar Ravi. "Predicting Traffic Accidents and their Injury Severities Using Machine Learning Techniques", International Journal of Transport Development and Integration, 2022 Publication	<1 %
38	Khabat Khosravi, Zohreh Sheikh Khozani, James R. Cooper. "Predicting stable gravel-bed	<1 %

river hydraulic geometry: A test of novel, advanced, hybrid data mining algorithms", Environmental Modelling & Software, 2021
Publication

39	Yingzhe Kan, Dongye Sun, Datong Qin, Yong Luo, Junren Shi, Ke Ma. "Design method of power matching for reflux hydrodynamic transmission system", Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2021 Publication	<1%

Exclude quotes	On	Exclude matches	< 3 words
Exclude bibliography	On		