

Sarat

Shankar_ResearchProposal_Fin alReport_Compressor- Predictive Maintanance_V01- 1.docx

by Sarat Shankar

Submission date: 14-Oct-2025 06:43PM (UTC+0100)

Submission ID: 265496364

File name: Sarat_Shankar_ResearchProposal_FinalReport_Compressor-Predictive_Maintanance_V01.docx
(1.98M)

Word count: 14797

Character count: 86798

PREDICTIVE MAINTANANCE & ANOMALY DETECTION FOR
SINGLE STAGE SCREW AIR COMPRESSORS

SARAT SHANKAR

3
Final Thesis Report

OCTOBER 2025

• TABLE OF CONTENT

□ <i>DEDICATION</i>	4
□ <i>ACKNOWLEDGEMENTS</i>	4
□ <i>ABSTRACT</i>	5
□ <i>LIST OF TABLES</i>	6
□ <i>LIST OF FIGURES</i>	6
□ <i>LIST OF ABBREVIATIONS</i>	6
□ <i>CHAPTER1 - INTRODUCTION</i>	8
<i>1.1. Background of the Study</i>	8
<i>1.2. Problem Statement</i>	13
<i>1.3. Aim and Objectives</i>	14
<i>1.4. Research Questions</i>	14
<i>1.5. Scope of the Study</i>	15
<i>1.6. Significance of the Study</i>	15
<i>1.7. Structure of the Study</i>	16
□ <i>CHAPTER2 - LITERATURE REVIEW</i>	17
<i>1.8. Introduction</i>	17
<i>1.9. History of Industrial Maintenance</i>	17
<i>1.10. Types of Maintenance</i>	20
<i>1.11. Predictive Maintenance & Anomaly Detection</i>	20
<i>1.12. Sensor-based Data Acquisition</i>	22
<i>1.13. Technology use in Predictive Maintenance</i>	24
<i>1.14. Related Research Publications:</i>	25
<i>1.15. Discussion & Summary:</i>	30
□ <i>CHAPTER 3: RESEARCH METHODOLOGY</i>	31
<i>1.16. Introduction</i>	31
<i>1.17. Research Methodology</i>	31
<i>1.17.1. Introduction</i>	31
<i>1.17.2. Architecture for Proposed System</i>	31
<i>1.17.3. Data Preparation</i>	34
<i>1.17.3.1. Dataset Selection</i>	34
<i>1.17.3.2. Dataset Preparation</i>	34
<i>1.17.3.3. Dataset Usage</i>	38

<i>1.17.4. Proposed Method</i>	39
<i>1.17.4.1. Algorithms & Techniques</i>	39
<i>1.17.4.1.1. Adaptive Random Forest</i>	40
<i>1.17.4.1.2. LSTM</i>	40
<i>1.17.4.1.3. NiceGUI</i>	41
<i>1.17.4.2. Proposed Research System</i>	41
<i>1.17.4.2.1. Stage1</i> :.....	41
<i>1.17.4.2.2. Stage 2</i>	42
<i>1.17.5. Evaluation Method</i>	43
<i>1.17.6. Visualization & Monitoring</i>	46
<i>1.17.6.1. Tools & Technology</i>	46
<i>1.17.6.2. Anomaly Detection</i>	46
<i>1.17.6.3. Condition Monitoring</i>	47
<i>1.17.6.4. Timeseries Forecasting</i>	47
<i>1.17.6.5. System Simulator</i>	48
□ <i>CHAPTER 4: RESEARCH ANALYSIS</i>	50
<i>1.17.7. Introduction</i>	50
<i>1.17.8. Dataset Description</i>	50
<i>1.17.9. Dataset Processing</i>	50
<i>1.17.9.1. Elimination of Variables</i>	50
<i>1.17.9.2. Data Transformation</i>	51
<i>1.17.9.3. Feature Selection:</i>	51
<i>1.17.9.4. Data Scaling</i> :.....	53
<i>1.17.9.5. Splitting of Original Dataset</i>	53
<i>1.17.9.6. Exploratory Data Analysis</i>	54
<i>1.17.9.6.1. Univariate Analysis</i>	54
<i>1.17.9.6.2. Correlation analysis</i>	54
□ <i>CHAPTER 5: RESULTS & DISCUSSIONS</i>	57
<i>1.17.10. Introduction</i>	57
<i>1.17.11. Evaluation Results</i>	57
<i>1.17.12. Interpretation of Visualizations</i>	60
<i>1.17.13. Summary</i>	61
□ <i>CHAPTER 6: CONCLUSIONS & RECOMMENDATIONS</i>	62

<i>1.17.14. Introduction</i>	62
<i>1.17.15. Discussion and Conclusion</i>	62
<i>1.17.16. Future Recommendation</i>	62
□ <i>CHAPTER 7: RESOURCE REQUIREMENTS</i>	63
<i>1.17.17. Hardware Requirements</i>	63
<i>1.17.18. Software Requirements</i>	63
□ <i>CHAPTER 8: DATA MANAGEMENT PLAN</i>	64
□ <i>REFRENCES</i>	65
□ <i>APPENDIX A : RESEARCH PLAN</i>	69
<i>1.17.19. Introduction</i>	69
<i>1.17.20. Research Project Plan</i>	69
<i>1.17.21. Risk Mitigation and Contingency Plan</i>	69
□ <i>APPENDIX B: RESEARCH PROPOSAL</i>	71
□ <i>APPENDIX B: ETHICS FORMS</i>	72

• **DEDICATION**

I would like to thank my family for their continuous moral and emotional support in this journey. I also appreciate all the support I received from the other students in my batch whose enthusiasm motivated me to finish this thesis.

• **ACKNOWLEDGEMENTS**

I would like to thank all the guidance, help and advice provided by my thesis supervisor Mr Nirav Bhatt in this journey which helped me to better understand key points of this thesis. I would like to thank the LJMU and Upgrad for the studentship that allowed me to conduct this thesis.

• ABSTRACT

Predictive Maintenance using ML algorithms has become a very critical operation in industry where vast amounts of data generated by different equipments are analyzed to predict potential failures. By leveraging historical and real-time data, along with the power of Machine Learning enable maintenance activities to be precisely timed, ensuring equipment is serviced only, when necessary, which reduces overall cost of maintenance, improves useful life of equipment and also improves efficiency of maintenance engineers.

This research paper presents a review on Predictive Maintenance and Anomaly Detection of an Industrial Single Stage Screw air compressor using ML algorithms where we analyze key historical operational parameters and monitor real-time equipment data to detect anomalies for Single Stage Air Compressor. The study focuses on having an end-to-end solution that can be scaled and implemented across the shop floor. Predictive Maintenance of Air Compressor prevents unplanned downtime, prevents costly capital expenditure, and improves the overall equipment efficiency.

Air compressors are important equipment for a variety of industries which use kinetic energy to create pressurized air. This air which is pressurized and stored in a storage air tank is used to power various industrial tools and processes.

Currently, in majority of manufacturing industries, maintenance adopted is based on periodic and corrective maintenance policies. There are many research papers written on predictive maintenance of industrial equipment, but actual fruitful implementation of these on the shop floor is very limited.

This research focuses on detecting anomalies associated with key components of air compressors. In this study, we focus on a single component and examine the key parameters influencing its behaviour. To achieve this, we employ Adaptive Random Forest models to analyse real-time data generated by air compressors during production and predict any probable anomaly based on datapoints that deviate significantly from expected patterns. Also, we are assessing forecasting approaches for identified key parameters like temperature, airflow etc. and we are also exploring what a minimalistic simple, user-friendly dashboard or screen should be available with maintenance engineers that can be scaled and implemented across the shop floor

The results of this research will help us prove that advanced machine learning techniques can be applied to single stage screw air compressors and will support in defining policies for predictive maintenance replacing the currently used periodic and corrective maintenance policies.

Additionally, the conclusion of this research can scale to other many other equipment and instruments in an industrial ecosystem.

• LIST OF TABLES

Table 1 - List of Abbreviations.....	7
Table 2 - Evaluation of Maintenance Strategies - (Michael Nsor, 2024).....	19
Table 3 - Correlation -Industrial Revolution & Maintenance (Peter Poor, Davide Zenesik and Josef Basl, 2019)	20
.....	20
Table 4 - Research Publication Table	29
Table 5 - List of Dataset Parameters	38
Table 6 - Labelled Fault List	38
Table 7 - Architecture diagram on Proposed Research Structure Stage1.....	41
Table 8 - Architecture diagram on Proposed Research Structure Stage2.....	42
Table 9 - Confusion Matrix	44
Table 10 - Visualization – Simulator (VIsualization).....	48
Table 11 - Removed Dataset Features	50
Table 12 - Selected Feature List	53
Table 13 - Dataset Split %	54
Table 18 - Data Management Plan.....	64
Table 19 - Risks & Contingency.....	70

• LIST OF FIGURES

Figure 1 - Single Stage Screw Air Compressor Working Diagram Nikhil Technochem (2024)	9
Figure 2 - Single Stage Screw Air Compressor Block Diagram (Engineer Waqar, 2020).....	10
Figure 3 - Architecture Diagram for proposed system	31
Figure 4 - ROC Curve Plot.....	45
Figure 5 - Anomaly Detection Plot (Visualization).....	46
Figure 6 - Anomaly Detection Details (Visualization)	47
Figure 7 - Condition Monitoring Anomaly Detection Details (Visualization)	47
Figure 8 - Timeseries Forecasting.....	48
Figure 9 - Feature Missing Values	51
Figure 10 - Timestamp Assignment	51
Figure 11 - Feature Importance	52
Figure 12 - Stage1 - Evaluation Matrix	52
Figure 13 - Stage 1 - Data Scaling	53
Figure 14 - Bearing Fault Distribution Chart	54
Figure 15 - Correlation Natrix Heatmap	55
Figure 16 - Corelation Chart - Bearing Fault.....	56
Figure 17 - Stage 1 - Evaluation Matrix	57
Figure 18 - Stage 1 - Precision, Recall & F1 Plot	58
Figure 19 - Stage2 - Evaluation Metrix Batch1	59
Figure 20 - Stage2 - Evaluation Metrix Batch2	59
Figure 21 - Stage2 - Evaluation Metrix Batch3	59
Figure 22- Stage2 - Evaluation Metrix Batch4	60
Figure 23 - Project Plan Gant Chart.....	69

• LIST OF ABBREVIATIONS

Below table (Table1) shows the list of abbreviations used in this proposal.

Abbreviations	Description
---------------	-------------

ML	Machine Learning
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
GPU	Graphics Processing Units
KPI	Key Performance Indicator
RF	Random Forest
RPM	Rotation Per Minute
KPI	Key Performance indicators
RM	Reactive Maintenance
PM	Preventive Maintenance
PdM	Predictive Maintenance
AI	Artificial Intelligence
IIOT	Industrial Internet of Things
IDE	Integrated Development Environment
EDA	Exploratory Data Analysis
LSTM	Long Short term Memory
RUL	Remaining Useful Life
CUDA	Compute Unified Device Architecture

Table 1 - List of Abbreviations

- CHAPTER1 - INTRODUCTION

1.1. Background of the Study

Predictive Maintenance is a preventative approach that forecasts the probability of industrial equipment failure using data and algorithms. ML models can predict possible problems before they arise by examining patterns in data gathered from sensors and other sources. Understanding Operational Efficiency and the reliability of all industrial systems in a manufacturing domain are very important to minimize downtime and improve efficiency and quality of production.

During **first Industrial Revolution** ¹ the most characteristic maintenance form during this period was the breakdown maintenance (also known as “**reactive maintenance**” or “**corrective maintenance**”) where repairs are done only after the breakdown

With the peak of the **Industrial Revolution 2.0**. Machines became more complex, and production grew rapidly. Breakdowns caused higher expenses and therefore first attempts at **preventive maintenance** (also known as planned maintenance) appeared

With **Industrial Revolution 3.0 Productive Maintenance** (also referred to as PM) ¹ started appearing after the second world war. This new approach towards maintenance **combines Corrective Maintenance and Preventive Maintenance** with a data-driven, analytical approach and is performed to increase the broad economic efficiency of production. Methods like TPM and RCM were developed during this period.

In **Industrial Revolution 4.0** circa 2010 we started seeing use of Predictive maintenance in industries, and with **Industrial Revolution 5.0** advancement continues.

Research and understanding of these critical parameters and the relation of these parameters with overall efficiency and health of equipment is the key in predicting the performance of any equipment, and with this will help maintenance engineers to maintain the equipment and improve RUL (Remaining Useful Life).

This research will focus on predictive maintenance and anomaly detection of an industrial Single stage screw compressor with the use ML algorithms.

What is Industrial Single Stage Screw Compressor?

As referred in (Shivansh Sabhadiya,2021), Air compressor is an important equipment for a variety of industries which uses kinetic energy to create pressurized air. This air which is pressurized and stored in a storage air tank is used to power various industrial tools and processes.

In this research we will be focusing on Industrial Single Stage Screw Compressor. A screw compressor works on positive displacement mechanism. There are 2 Spiral rotors in the systems, one is called as male rotor, and another is called as female rotor. The male rotors are powered by an electric motor and both these rotors mesh and cause generation of air.

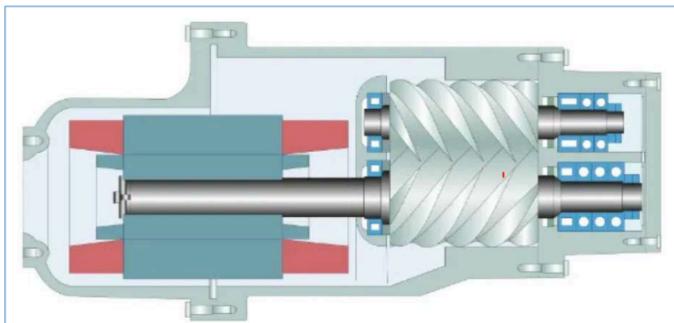


Figure 1 - Single Stage Screw Air Compressor Working Diagram Nikhil Technochem (2024)

Above figure (Figure 1) shows the working principal of screw air compressor (figure taken for reference from cited as: Nikhil Technochem (2024)

Schematic Block Diagram of Screw Air Compressor Unit

Below figure (Figure 3) shows Schematic block diagram of Screw Air Compressor unit figure taken for reference from cited as: Mechstudies (2021)

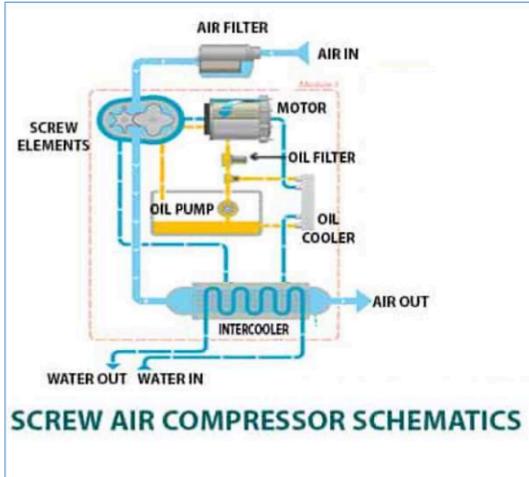


Figure 2 - Single Stage Screw Air Compressor Block Diagram (Engineer Waqar, 2020)

Screw Air Compressor consists of following Major parts which direct quotation from website (Engineer Waqar, 2020)

- **Air In**

Air is taken into the air compressor from this Inlet. Air taken in from the outside environment is passed to the Inlet Air filter

- **Air Filter @ Inlet**

Ambient air from environment is not clean, it contains pollutants, dust and mist and it is required to be removed from the air coming from inlet. Filters are available at air inlet to clean the air. These filters need to be cleaned frequently to maintain efficiency of air compressor.

- **Screw Element**

Air inlet in air compressor compresses the air. There are 2 rotors/screws in Screw air compressor, large screw is called female rotor and the other one is known as male rotor. The male rotors are powered by an electric motor and both these rotors mesh and cause generation of air.

Rotors need frequent oil treatment for preventing or decreasing friction within parts

- **Motor**

Electric Motor is the source of power source for air compressor; this is used to power and rotate the male and female screw elements.

- **Oil Filters**
- In air compressor one of the important operations is lubrication, there are multiple types of lubricant used in air compressor and one of the most famous is Oil based lubrication, this is required as there are many rotating parts in compressor, to filter the lubrication oil – oil filters are used.
- **Oil Pump & Oil Coolers**
Oil pumps are used to pump oil to screw element as shown in figure(Figure 3) and coolers are used to cool down oil coming from screw element.
- **Bearings**

Bearing is another mechanical element in air compressor, these are anti friction rollers which are used to:

- Bearings are used rotating parts like male and female rotor.
- Bearing reduces friction between moving parts.
- With reduced friction it helps reduce the losses caused by friction.
- It helps free rotation of Screw in air compressor.

What are the major causes of failure in Industrial Single Stage Screw Air Compressor?

While there can be a lot of reasons related to mechanical, electrical and others

Bearing Fault:

Bearings, though small, play a crucial role in the operation of any air compressor. Inadequate lubrication can cause the motor to overheat, deteriorate, and potentially fail ahead of its expected lifespan. Regularly maintaining clean, well-lubricated bearings is a simple yet effective way to ensure smooth and reliable air compressor performance. In this study we are focussing on Bearing fault.

Water Pump Fault:

A water pump is a mechanical device engineered to circulate water or coolant throughout a system, ensuring efficient flow and operational functionality. Common faults in water pumps may arise from wear and tear, overheating, corrosion, rust, contaminated fluids, or failures in electrical components.

Radiator Fault:

The air compressor radiator plays a role in cooling the air compressor oil and is a very important component of the screw air compressor. Radiator fault can occur due to oil blockage, dust blocking, carbon blockage etc.

Motor Fault:

Electric Motor is used in Fcompressor to drive piston and is the most important part of air compressor. There are a lot of reasons for motor to stop working ranging from electrical to mechanical to environmental issues.

Expansion Valves:

Air Compressor valves are essential for managing the flow of air within the system. They operate in coordination with the piston strokes, precisely opening and closing to regulate the intake and discharge. Common issues in fault includes corrosion, insufficient lubrication, heat, wear & tear etc.

Major issues faced in the field of Predictive maintenance of compressors are the availability of data from air compressors, in majority of the cases there are not enough sensors available with air compressors to provide key parameters, another issue is lack of labeled data where there is not adequate labels available for past failures, this makes training accurate models very difficult.

Most of the studies in this area of  Predictive Maintenance systems operate on historical data but fail to implement real-time monitoring and anomaly detection mechanisms and provide a minimum required user interface to visualize & monitor critical parameters & anomalies.

To alleviate this gap, this research paper focuses on detecting anomalies linked to critical air compressor components, and this study concentrates on a single component to analyse the key parameters influencing its performance. **Adaptive Random Forest** models are utilized to process real-time production data from air compressors, enabling the prediction of potential anomalies by identifying data points that significantly deviate from expected patterns. Additionally, we evaluate forecasting techniques for essential parameters such as temperature and airflow. Parallel to this, we explore the design of a minimalistic, user-friendly dashboard tailored for maintenance engineers - one that is scalable and deployable across the shop floor.

1.2. Problem Statement

Unplanned Downtime is a big issue for any manufacturing industry; there is a huge cost burden for each unplanned downtime. Direct Quotation from Doug White, Emerson Industry Expert – says that based on current economics of any refinery, every 1% gain in availability is worth \$84 million of additional margin capture per year in a typical 200,000 bpd refinery. Also, recent studies show that in a typical manufacturing industry, the downtime of machines due to unplanned reasons cost the organization an estimated \$60 billion each year.

With time, we saw a revolution in maintenance of Industrial Systems - reflects a transition or revolution of maintenance strategies from Reactive Maintenance to Preventive Maintenance to Predictive Maintenance

Adoption of Predictive Maintenance has become a necessary objective for all manufacturing industries to stay comparative as unexpected equipment failure causes expensive repairs, costly downtime, loss of production due to extended downtime events and increased need of skilled resources,

Most of the studies in this area of Predictive Maintenance for Compressors ⁴ systems operate on historical data but fail to implement real-time monitoring and anomaly detection mechanisms which also caters to live streaming data coming from compressors while effectively adapting to concept drift scenarios.

Also, there are limited studies available which focuses on easy to implement User Interface to do condition Monitoring , forecasting for key parameters and anomaly detection which can be used by medium and small-scale industries.

Also, there is very little research available that focuses on cost effectiveness, which is very important in the real adoption of such systems in Industrial floors.

With advancement in technology, solving this problem with right set of approaches and technologies in key for Industry 5.0 goals.

1.3. Aim and Objectives

Aim: of this research is to develop a system which can detect anomalies in **Single Stage** Screw Air Compressor while analyzing live data coming from sensors as live streaming data, additionally aim is to implement condition monitoring and forecasting of critical operational parameters using simple to use UI framework

of this research is to forecast critical parameters coming from Sensors mounted on a Single Stage Screw Air Compressor and detect anomalies using advanced machine learning algorithm using simple to use UI framework

Objectives:

- To conduct a comprehensive review of available literature regarding Anomaly Detection and Predictive Maintenance in Industrial Air Compressors.
- To forecast critical features of Industrial Single stage Screw AirCompressor.
- To detect anomalies using a machine learning algorithm by monitoring data in real-time.
- To provide an optimized value-driven driven User-Interface to maintenance engineers for real-time monitoring of parameters and detected anomalies.

1.4. Research Questions

This thesis tries to answer the following questions:

- Can an Adaptive Random Forest be effective in predicting detecting Anomalies for a Single Stage Screw Air Compressor
- How can model help with real-time monitoring of Air Compressor sensor inputs and detect anomalies (based on threshold values)? This will help the maintenance team take immediate action.
- How and what information related to condition monitoring and anomaly detection can be visually provided to the Maintenance team? so that they can root cause the issue relatively faster with minimum skills required.
- How can a simple time series forecasting of key parameters be helpful in fault detection
- Can the research conducted on the Air Compressor be easily scaled to other industrial equipment on a plant floor (theoretical analysis)?
- How an optimized value-driven driven User-Interface can be useful for Maintenance Engineers.
- Based on Literature review and attained results from the study, what can be further improvements made to the studied approach.

1.5. Scope of the Study

The scope of this research is defined as follows:

- This research work should be completed as per the research plan refer: [Research Plan](#).
- This study is focused on Single Stage Screw Air Compressors which are used in Manufacturing domain.
- Open-Source models and software will be used to conduct all types of experiments. We are using Adaptive random Forest Classifier from River
- We will be using Google-Colab which is a publically available GPU for all experimentations.
- We will be only using recognized available standard metrics for evaluation of models.
- To keep the scope of study limited we will be studying the parameter pattern and anomaly detection capability using Machine Learning for one component of air compressor that is – Bearing.
- UI visualization planned for this research will be very high level and will not cover all parameters for raw data, for predicted data and for anomaly detection.
- For UI Visualization we are using NiceGUI which is python based easy to use UI framework
- This research will be using open-source dataset (refer next section) and may not be having all features which is generally available from air compressor.

1.6. Significance of the Study

In this study I am trying to address issue of unplanned downtime for industrial single stage air screw air compressor. There is a huge cost burden for manufacturers with each unplanned downtime. As I mentioned in the related research section, Doug White, Emerson Industry Expert – Based on current refinery economics, says Every 1% gain in availability is worth \$84 million of additional margin capture per year in a typical 200,000 bpd refinery. Also, recent studies show that in a typical manufacturing industry, downtime of machines due to unplanned reasons cost the organization an estimated \$60 billion each year.

This study will help Maintenance Engineers in any Manufacturing Industry to visualize forecasted values for key parameters and detect anomalies in real time, which will be used by Maintenance engineers to take appropriate actions and avoid unplanned downtime.

Output of this research can be used by any manufacturing industry to monitor and visualize detected anomalies, parameter prediction and condition monitoring parameters for any industrial single stage air screw air compressor

1.7. Structure of the Study

The remainder of the paper is structures in this way:

CHAPTER1 INTRODUCTION: Provides brief background of the study with problem statement. Aims and Objectives for the study is detailed with research questions, scope and significance of the study.

CHANPTER2 LITERATURE REVIEW: This chapter mentions related Literature Reviews and research work carried out around Predictive Maintenance and Anomaly Detection.

CHAPTER 3 RESEARCH METHODOLOGY: This section covers proposed structure and research methodology in detail, where description of dataset is provided, data processing techniques are detailed, model consideration and development are explained and data validation and performance measures are described,

CHAPTER 4 ANALYSIS: This chapter a detailed Analysis of Dataset is provided including Data Preparation, Exploratory Datata Analysis and Visualization.

CHAPTER 5 RESULTS & DISCUSSIONS: This chapter summarizes the results of Study with detailed representation of test results.

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS: This chapter presents the conclusions drawn from the study and outlines key recommendations to guide future research in this domain.

CHAPTER 7 REQUIRED RESOURCES: This section details resource utilized for this research including hardware, software, datasets and others.

CHAPTER 8 DATA MANAGEMENT PLAN: This section covers Data Management Plan and describes how data required for this research are collected and stored in most ethical manner.

REFERENCES: This section provides citations and references

Appendix A: This section details research plan which was put together at the start of research with actual completion dates

- CHAPTER2 - LITERATURE REVIEW

1.8. Introduction

Predictive maintenance is not a new field and there is a lot of work going on in this field from many years, but the technological advancements in recent times has boasted this area with very high potential and unexplored opportunities where systems used machine learning (ML) algorithms and condition monitoring tools to predict potential failures and predict deteriorating health of an asset or equipment. Timely handling of these causes, which can lead to unplanned downtime, is a game-changer in equipment maintenance.

1.9. History of Industrial Maintenance

As mentioned by authors (Peter Poor, Davide Zenesik and Josef Basl, 2019) Each industrial revolution has brought significant shifts in technological, socioeconomic, and cultural dimensions. Among the technological advancements, maintenance management has evolved notably, with approaches to equipment upkeep transforming across successive eras

As mentioned in the background study with Industrial 1.0 circa 1760–1840, machines are attended when it is down this type of maintenance is known as “reactive maintenance” or “corrective maintenance”) where repairs are done only after the breakdown

Authors (Peter Poor, Davide Zenesik and Josef Basl, 2019) that the earliest maintenance strategy adopted by humankind was the reactive approach—allowing equipment to operate until failure occurred. At first glance, this method appeared to be the simplest and most intuitive, especially given the basic design of early machines, which required little to no specialized knowledge for repairs. Even today, reactive maintenance remains the most practiced method, accounting for over 55% of maintenance activities. However, with the growing complexity of machinery—particularly following the onset of the First Industrial Revolution—a shift began to emerge across industries toward more proactive and systematic maintenance strategies.

During the height of Industry 2.0, roughly between 1870 and 1914, machinery became increasingly sophisticated, and production volumes surged. As equipment failures led to rising costs, industries began adopting early forms of preventive maintenance—also referred to as planned maintenance—to reduce downtime and manage expenses more effectively.

Factories progressively substituted human labour with machines, driving increased automation across industrial processes.

With Industry 3.0 circa 1969–2000s Productive Maintenance (also referred to as PM) started appearing after the second world war. This new approach towards maintenance combines Corrective Maintenance and Preventive Maintenance with a data-driven, analytical approach and is performed to increase the broad economic efficiency of production. Methods like TPM and RCM were developed during this period.

As mentioned in (FTMaintanance,2019) the advent of electronics in the latter half of the 20th century ushered in a transformative era of industrial automation. Production processes became increasingly automated through the integration of programmable logic controllers (PLCs) and robotic systems. As equipment performance improved, employee safety emerged as a critical maintenance concern due to heightened risks of accidents. Large enterprises began utilizing punch card-based computerized maintenance management systems (CMMS) to prompt technicians with routine maintenance tasks. Subsequently, technicians documented their work on paper forms, which were then processed by data-entry clerks who input the information into mainframe computers to monitor maintenance activities for each asset.

In Industry 4.0 circa 2010-2020 we started seeing use of Predictive maintenance in industries.

Authors (Peter Poor, Davide Zenesik and Josef Basl, 2019) in their research mentions that a well-functioning predictive maintenance program can mean savings of 8% to 12%. Depending on the equipment and material conditions, it is possible to save 30% to 40% (US Department of energy,2010).

The following savings resulting from the use of predictive maintenance are:

- Return on investment: 10 times
- Reduction of maintenance costs: 25% to 30%
- Troubleshooting: 70% to 75%
- Reduction of downtime: 35% to 45%
- Increased production: 20% to 25% (Bloch, H. P., and Geitner, F. K. 1997).

But the start of predictive maintenance is not cheap. A large part of the equipment requires a cost of more than € 30,000. Staff training also requires additional funding.

As author(Michael Nsor, 2024) says advent of digital technologies and the rapid expansion of the Industrial Internet of Things (IIoT) have revolutionized the implementation and scalability of maintenance strategies. IIoT integrates physical equipment, embedded sensors, data processors, and cloud-based platforms into a cohesive and intelligent network. This interconnected system enables continuous, real-time tracking of key parameters such as temperature, vibration, pressure, and voltage,

providing valuable insights into the operational health of machinery. This also makes predictive maintenance scalable across borders

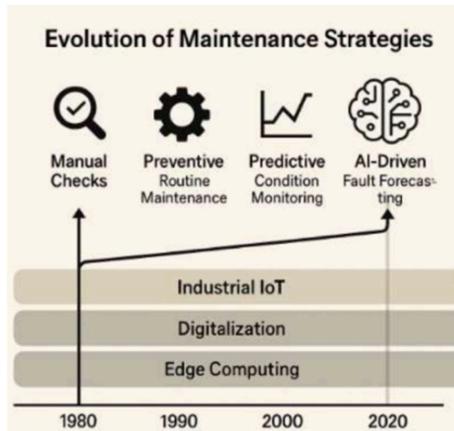


Table 2 - Evaluation of Maintenance Strategies - (Michael Nsor, 2024)

with Industry 5.0 advancement continues. Modern maintenance management focuses on minimizing both unscheduled and scheduled downtime to enhance operational efficiency.

Industry 5.0 places renewed emphasis on the human element, shifting focus from the automation-centric approach of previous industrial eras to the unique value of human insight and creativity. In the context of maintenance, this evolution means that decision-making is not solely driven by data but also informed by human experience and intuition. This human-centric perspective fosters maintenance strategies that are both efficient and attentive to the safety and well-being of workers.

Authors (João Barataa and Ina Kayserb, 2023) concluded with clear correlation between advancement in maintenance activities with Industrial revolutions

Industry revolution	Industry 1.0	Industry 2.0	Industry 3.0	Industry 4.0	Industry 5.0
Characteristics of the industrial revolution	Mechanization, steam power, weaving loom	Mass production, assembly lines,	Automation, computers, electronics	Cyber Physical Systems, IoT,	Cyber Physical Systems, IoT,

		electrical energy		networks, cloud,	networks, cloud,
Type of maintenance	Reactive maintenance	Planned maintenance	Productive maintenance	Predictive maintenance	Predictive maintenance
Inspection	Visual inspection	Instrumental inspection	Sensor monitoring	Predictive analysis	Predictive analysis
OEE	<50%	50-75%	75-90%	>90%	>90%
Maintenance team reinforcement	Trained craftsmen	Inspectors	Reliability engineers	Data scientists	Data scientists

Table 3 - Correlation -Industrial Revolution & Maintenance (Peter Poor, Davide Zenesik and Josef Basl, 2019)

1.10. Types of Maintenance

Reactive or Corrective maintenance – Whenever there is any unplanned downtime due to any issue of it is reported that machine is not running in ideal condition, corrective measures are taken to return equipment to a defined state. Here we can expect unexpected equipment failures, Expensive repairs, Costly unplanned downtime and Lost production during extended downtime events

Preventive maintenance – All maintenance activities are carried out as per a scheduled periodically, there will be a checklist of items to be completed to keep the machine health and efficient. Here we can expect Unnecessary ,frequent planned downtimes, Increased need for maintenance resources and Expensive asset maintenance or replacements

⁶ **Predictive maintenance** – the use of modern measurement and signal processing methods to accurately predict and diagnose items/equipment condition during operation. As the technology has advanced, sophistication of all man-made machines and systems has grown and, with that, the nature and needs of maintenance have drastically changed. Maintenance function has become not only more technical, more scientific and more complicated, but also more prominent, more pressing and more paying. Gone are the times when maintenance was considered “a necessary evil” or managers were contented even if all the profits went to maintenance.

1.11. Predictive Maintenance & Anomaly Detection

Authors (João Barataa and Ina Kayserb,2023) talks about Industry 5.0 as a vision of technological transformation balancing the current and future needs of workers keeping sustainability in mind by optimizing product lifecycle and author in (Daniel A Et al,2025) talks about adoption, expansion, and implementation of Artificial Intelligence (AI)-enabled hardware, tools, methods, and semiconductor technologies in the journey towards Industry5.0.

In this world of Industry 5.0, sustainability is top of everyone's mind and (Daniel A Et al,2025) adopting predictive practices not only helps being more efficient operationally and reduces costs but also help us being more sustainable. Detecting potential issues at the right time makes us more energy efficient, this increases the remaining useful life (RUL) of equipment and reduces the need for replacements. The author in (Philip Stahmann Et al,2025) talks about how IIOT and advancement in technology has helped to connect all equipment's and communicate data in real-time with high response and accuracy. This advancement enhances the monitoring of processes and states in industrial engineering, how things that were not connected are connected now and how these data can be used in predicting useful insights.

All these technical advancements have opened new doors for smart manufacturing practices, and every industry is focusing on ways to be more sustainable, efficient, and lean ways to manufacturing goods. Artificial Intelligence, with Machine learning and other subsets, are playing a key role in this advancement

The author in (S. Arena Et al,2022) mentioned that ² Predictive Maintenance (PdM) based on Machine Learning (ML) is one of the most prominent data-driven analytical approaches for monitoring industrial systems, aiming to maximize reliability and efficiency.

Another author, (Falsk Raja,2023) talks about the importance of AI-driven predictive maintenance and how this will revolutionize the field of maintenance and ensure increased equipment uptime, enhanced operational efficiency, and improved asset management.

Also, authors (Lei Y Et al,2020), emphasized the effectiveness of ML in processing multi-parameter data for fault detection, where the authors also mentioned that the integration of ML not only improves diagnostic accuracy but also enables predictive maintenance, reducing downtime and enhancing the overall reliability and efficiency of industrial processes.

In Predictive Maintenance field - Anomaly detection identifies unusual patterns in the behavior of equipment which suggest possible issues with the equipment and will lead to potential failure. Is such case system analyzes data coming from sensors connected to this equipment and establishes a baseline of normal behavior and then flag deviations.

Author talk about the importance of anomaly detection mechanisms in real-time in research (Parthajit Bisal and Prithviraj Jana,2025), In this paper author also talks about

an interesting way to maintain the health status of air compressor using a scoring mechanism.

Authors (Yuvraj Jivan Jadavi and Dr. Bhagya,2024) mention similar approaches in his papers and talk about how we can use the threshold approach for anomaly detection.

Author (Ahmad Et al, 2024) also explains the computational usage and deployment of algorithms in real time systems and solutions due to the quantity of data getting created every second, and provide some interesting points related comparison of computational needs vs performance,

(Pooja Kamat and Rekha Sugandhi Zope,2024) talk about common issues in anomaly detection, which need to be taken care of, data containing noise that could be alerted as an anomaly, the characteristics of an anomaly and of a normal frequently vary, and the anomaly pattern is mostly based on seasonality. All such issues need to be taken care of to have a robust anomaly detection system. Another paper's authors (David Valdivieso López Et al,2024) talk about a common issue where identifying anomalies is very complex as there will be a lot of false positives due to which solution often label standard data as anomaly and in such false possible mitigation is important where the key task is to reduce the number of false positives tagged by the anomaly

1.12. Sensor-based Data Acquisition

Sensors are the foundation of any Predictive Maintenance System (Michael Nsor, 2024), Condition Monitoring and analysis of Sensor Data provide insight to action condition of data. Sensors translate physical conditions—like heat, motion, sound, and pressure—^s into quantifiable electrical signals that enable diagnostics, enhance performance, and facilitate early fault detection. Selecting the appropriate sensor type is crucial, as it directly influences the precision, range, and reliability of the data used in analytical modelling.

As mentioned in paper from some of the most important:

RPM - Motor speed refers to the number of rotations or revolutions a motor's shaft completes in each amount of time. It is usually expressed in RPM (Revolutions Per Minute). Motor speed determines how fast the motor is turning and is a critical factor in its operation, as it affects the performance of the system or machine it powers. We get this data directly from drives.

Motor Power - Motor power represents the energy converted by the motor per unit time and is an important indicator of the motor's working ability. The motor current determines both the input power and the output power of the motor. We get this data directly from drives.

Outlet Pressure Transmitter - In a compressor, the outlet pressure (also called discharge pressure) refers to the pressure of the gas or air after it has been compressed and exits the air compressor. Pressure sensor or pressure transmitter is used.

Outlet Temperature Probe - The outlet temperature in a air compressor refers to the temperature of the gas or air after it has been compressed and exits the compressor.

Air Flow Meters - In a air compressor, air flow refers to the volume or mass of air that the air compressor moves or processes over a given period of time

Noise Sensors - In a air compressor, noise refers to the unwanted sound produced during its operation. This noise can come from various sources such as:

Mechanical vibrations (from moving parts like pistons or rotors)

Air turbulence (especially at the intake and exhaust)

Motor or engine noise

Structural resonance of the air compressor housing

Vibration Monitoring Sensors – Air Compressor, vibration parameters are critical indicators of mechanical health and performance. Monitoring these helps detect issues like imbalance, misalignment, bearing wear, or looseness before they lead to failure.

Vibration Acceleration at X Axis = Measures how quickly the vibration speed changes in x axis

Vibration Acceleration at Y Axis = Measures how quickly the vibration speed changes in Y axis

1.13. Technology use in Predictive Maintenance

Numerous survey studies have previously been published on predictive maintenance across various industries, conducted by different researchers employing a range of methodologies.

The key here is to how best we can use Sensor data coming from Air Compressor and how Deep learning can forecast equipment breakdowns by analysing historical sensor data to detect patterns and predict failures before they occur.

An investigation (Nandini Nalawade Et al,2024) delves into predictive maintenance for electrical machinery by integrating condition monitoring data with advanced machine learning techniques. Through the application of models such as random forests and gradient boosting, the study enhances the accuracy of fault detection and provides deeper insights into equipment health. Despite these advancements, the research underscores significant barriers to implementation, notably the substantial costs and extensive data requirements involved.

In one of the earlier papers Author (Gian Antonio Susto Et al,2015) introduced a methodology applied to a maintenance task within semiconductor manufacturing, specifically related to implant processes. The approach demonstrated better performance in comparison to traditional preventive maintenance (PvM) strategies and a distance-based predictive maintenance (PdM) method using a single SVM classifier. The case study further revealed that Support Vector Machines (SVMs) outperform k-Nearest Neighbors (kNN) classifiers in the implementation of Multi-Class Predictive Maintenance (MC-PdM), and that MC-PdM consistently delivers better results than conventional PvM techniques.

(Parthajit Bisal and Prithviraj Jana,2025) talks another approach where Fuzzy logic offers a sophisticated approach to anomaly detection by evaluating the interrelationships among variables like suction pressure, discharge pressure, and temperature. Instead of binary classifications, it produces a continuous anomaly score that reflects the degree of deviation from normal operating conditions.

Another author (Sai Takawale Et al,2015) mentions about the advantage of using an hybrid approach where LSTM is used for capturing temporal dependencies in sequential data and Random Forest capturing temporal dependencies in sequential data and using Random Forest for feature importance ranking and baseline classification and Autoencoder for unsupervised anomaly detection.

In our case the sensor dataset arrives in continuous stream and entire training dataset is generally not available at the time of model creation. To keep pace with dynamic and ever-changing data environments, it is essential to develop machine learning models capable of real-time learning and adaptation. These models must continuously ingest

new data, recalibrate their understanding, and respond to shifts in the statistical patterns of incoming information. In such fluid scenarios, traditional static learning methods fall short necessitating the use of online learning techniques that evolve alongside the data itself. Author (AYHAM ALKAZAZ,2020) mention of challenges with such data availability and the need to have model with real-time learning and adaptation, author further adds that how Concept drift is a major challenge in such data streams where Adaptive Random Forest method can be helpful.

²Author Doan Ngoc Chi Nam Et al,2019) in their paper mentions challenges in calculation of Remaining useful life which is the measure of how much life is remaining with the machine to run efficiently. One of the main challenge is Concept Drift where data patterns in machines change with time due to wear& tear and how concept drift handler utilizing a gradient descent weighting method. Is used to handle concept drift

Author (Mohammed Amine Moustakim Et al,2020) mentions how LSTM model can be used for Medium-Term Load Forecasting (MTLF) and Very Short-Term Energy Forecasting (VSTEF) and how in there experiment good forecasting accuracy

1.14. Related Research Publications:

Title	Reference	Related Work	Methodology Used
Industry 5.0 – Past, Present, and Near Future	(João Barataa and Ina Kayserb,2023)	Changes in Industrial Revolution and its effect in manufacturing with technological advancement	--
Predictive Maintenance and Smart Sensors Aiming Sustainability: A Perspective from a Bibliometric Analysis	(Daniel A Et al,2025)	Predictive Maintenance and its importance in building sustainable future	--
AI-based real-time anomaly detection in industrial engineering: A structured literature review, taxonomy,	(Philip Stahmann Et al,2025)	Structured literature review, systematically decomposed implementation options of real-time anomaly detection	--

and research agenda			
A novel decision support system for managing predictive maintenance strategies based on machine learning approaches	(S. Arena Et al,2022)	Predictive Maintenance using Data Driven Analytics approach	DDS Based on Decision tree
AI for Predictive Maintenance in Industrial Systems	(Falsk Raja,2023)	AI-driven PdM, from its historical evolution to the technologies and tools involved	Long Short-Term Memory (LSTM) for timeseries.
² Applications of machine learning to machine fault diagnosis: A review and roadmap	(Lei Y Et al,2020),	Fault Diagnostics & relationship between the monitoring data and the health states of machine	artificial neural networks (ANN), support vector machine (SVM), and deep neural networks (DNN)
Air Compressor health monitoring and predictive maintenance by anomaly detection using fuzzy logic and random forest model by machine learning process through Matlab	(Parthajit Bisal and Prithviraj Jana,2025),	Maintain the health status of a air compressor using a scoring mechanism and Integration of Fuzzy Logic with Machine learning.	Fuzzy Logic with Random Forest
Real-Time Anomaly Detection in Air Compressors Using Machine Learning	(Yuvraj Jivan Jadavi and Dr. Bhagya,2024)	Anomaly Detection based on threshold using regression	Random Forest regression models
A Machine Learning Implementation to Predictive Maintenance and Monitoring of	(Ahmad Et al, 2024)	Focuses on comparison of computational needs vs performance	Simple Linear Regression

Industrial Air Compressors			
5 Anomaly Detection for Predictive Maintenance in Industry 4.0-A survey	(Pooja Kamat and Rekha Sugandhi Zope.2024)	Condition Monitoring and Anomaly Detection in Predictive Maintenance	AutoEncoders and Long Short Term Memory (LSTM) deep learning model
Fusing anomaly detection with false positive mitigation methodology for predictive maintenance under multivariate time series Fusing anomaly detection with false positive mitigation methodology for predictive maintenance under multivariate time series	(David Valdivieso López Et al,2024)	Anomaly Detection and innovation ways to identify false positives	KNN ,SVM,RF,XGBOOST algorithm
1 Historical Overview of Maintenance Management Strategies: Development from Breakdown Maintenance to Predictive Maintenance in Accordance with Four Industrial Revolutions	(Peter Poor, Davide Zenesik and Josef Basl, 2019)	History of Maintenance over Industrial Revolutions	--
History of Maintenance: The Evolution of	FTMaintanance	History of Maintenance over Industrial Revolutions	--

Industrial & Facility Maintenance Practices			
Operations & Maintenance Best Practices	US Department of Energy	Maintenance Best Practices	
Machine Learning for Predictive Maintenance: A Multiple Classifier Approach	(Gian Antonio Susto Et al,2015)	Machine Learning for Predictive Maintenance: A Multiple Classifier Approach	Support Vector Machines (SVMs)
Predictive Maintenance for Industrial Equipment using IIOT, AI and ML	(Nandini Nalawade Et al,2024)	Predictive Maintenance for Industrial Equipment using IIOT, AI and ML	LSTM
Predictive Maintenance Using Machine Learning for Engineering Systems Through Real-Time Sensor Data and Anomaly Detection Models	(Michael Nsor, 2024)	Predictive Maintenance Using Machine Learning for Engineering Systems Through Real-Time Sensor Data and Anomaly Detection Models	Model Comparison - Random Forest, XGBoost, SVM, etc.
AI-Powered Predictive Maintenance Using Machine Learning for Industrial IoT Systems	(Sai Takawale Et al,2015)	AI-Powered Predictive Maintenance Using Machine Learning for Industrial IoT Systems	Hybrid Model
Evaluation of Adaptive random forest algorithm for classification of evolving data stream	(AYHAM ALKAZAZ,2020)	Evaluation of Adaptive random forest algorithm for classification of evolving data stream	Adaptive Random Forest
Machine learning based concept drift detection for	(Jan Zenisek Et al,2019)	Concept Drift in machines and	--

predictive maintenance			
A Method to Handle Concept Drift in Predicting Remaining Useful Life	Doan Ngoc Chi Nam Et al,2019)	Concept Drift in Predicting Remaining Useful Life	--
Prediction of electric power and load forecasting using LSTM technique for EMS	Mohammed Amine Moustakim Et al,2020	Prediction of electric power and load forecasting using LSTM technique for EMS	LDTM

Table 4 - Research Publication Table

1.15. Discussion & Summary:

With the references mentioned in above sections it is very evident that Predictive Maintenance is an important area of research in manufacturing for Industrial systems, and their big opportunity in research and innovation, which will add value to this field

4
Most of the studies in this area of Predictive Maintenance systems operate on historical data but fail to implement real-time monitoring and anomaly detection mechanisms and provide a minimum required user interface to visualize & monitor critical parameters & anomalies.

In majority of the studied it is expected that we have a large historical dataset labeled available to us for training, which is not the case in majority of the actual scenarios thus the focus here is to find appropriate method which can work with limited historical dataset and can learn on the fly as and when we stream of data is received from field.

- CHAPTER 3: RESEARCH METHODOLOGY

1.16. Introduction

This study investigates anomaly detection in critical elements of a Single Stage Screw Air Compressor, with particular emphasis on the bearing system. By analyzing the parameters that govern bearing performance, the research aims to uncover patterns indicative of abnormal behavior. The following sections outline the foundational concepts and computational techniques that support the methodology presented in this thesis.

1.17. Research Methodology

1.17.1. Introduction

1.17.2. Architecture for Proposed System

Below is the architecture diagram of the proposed system for Predictive Maintenance & Realtime Anomaly Detection for Single Stage Air Compressor.

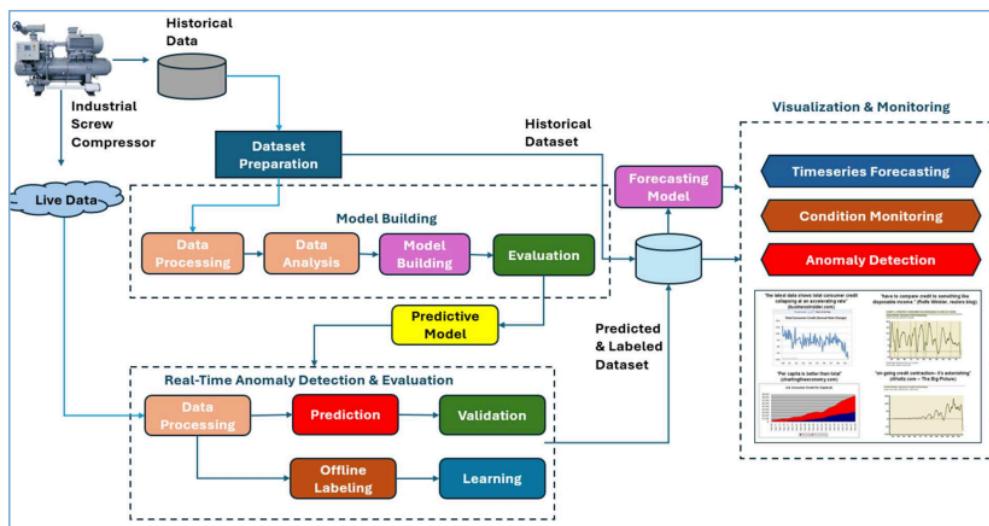


Figure 3 - Architecture Diagram for proposed system

Here, for this research, we are employing Adaptive Random Forest Classifier models to detect anomalies using for bearing (one of the major components of Single Stage Screw Air Compressor).

Major Components of Architecture are given below:

Single Stage Screw Air Compressor

Our Paper's Aim is to detect anomalies in components of Air Compressor and forecast key parameters using advanced machine learning algorithms. As referred in (Shivansh Sabhadiya,2021), Air compressor is an important equipment for a variety of industries which uses kinetic energy to create pressurized air. This air which is pressurized and stored in a storage air tank is used to power various industrial tools and processes. For more details, please refer 1.1

Historical Data

Historical Data in context of air Compressor includes sensor-based data for a considerable amount of time labeled with faults. For this study we have sample Open-Source data available for a period of time (a part of this data set is used as historical data and remaining is used as Live Data). All data are labeled for faults. Detailed information on available sensor values and fault label please refer **Error! Reference source not found.**

Live Data

Live data in this context refers to real-time data coming from air compressors during production, for this study we are using the same sample Open-Source data mentioned in above section, labeled data information is removed from this source before including as part of LiveData. Detailed information on available sensor values and fault label please refer **Error! Reference source not found.**

Data Preparation

Data processing involves converting unrefined data into an organized, standardized format suitable for analytical tasks or predictive modelling. This essential phase plays a pivotal role in ensuring that the insights derived are both accurate and dependable. please refer **Error! Reference source not found.**

Data Processing

In this stage cleaned, transformed and structured for analysis and creation of model, please refer 1.17.9

Model Building

Model Building involves creation of initial models using Historical dataset from air compressors. In this study we are using Adaptive random forest (ARF) which Adaptive Random Forest (ARF) is an enhanced version of the traditional Random Forest algorithm, specifically tailored to handle continuously changing data streams. It integrates the core principles of Random Forest with dynamic strategies that respond effectively to various types of concept drift, allowing the model to maintain accuracy as data patterns evolve over time.

Evaluation

Created Models are evaluated using popular and accepted evaluation metrics. This helps access the overall performance of created model. For details on evaluation metrics, please refer to **Error! Reference source not found.**

Predictive Model

Once final model is created and evaluated for optimal fit, Model Object is saved for detecting anomaly from new batch of dataset from live data. Further details on Predictive Model is given in later sections.

Realtime Prediction & Evaluation

Realtime data is received in batches from air compressor (refer Live Data) and Predictive Model is used for detecting Anomaly. During prediction overall performance of Model is evaluated for each batch

Realtime Data Offline Labeling

Every newly received batch goes for labelling where we label every batch for any anomaly physically detected by engineer. For this study we are using the labelled information available with the dataset for labelling.

Realtime Data Learning

Once the new batch is labelled, it is passed to Predictive Model for learning.

Visualization & Monitoring

Visualization and Monitoring is a UI application which meets the objective of providing an optimized value-driven driven User-Interface to maintenance engineers for real-time monitoring of parameters and detected anomalies

- **Timeseries Forecasting** – Univariate forecasting of critical parameters is done using LSTM method and visualized using application.
- **Condition Monitoring** – Key critical parameters are monitored for a period via timeseries plots to assess the health and performance of air compressor components.
- **Anomaly Detection:** Predicted values/Anomalies are plotted over timeseries chart and visualized using application

The Adaptive Random Forest classifier is trained on a pre-processed, labelled dataset and employed to detect anomalies in batches of live sensor data. In this research, an existing dataset is repurposed to emulate real-time monitoring conditions.

Additionally, we are using Simple Forecasting LSTM Models to forecast key critical parameters from air compressor

1.17.3. Data Preparation

1.17.3.1. Dataset Selection

Data for any industrial air compressor is collected from multiple sensors and equipment mounted on air compressor unit. Data from these sensors and equipments are collected on real-time basis for condition monitoring.

Sensor data is typically recorded and stored in a time-series format. In this study, we utilize an open-source CSV dataset collected from an industrial screw air compressor. The dataset includes key operational parameters and is labelled with fault information corresponding to various air compressor components.

Dataset used is from (Neuraldesigner,2024) available at <https://www.neuraldesigner.com/wp-content/uploads/2023/10/aircompressor.csv>

1.17.3.2. Dataset Preparation

List of parameters from data set is given below in table (Table2).

Parameter Name	Data Type	Unit Of Measure	Description
RPM	Numerical	r/min	Motor speed refers to the number of rotations or revolutions a motor's shaft completes in each amount of time. It is usually expressed in RPM (Revolutions Per Minute). Motor speed determines how fast the motor is turning and is a critical factor in its operation, as it affects the performance of the system or machine it powers.
Motor Power	Numerical	kW(Kilo Watt)	Motor power represents the energy converted by the motor per unit time and is an important indicator of the motor's working ability. The motor current determines both the input power and the output power of the motor

Outlet Pressure	Numerical	bar	In a air compressor, the outlet pressure (also called discharge pressure) refers to the pressure of the gas or air after it has been compressed and exits the compressor
Outlet Temperature	Numerical	Degrees Celsius (°C)	The outlet temperature in a air compressor refers to the temperature of the gas or air after it has been compressed and exits the compressor.
Air Flow	Numerical	CFM – Cubic Feet per Minute	In a air compressor, air flow refers to the volume or mass of air that the compressor moves or processes over a given period of time
Noise	Numerical	decibel (dB).	<p>In a air compressor, noise refers to the unwanted sound produced during its operation. This noise can come from various sources such as:</p> <p>Mechanical vibrations (from moving parts like pistons or rotors)</p> <p>Air turbulence (especially at the intake and exhaust)</p> <p>Motor or engine noise</p> <p>Structural resonance of the air compressor housing</p>
Water Pump, Inlet Pressure	Numerical	bar	Water Pump's Inlet pressure. Water Pump is used in air compressor to circulate water and coolant through air Compressor to reduce heat
Water Pump, Outlet Pressure	Numerical	bar	Water Pump's Outlet pressure.
Water Pump Power Consumption	Numerical	kW(Kilo Watt)	Power Consumed by WaterPump
Water Flow	Numerical	m ³ /h(Cubic meters per hour)	Flow Rate of water or coolant in air Compressor

Oil Pump Power Consumption	Numerical	kW(Kilo Watt)	<p>Power Consumed by Oil Pump</p> <p>Oil Pump distribute lubricating oil across the machine's components to maintain efficient performance, minimize friction, and regulate temperature.</p>
Oil Tank temperature	Numerical	Degrees Celsius (°C)	Temperature captured at Oil tank through thermostat,
Vibration Ground Acceleration @ X	Numerical	m/s ² (meters per second squared)	<p>Air Compressor, vibration parameters are critical indicators of mechanical health and performance. Monitoring these helps detect issues like imbalance, misalignment, bearing wear, or looseness before they lead to failure.</p> <p>Ground Acceleration at X Axis = Taken at the base and measures how quickly the vibration speed changes in x axis</p>
Vibration Head Acceleration @ Y	Numerical	m/s ² (meters per second squared)	<p>Air Compressor, vibration parameters are critical indicators of mechanical health and performance. Monitoring these helps detect issues like imbalance, misalignment, bearing wear, or looseness before they lead to failure.</p> <p>Ground Acceleration at Y Axis = Taken at the base and measures how quickly the vibration speed changes in y axis</p>
Vibration Acceleration @ Z	Numerical	m/s ² (meters per second squared)	<p>Air Compressor, vibration parameters are critical indicators of mechanical health and performance. Monitoring these helps detect issues like imbalance, misalignment, bearing wear, or looseness before they lead to failure.</p>

			Ground Acceleration at Z Axis = Taken at the base and measures how quickly the vibration speed changes in z axis
Vibration Head Acceleration @ X	Numerical	g (acceleration due to gravity)	Air Compressor, vibration parameters are critical indicators of mechanical health and performance. Monitoring these helps detect issues like imbalance, misalignment, bearing wear, or looseness before they lead to failure. Head Acceleration at X Axis = Taken at the head and measures how quickly the vibration speed changes in x axis
Vibration Head Acceleration @ Y	Numerical	g (acceleration due to gravity)	Air Compressor, vibration parameters are critical indicators of mechanical health and performance. Monitoring these helps detect issues like imbalance, misalignment, bearing wear, or looseness before they lead to failure. Head Acceleration at Y Axis = Taken at the head and measures how quickly the vibration speed changes in y axis
Vibration Head Acceleration @ Z	Numerical	g (acceleration due to gravity)	Air Compressor, vibration parameters are critical indicators of mechanical health and performance. Monitoring these helps detect issues like imbalance, misalignment, bearing wear, or looseness before they lead to failure. Head Acceleration at Z Axis = Taken at the head and measures how quickly the vibration speed changes in z axis
Bearing Fault	Categorical	Bool	Bearing Fault is the categorical value where 1 means bearing fault and 0 means no fault

Water Pump Fault	Categorical	Bool	Water Pump Fault is a categorical value where 1 means Water Pump fault and 0 means no fault
Radiator	Categorical	Bool	Label for Radiator Failure where 1 means radiator fault and 0 means no fault
ExValve	Categorical	Bool	Label for Valve Failure where 1 means valve fault and 0 means no fault
acmotor	Categorical	String	Label for AC Motor Failure where “Not Stable” means AC Motor fault and “Stable” means no fault

Table 5 - List of Dataset Parameters

1.17.3.3. Dataset Usage

Here in this study, we are focussing on 1 air coprocessor component to keep the overall scope small. We have labelled data available for 5 components in dataset out of which we chose too monitoring Bearing Fault.

Bearing Fault
Water Pump Fault
Radiator Fault
ExValve Fault
Acmotors Fault

Table 6 - Labelled Fault List

Bearing Fault:

Bearings, though small, play a crucial role in the operation of any air compressor. Inadequate lubrication can cause the motor to overheat, deteriorate, and potentially fail ahead of its expected lifespan. Regularly maintaining clean, well-lubricated bearings is a simple yet effective way to ensure smooth and reliable air compressor performance. In this study we are focussing on Bearing fault.

The dataset contains 1000 records captured at a defined period or when any major event happened.

We will be reusing the existing data for testing real-time data-related workflows. We are using 60% of dataset as Historical Data and remaining data is converted into small batches for Live Data Simulation.

1.17.4. Proposed Method

Random Forests remain among the most widely adopted machine learning algorithms for batch-based, non-streaming applications. Their popularity stems from strong predictive performance and minimal requirements for data preprocessing or hyperparameter optimization. However, when applied to dynamic and continuously evolving data streams, Random Forests fall short of being considered cutting-edge—especially when compared to more adaptive methods like bagging and boosting algorithms.

Over time, single-stage screw air compressors naturally undergo aging, experiencing wear and tear, shifts in ambient operating temperatures, and gradual changes in overall efficiency. As a result, the sensor data stream associated with these machines evolves continuously.

This dynamic nature of data poses a challenge for predictive models that rely on fixed input-output relationships, often leading to diminished accuracy and declining performance over time.

Concept Drift

In machine learning and data mining, concept drift describes the gradual shift in the relationship between input features and output targets as the underlying data patterns evolve over time.

Concept drift can occur in supervised learning scenarios where predictions are continuously generated, and data is gathered over time. Such situations are commonly referred to as online learning problems, due to the anticipated evolution in data patterns as time progresses.

To mitigate the issue of evolving data and also to cater to streaming data, In this paper we are studying the performance of Adaptive Random Forest for predicting Anomaly in Single Stage Screw Air Compressor.

1.17.4.1. Algorithms & Techniques

In this section we will be discussing the Algorithms and Techniques used for the proposed system.

1.17.4.1.1. Adaptive Random Forest

Adaptive Random Forest (ARF) is a modified version of the classic Random Forest algorithm, tailored to handle continuously evolving data streams in real-time environments.

It integrates core principles of Random Forests with dynamic strategies designed to manage various forms of concept drift. Unlike standard Random Forests, which operate as batch ensembles trained on static datasets, ARF is built for online learning environments where data arrives sequentially. To enable this real-time adaptability, ARF replaces conventional decision trees with Hoeffding trees—incremental learners capable of processing large-scale streaming data efficiently.

Adaptive Random Forest (ARF) incorporates a robust resampling strategy along with adaptive mechanisms that effectively handle various forms of concept drift, all without requiring intricate tuning across different datasets.

Here we are using ARF to detect Anomaly(bearing fault) on every received batch of sensor dataset.

1.17.4.1.2. LSTM

² Long Short-Term Memory (LSTM) networks mark a significant evolution in the architecture of recurrent neural networks (RNNs) within the field of deep learning. These advanced models are specifically designed to overcome limitations in capturing long-range dependencies, making them highly effective for tasks involving sequential data such as natural language processing and speech recognition. LSTMs excel across diverse applications by functioning as intelligent systems that selectively retain and discard information over time. Their nuanced handling of temporal patterns enables superior performance in scenarios where understanding the order and timing of data is crucial.

(LSTM) networks, a specialized form of recurrent neural networks (RNNs), have proven highly effective for analysing and predicting patterns in time series data. Their strength lies in the ability to capture and retain long-range dependencies within sequential inputs, enabling them to process entire data sequences with precision. This capability has positioned LSTMs as a preferred solution for time series forecasting across various domains.

1.17.4.1.3. NiceGUI

NiceGUI is a user-friendly UI framework built with Python that runs directly in your web browser. It allows you to effortlessly build interactive elements like buttons, dialogs, Markdown content, 3D visualizations, charts, and more.

Ideal for lightweight web applications, dashboards, robotics interfaces, and smart home systems, NiceGUI also proves valuable during development tasks—such as fine-tuning machine learning models or adjusting motor control parameters.

We are using NICEGUI here to build a simple visualization application as part of one of the objective to study simple optimized easy to use application with capability of timeseries forecasting, Condition Monitoring and Anomaly Detection.

1.17.4.2. Proposed Research System

As mentioned in Architecture above There are 2 major stages in the proposed system.

1.17.4.2.1. Stage1:

The objective here is to initialize and develop a predictive model using historical compressor sensor data. This involves a workflow where the data is first cleaned, processed, and analyzed. Key features are then selected based on relevance and impact, forming the foundation for building an effective and accurate model.

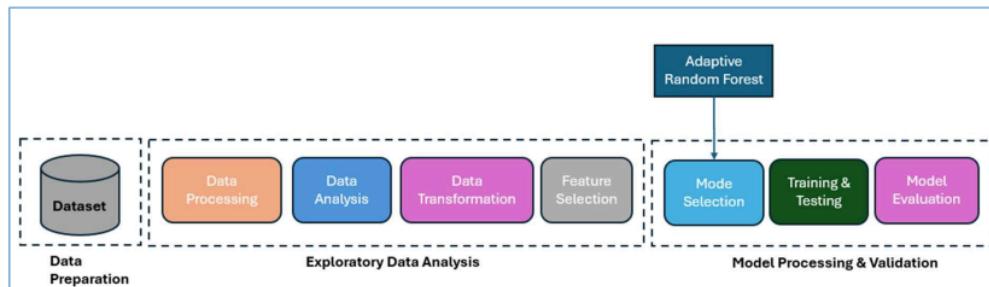


Table 7 - Architecture diagram on Proposed Research Structure Stage1

- **Data Preparation**:- Get Sensor based dataset . refer for more details
- **Data Processing(pre-EDA)** : Process Sensor based data. refer for more details
- **Exploratory Data Analysis**: involves examining data to uncover underlying distributions, detect trends and patterns, and reveal hidden anomalies or potential issues that may impact further analysis or modelling. refer for more details

- **Feature Selection:** Feature importance was evaluated during model development, and the process was repeated multiple times to identify and retain the features that contributed most to optimal performance. refer for more details
- **Model Creation:** A predictive model has been designed and trained using historical air compressor data.
- **Model Evaluation:** The model's performance has been assessed; for a detailed overview of the evaluation metrics applied, please refer to the corresponding section.
- **Save Model:** The model is stored for ongoing use in predicting live batch data generated by the air compressor system.

1.17.4.2.2. Stage 2

The objective here is to fetch Live Sensor data coming from compressor and pass data to a workflow where the data is first cleaned, processed and model from Stage 1 is used to predict anomalies, after this data goes to a holding area where data is labelled manually and then model is trained for new received batch.

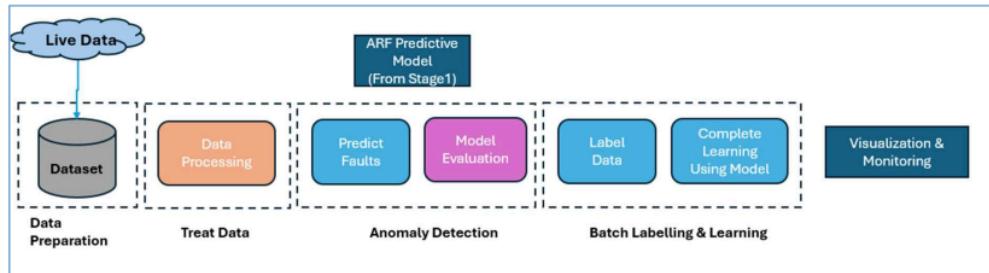


Table 8 - Architecture diagram on Proposed Research Structure Stage2

- **Data Preparation:-** Get Sensor based batch dataset in real-time . refer for more details
- **Treat Data:-** Treat & process data based on Stage1
- **Predict Faults:-** Detect Anomalies using Predictive Model from Stage1
- **Model Evaluation:-** Evaluate model performance after prediction

- **Label Data:** Newly received batch of data is labelled manually. When the model should be re-trained depends on many factors, in many cases Data scientists might periodically retrain models (e.g., weekly, monthly) in others when monitoring systems detect a significant drop in performance.
-
- **Learning :** Complete learning using a predictive model, Use Same model from Stage1

1.17.5. Evaluation Method

Models are evaluated Using many metrics

Accuracy:

Accuracy is a widely used metric for assessing the effectiveness of classification models. It represents the percentage of correct predictions made out of all predictions generated by the model.

$$\text{Accuracy} = (\text{True Positives} + \text{True Negatives}) / \text{Total Predictions}$$

F1

The F1 Score is a performance metric that harmonizes Precision and Recall into one comprehensive value, providing a balanced assessment of a model's accuracy particularly valuable in scenarios with imbalanced datasets.

$$F1 \text{ Score} = 2 \times ((\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}))$$

Precision

Precision measures the accuracy of positive predictions made by a model. It is defined as the ratio of true positive outcomes to the total number of instances predicted as positive. In other words, it indicates how many of the model's "yes" predictions were actually correct. A higher precision means fewer incorrect positive predictions, also known as false positives (FP). The formula for precision is:

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives})$$

Recall:

Also known as Sensitivity is a metric that quantifies the proportion of actual positive instances that are correctly identified by a predictive model. It essentially evaluates how well the model captures the true positive cases using its predicted positive rule (+P). A key advantage of recall is its ability to indicate the extent to which relevant cases are successfully retrieved. However, in the field of Information Retrieval, recall is often given less emphasis due to assumptions such as the abundance of relevant documents, the irrelevance of which specific subset is retrieved, and the inability to assess the relevance of documents that remain unreturned.

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$$

Kappa

Cohen's Kappa, commonly referred to as Kappa, is a statistical measure that evaluates the level of agreement between two raters or classification methods, while adjusting for the likelihood of agreement happening by chance.

$$K = (Po - Pe) / (1 - Pe)$$

- PoP_o = Observed agreement (the proportion of times both raters agree)
- PeP_e = Expected agreement by chance
-
- $\kappa=1$ kappa = 1: Perfect agreement
- $\kappa=0$ kappa = 0: Agreement no better than chance
- $\kappa<0$ kappa < 0: Less agreement than expected by chance

Confusion Matrix

A Confusion Matrix is a structured table that assesses the effectiveness of a classification model by contrasting its predicted labels with the actual outcomes. It offers a granular view of both accurate and erroneous predictions across various categories.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 9 - Confusion Matrix

ROC Curve

A Receiver Operating Characteristic (ROC) curve is a visual representation that demonstrates how a binary classification model performs across a range of threshold values. It highlights the balance between the True Positive Rate (TPR) and the False Positive Rate (FPR), helping to assess the model's ability to distinguish between classes at different decision boundaries.

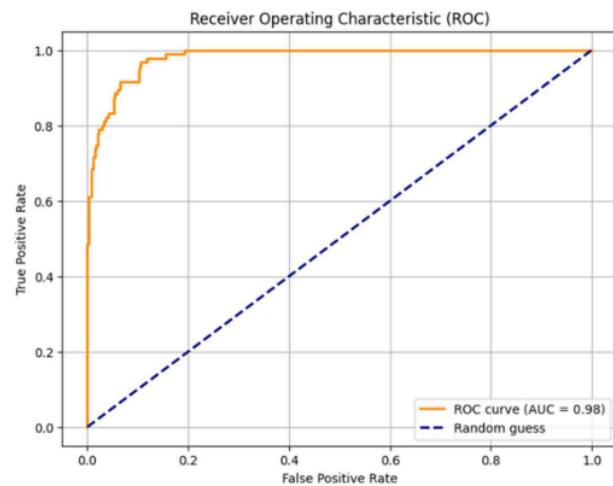


Figure 4 - ROC Curve Plot

1.17.6. Visualization & Monitoring

Visualization and Monitoring is a UI application which meets the objective of providing an optimized value-driven User-Interface to maintenance engineers for real-time monitoring of parameters and detected anomalies.

1.17.6.1. Tools & Technology

We are using NiceGUI which is a user-friendly UI framework built with Python that runs directly in your web browser. It allows you to effortlessly build interactive elements like buttons, dialogs, Markdown content, 3D visualizations, charts, and more.

NiceGUI is Open-source with MIT License

1.17.6.2. Anomaly Detection

Anomaly detection identifies data points predicted by the model that significantly diverge from typical or expected patterns. In this study, In this paper we are predicting Bearing Fault. These anomalies are visualized on a line chart, with the detected fault points highlighted in red.

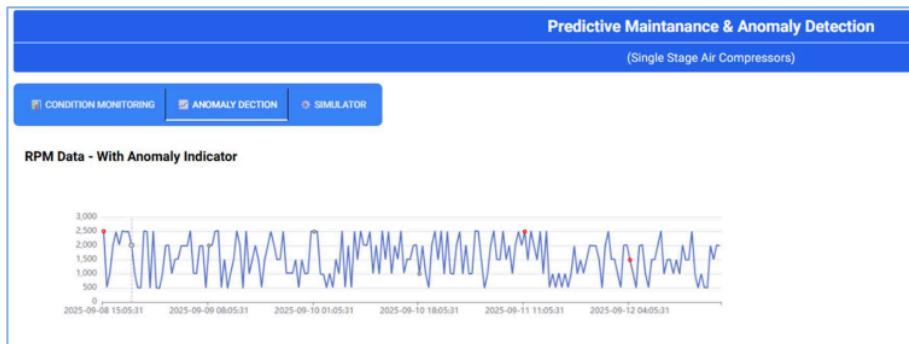


Figure 5 - Anomaly Detection Plot (Visualization)

Details are also shown in tabular format with anomalies highlighted in Red color.

Predicted Anomaly Details										
Detection Timestamp	RPM	Air Flow	Noise(db)	Water Outlet Temp	Water Flow	GACCK	HACCK	Record Timestamp	Predicted Value	
2025-10-12 19:05:31	1983	1226.747468	54.86734649	54.86734649	54.86734649	0.567782698	1.057036723	2025-09-09 07:05:31	0	
2025-10-12 19:05:31	1501	891.1224514	55.05153435	55.05153435	55.05153435	0.579689772	1.075257974	2025-09-09 19:05:31	0	
2025-10-12 19:07:05	1988	1211.278282	63.93713263	63.93713263	63.93713263	0.550445274	1.05636955	2025-09-08 19:07:05	1	
2025-10-12 19:05:33	499	300.4365268	43.83011576	43.83011576	43.83011576	0.719024871	1.218585675	2025-09-08 20:05:33	0	
2025-10-12 19:05:33	992	188.7662612	45.10659017	45.10659017	45.10659017	0.617680656	1.110620912	2025-09-10 11:05:33	0	

Figure 6 - Anomaly Detection Details (Visualization)

1.17.6.3. Condition Monitoring

Visualization of real-time and historical data for analysis help prevent failures and downtime, following key parameters are plotted for analysis

Power Consumption – Time series trend with Upper and Lower Control Limits.

Air Flow – Time series trend with Upper and Lower Control Limits.

Noice DB – Time series trend with Upper and Lower Control Limits

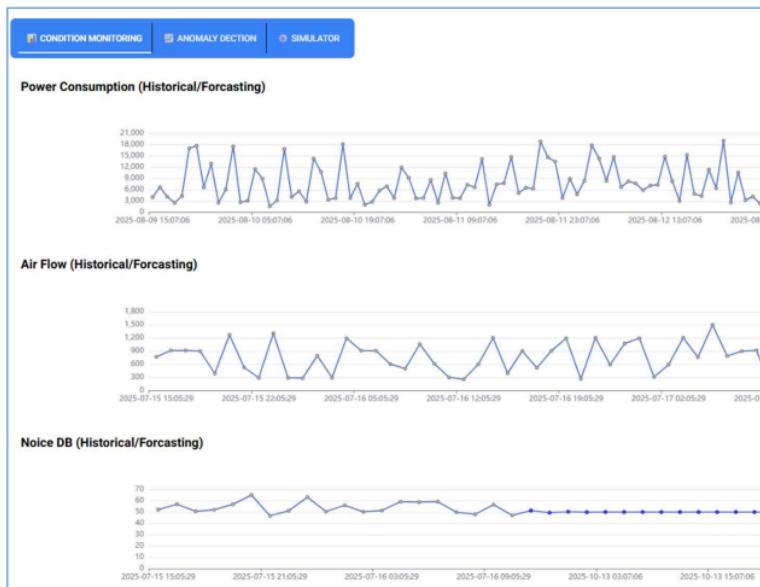


Figure 7 - Condition Monitoring Anomaly Detection Details (Visualization)

1.17.6.4. Timeseries Forecasting

Univariate forecasting of critical parameters is done using LSTM method and visualized using application.

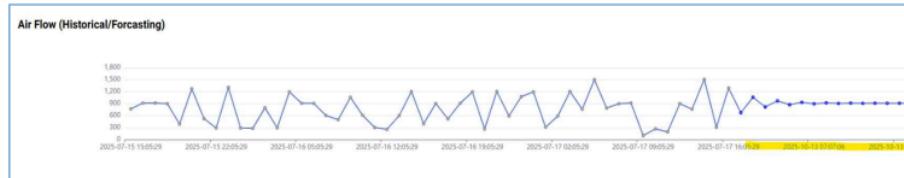


Figure 8 - Timeseries Forecasting

1.17.6.5. System Simulator

To simulate proposed system, we have created a Simulator which simulate end to end execution of both Stage 1 & Stage 2.

Here we can create Model , simulate live batches and monitor overall performance of model

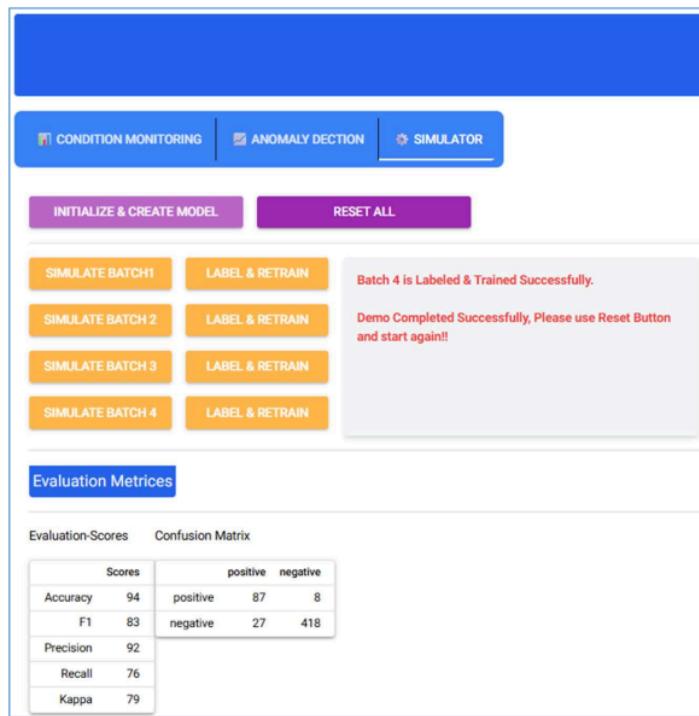


Table 10 - Visualization – Simulator (VI) visualization

Initialize Model

The first step of Simulation is to Initialize Model, here 60% of raw data is considered as Historical Data and model is created after data processing and evaluation results are updated.

Simulate Batch 1

Live data is simulated by taking 10% of raw data, record set are processed and initialized model is used to detect anomalies, anomalies are potted in Anomaly dashboard.

Label & Retrain Batch1

In this experiment, Batch 1 consists of pre-labeled simulated data. Since labeled data is already available, we simulate the learning process by feeding this data into the initialized model, allowing it to train and adapt based on the known patterns and classifications.

Simulate, Label & Retrain remaining batches

Same process above is repeated for Batch 2,3 and 4

Evaluation Matrix Scores

Evaluation Matrix shows evaluation scores for Accuracy, F1, Precision, Recall and Kappa, Evaluation result is updated for initial model and at prediction of every batch.

Confusion Matrix

Confusion Matrix is shown in tabular format, this table is updated for initial model and at prediction of every batch.

Reset All

This will reset all parameters related to simulation and make system ready for next cycle of execution

- **CHAPTER 4: RESEARCH ANALYSIS**

1.17.7. Introduction

This section presents a detailed examination of air Compressor datasets to summarize their main characteristics. The objective is to gain insights into the data's structure, uncover patterns, and identify anomalies prior to implementing any modelling approaches.

1.17.8. Dataset Description.

For this study, we utilize the open-source air compressor dataset referenced in earlier chapters. The dataset is partitioned to serve dual purposes: as historical data and as four separate batches of live sensor data.

1.17.9. Dataset Processing

Each entry in the dataset captures information recorded either at specific time intervals or during significant events in the air compressor's operation. These time-stamped records, containing multiple parameters, collectively offer a detailed overview of the air compressor's performance and health status over time.

1.17.9.1. Elimination of Variables

Data Cleaning of the raw dataset is done to take care of missing values, duplicate data, and fix inconsistencies like formatting, typos, etc.

- Following features are removed.

Feature Name	Reason
id	Identified not required for analysis
Acmotors Fault	This has a single label for all records, not helpful in classification
Water Pump Fault	In this study we are focusing on single classification label i.e.. Bearing fault
Radiator Fault	In this study we are focusing on single classification label i.e.. Bearing fault
ExValve Fault	In this study we are focusing on single classification label i.e.. Bearing fault

Table 11 - Removed Dataset Features

- There are no missing values.

Number of missing values	
id	0
rpm	0
motor_power	0
torque	0
outlet_pressure_bar	0
air_flow	0
noise_db	0
outlet_temp	0
wpump_outlet_press	0
water_inlet_temp	0
water_outlet_temp	0
wpump_power	0
water_flow	0
oilpump_power	0
oil_tank_temp	0
gaccx	0
gaccy	0
gaccz	0
haccx	0
haccy	0
haccz	0
bearings	0
wpump	0
radiator	0
exvalve	0
acmotor	0

Figure 9 - Feature Missing Values

1.17.9.2. Data Transformation

- Timestamp is added to each record of dataset indicating the data collected in a defined frequency. This is used in the time series analysis of key features.

	rpm	air_flow	noise_db	water_outlet_temp	water_flow	gaccx	haccx	bearings	wpump	radiator	exvalve	Timestamp
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.0000	1000.0000	1000.0000	1000.0000	1000
mean	1499.516000	754.673930	53.412121	96.636589	53.708176	0.601789	1.101253	0.2000	0.2000	0.2000	0.2000	2025-09-19 01:36:01 999999744
min	480.000000	95.058952	39.879017	47.259238	38.570791	0.540045	1.039910	0.0000	0.0000	0.0000	0.0000	2025-08-29 08:06:02
25%	988.750000	305.388152	46.490976	81.365268	50.459245	0.558092	1.057717	0.0000	0.0000	0.0000	0.0000	2025-09-08 17:51:02
50%	1500.000000	608.626411	52.708480	96.084809	58.112667	0.576814	1.075392	0.0000	0.0000	0.0000	0.0000	2025-09-19 03:36:02
75%	2010.000000	1193.266273	60.158248	110.989064	58.552897	0.615021	1.113696	0.0000	0.0000	0.0000	0.0000	2025-09-29 13:21:02
max	2520.000000	1539.911634	74.401986	161.123805	59.174992	0.726604	1.227134	1.0000	1.0000	1.0000	1.0000	2025-10-09 23:06:02
std	707.681915	442.743491	8.055363	20.557276	6.587259	0.056719	0.058546	0.4002	0.4002	0.4002	0.4002	Nan

Figure 10 - Timestamp Assignment

- All outliers are systematically identified and addressed to prevent distortion in the analysis, ensuring that any irregular values are properly managed to maintain data integrity.

1.17.9.3. Feature Selection:

Feature selection is the process of identifying and eliminating irrelevant or redundant features, which plays a crucial role in enhancing the performance and interpretability of ML models.

We are using built in feature of Random Forest classifier to identify and retain most influential features. Throughout this process, the model inherently assesses each feature's impact on decreasing impurity—such as Gini index or entropy—making it well-suited for determining and ranking feature importance.

	Feature	Importance
5	noise_db	0.333711
11	water_flow	0.115802
4	air_flow	0.093639
14	gaccx	0.037514
0	rpm	0.035230
9	water_outlet_temp	0.033758
13	oil_tank_temp	0.031050
17	haccx	0.029365
7	wpump_outlet_press	0.028250
6	outlet_temp	0.028092
3	outlet_pressure_bar	0.027625
8	water_inlet_temp	0.025445
16	gaccz	0.025407
10	wpump_power	0.024963
19	haccz	0.024664
12	oilpump_power	0.022578
2	torque	0.022451
15	gaccy	0.022373
18	haccy	0.020476
1	motor_power	0.017607

Figure 11 - Feature Importance

Performance is evaluated while checking feature importance , this step is repeated multiple times to find the most influential feature with highest scores

```
Accuracy: 0.9466666666666667
Precision: 0.9483747276688452
Recall: 0.9466666666666667
F1 Score: 0.9438296112489661
Confusion Matrix:
[[240  1]
 [ 15  44]]
Classification Report:
precision    recall  f1-score   support
      0       0.94      1.00      0.97     241
      1       0.98      0.75      0.85      59
      accuracy                           0.95     300
      macro avg       0.96      0.87      0.91     300
      weighted avg    0.95      0.95      0.94     300
```

Figure 12 - Stage1 - Evaluation Matrix

As a result of Feature selection process following feature are identified as most influential features are remaining features are removed from dataset before creation of model.

Feature Name
rpm
air_flow
Noise_db
water_outlet_temp
water_flow
gaccx
haccx

Table 12 - Selected Feature List

1.17.9.4. Data Scaling:

Feature scaling is an essential preprocessing technique that balances the influence of all input features during model training. Without it, variables with larger numerical ranges can overshadow those with smaller ones, potentially skewing the model's learning and resulting in less accurate or biased outcomes.

Here we are using Standard Scaler from river, transforms the data so that each feature has:

- A **mean (μ)** of 0
- A **standard deviation (σ)** of 1

```
[[[-1.41 -1.01 -1.56 ... 0.82 1.87 1.92]
 [-1.39 -1.01 -1.61 ... 0.83 1.75 1.87]
 [-1.42 -1.01 -1.55 ... 0.81 1.76 1.87]
 ...
 [ 1.43 -0.63  0.87 ... 0.62 -0.78 -0.77]
 [ 1.4 -0.65  1.08 ... 0.61 -0.82 -0.84]
 [ 1.41 -0.64  1.05 ... 0.63 -0.93 -0.89]]]
```

Figure 13 - Stage 1 - Data Scaling

The scaled data includes **positive and negative values**, depending on how far each original value is from the mean.

1.17.9.5. Splitting of Original Dataset

Original Data Set is broken down into multiple smaller datasets to simulate multiple live streaming sensor data coming from air compressor.

As part of the simulation, all processed datasets along with their prediction results are stored in CSV files.

Dataset	Split %
---------	---------

Historical Data	60%
Batch1	10%
Batch2	10%
Batch3	10%
Batch4	10%

Table 13 - Dataset Split %

1.17.9.6. Exploratory Data Analysis

In this section, we will explore, summarize, and visualize the data to uncover patterns, trends, relationships, and potential issues. It helps you understand the structure and meaning of the data before applying any modeling or machine learning.

1.17.9.6.1. Univariate Analysis

The dataset was analysed for bearing fault distribution, revealing that 20% of the records are labelled as bearing fault cases.

There is class imbalance , which is expected as we are dealing with fault data. Bearing fault instances account for only 20% of the dataset, with the remaining 80% likely corresponding to normal or non-fault conditions. Such an imbalance can influence the effectiveness of classification models, potentially leading to biased predictions toward the majority class.

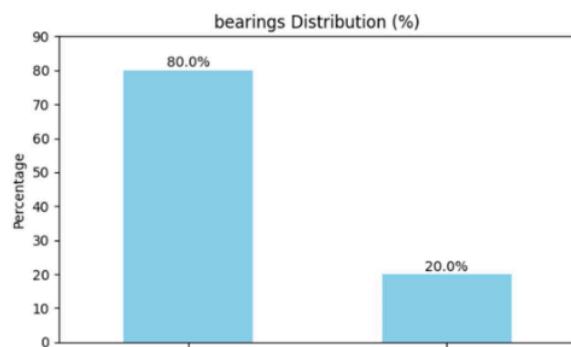


Figure 14 - Bearing Fault Distribution Chart

1.17.9.6.2. Correlation analysis

Correlation analysis evaluates how strongly and in what direction two or more variables are related. It uncovers patterns and interdependencies within the data, which are vital for tasks like feature selection, building predictive models, and gaining insights into underlying trends.

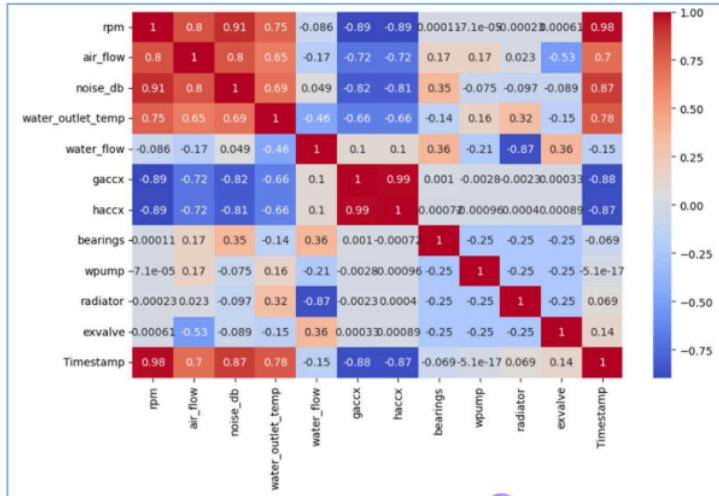


Figure 15 - Correlation Matrix Heatmap

High Positive Correlations:

- rpm vs. air_flow (0.98):** A near-linear relationship where increased engine speed closely corresponds with higher airflow.
- air_flow vs. noise_db (0.89):** Indicates that greater airflow tends to produce more noise.
- rpm vs. noise_db (0.87):** Suggests that higher rotational speeds result in louder operational sound.
-

Weak or Insignificant Correlations:

- gacx and hacx:** These variables exhibit very low correlation with others, suggesting they may operate independently of the main system dynamics or reflect isolated measurements.

Water outlet temperature shows moderate correlation with variables like rpm, air_flow, and noise_db. This implies that as the engine or pump speed increases, the water outlet temperature tends to rise, likely due to increased thermal output from mechanical activity.

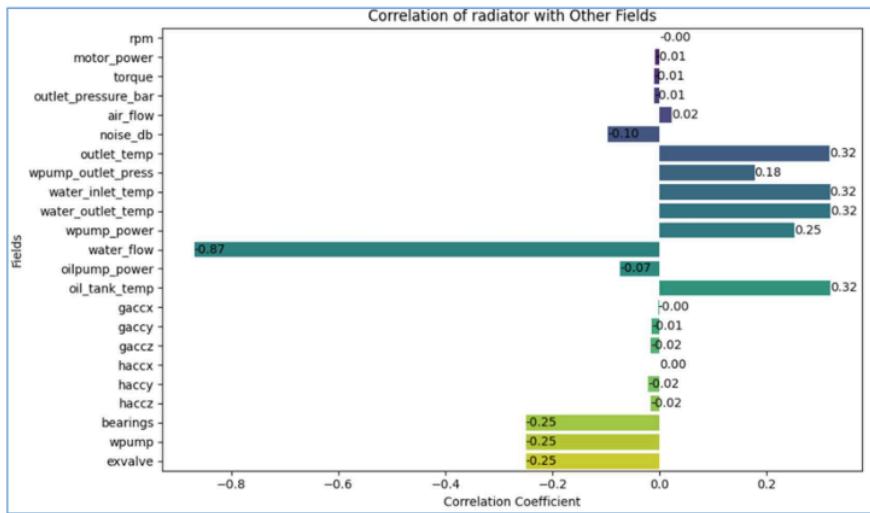


Figure 16 - Corelation Chart - Bearing Fault

We took a closer look at corelation of Bearing Fault as this is our dependent variable. This enabled us to determine which features play a significant role in influencing bearing conditions.

- CHAPTER 5: RESULTS & DISCUSSIONS

1.17.10. Introduction

This section presents a comprehensive analysis of the experimental results, offering detailed insights into the outcomes observed during the study. The findings discussed here serve as the foundation for drawing meaningful conclusions and formulating practical recommendations. By examining these results, we aim to validate our hypotheses, assess the effectiveness of the methodologies applied, and identify key patterns or implications that inform the final direction of the research.

1.17.11. Evaluation Results

Initial Model Creation:

Evaluation Metrics are captured during initial creation of model.

Evaluation Metrics				
Evaluation-Scores		Confusion Matrix		
	Scores	positive	negative	
Accuracy	91	positive	17	5
F1	74	negative	7	111
Precision	77			
Recall	71			
Kappa	69			

Figure 17 - Stage 1 - Evaluation Matrix

- **Accuracy (91%):** The model correctly classified 91% of all samples. This suggests strong overall performance, though it may not fully reflect effectiveness on minority classes if the dataset is imbalanced.
- **F1 Score (74%):** This score balances precision and recall, indicating the model maintains a reasonable trade-off between identifying true positives and minimizing false positives/negatives.

- **Precision (77%):** Of all instances predicted as positive (e.g., bearing faults), 77% were correct. This shows the model is fairly reliable in avoiding false alarms.
- **Recall (71%):** The model successfully identified 71% of actual positive cases. While decent, it suggests that nearly 29% of true positives were missed, which could be critical in fault detection scenarios.
- **Cohen's Kappa (69%):** This metric accounts for chance agreement between predicted and actual labels. A score of 0.69 indicates substantial agreement, though there's room for improvement.

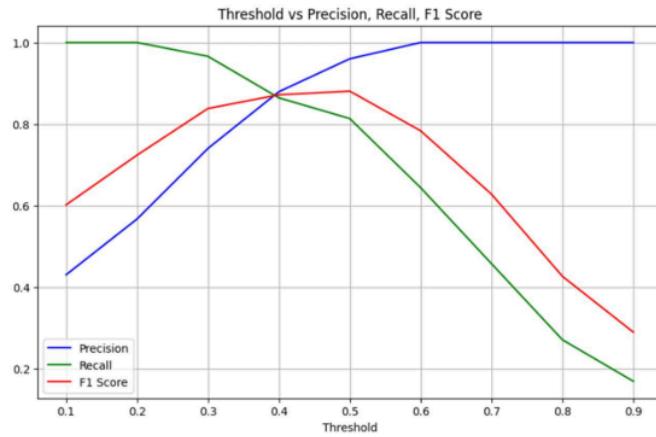


Figure 18 - Stage 1 - Precision, Recall & F1 Plot

The Adaptive Random Forest model is highly precise, which means it rarely raises false alarms. However, its recall is lower, so it may miss some actual faults, which could be risky depending on the actual implementation.

Detailed tuning can be implied to improve recall without sacrificing precision.

Live Data Streaming

Evaluation Metrix are captured and verified for every batch of live data coming from air compressor sensors.

Batch 1

Evaluation Metrics				
Evaluation-Scores		Confusion Matrix		
	Scores	positive	negative	
Accuracy	92	positive	32	7
F1	76	negative	13	188
Precision	82			
Recall	71			
Kappa	71			

Figure 19 - Stage2 - Evaluation Metrix Batch1

Batch 2

Evaluation Metrics				
Evaluation-Scores		Confusion Matrix		
	Scores	positive	negative	
Accuracy	92	positive	45	7
F1	77	negative	20	268
Precision	87			
Recall	69			
Kappa	72			

Figure 20 - Stage2 - Evaluation Metrix Batch2

Batch3

Evaluation Metrics				
Evaluation-Scores		Confusion Matrix		
	Scores	positive	negative	
Accuracy	93	positive	60	9
F1	79	negative	22	349
Precision	87			
Recall	73			
Kappa	75			

Figure 21 - Stage2 - Evaluation Metrix Batch3

Batch4

Evaluation Metrics				
Evaluation-Scores		Confusion Matrix		
	Scores	positive	negative	
Accuracy	94	positive	81	9
F1	83	negative	25	425
Precision	90			
Recall	76			
Kappa	79			

Figure 22- Stage2 - Evaluation Metrix Batch4

The steady improvement in performance metrics indicates that the model is effectively adjusting to live data streams. It's important to maintain ongoing surveillance for concept drift, with particular attention to changes in recall, to ensure sustained accuracy over time.

Although the model demonstrates strong precision, recall remains under 80%, indicating some fault cases are still being missed. To enhance fault detection while maintaining high precision, it may be beneficial to explore threshold adjustments or implement ensemble techniques.

Considering the positive performance trend, regularly retraining the model with the latest batch data appears to be an effective strategy. Continuing this retraining cycle will help sustain and potentially enhance model accuracy over time.

Considering the positive trajectory in performance, regularly updating the model with data from recent batches appears to be a beneficial approach for maintaining and enhancing its effectiveness.

1.17.12. Interpretation of Visualizations

Aim for this experiment was to create an Optimized UI which can we be used by maintenance Engineer to monitor condition of machine.

Conditioning Monitoring – In this experiment we have used simple plots and added to application to view trends of critical parameters, it is evident that such plots are easy to implement and manage at the same time enables maintenance engineers to proactively identify equipment anomalies at an early stage, helping to prevent unexpected breakdowns and minimize operational disruptions. By continuously tracking the health of machinery, it allows for more strategic and timely maintenance planning, which enhances system reliability, ensures workplace safety, and reduces overall maintenance costs.

Such functionality can be easily scaled to multiple features and visualization parameters

Anomaly Detection: Predicted values plotted on chart as datapoints where anomalies are annotated and shown in different colour is highly useful for any maintenance engineer for technical analysis and decision making.

This aids users in identifying root causes more efficiently, as visually highlighted anomalies often align with specific events—such as equipment malfunctions or sensor inconsistencies—making it easier to trace and diagnose issues quickly.

Timeseries Forecasting:

In this experiment, we implemented an LSTM model to perform time series forecasting on air compressor sensor data. The results were highly encouraging, demonstrating the model's capability to capture temporal patterns and predict future equipment behavior with accuracy. These findings support the feasibility of deploying this approach on an actual manufacturing shop floor for real-time monitoring and decision-making. Moreover, the experiment reinforces the value of time series forecasting in predictive maintenance—enabling early fault detection, anticipating performance shifts, and driving operational efficiency through timely, data-driven interventions.

1.17.13. Summary

The experiment yielded highly encouraging results using the Adaptive Random Forest model, demonstrating its strong potential for handling dynamic, real-time data streams effectively. The model's performance across multiple batches showed consistent accuracy and adaptability, making it a promising candidate for industrial applications. These findings open up a significant opportunity to conduct more comprehensive experiments using actual air compressor sensor data. By applying the model to real-world conditions, researchers and engineers can validate its robustness, fine-tune its predictive capabilities, and potentially integrate it into live monitoring systems for predictive maintenance and fault detection.

- **CHAPTER 6: CONCLUSIONS & RECOMMENDATIONS**

1.17.14. Introduction

In this section we will be concluding our experiment and will discuss about the contribution to domain knowledge and future recommendations.

1.17.15. Discussion and Conclusion

This experiment shows how combining machine learning with modern data collection tools and easy-to-use visualizations can improve predictive maintenance. It uses Adaptive Random Forest algorithms to detect faults in bearing systems by analyzing live data streams in real time. It also explores how LSTM models can be used for time series forecasting to monitor the health of industrial air compressors. By using multiple sensor inputs—like temperature, airflow, RPM, and vibration—the system can continuously track performance, spot unusual behavior, and detect early signs of problems. This contributes to smarter, data-driven maintenance and helps reduce downtime and improve efficiency.

1.17.16. Future Recommendation

Further Experiment with Air Compressors:

Next step should be to conduct this experiment on a physical Industrial Air Compressor where as a proof-of-concept visualization can we used by maintenance engineers to provide feedback on models' performance.

Based on fruitfulness of further experiment , we can integrate Alert system with this system where an alert in terms of email or SMS can be send to user directly.

This experiment can also be expanded to dynamically assign new labels within the system by tagging specific timeframes when physical failures occur. This approach enables the model to continuously learn from real-world events and improve its ability to predict similar failures in the future, enhancing its adaptability and long-term accuracy.

Optimization and Tuning of Model.

Experiment can be further continued monitoring the concept drift and tuning threshold and other parameters of model to maintain accuracy and improve recall.

- **CHAPTER 7: RESOURCE REQUIREMENTS**

1.17.17. Hardware Requirements

Following are the Hardware requirements for this research activity:

- A laptop or Desktop with good internet connection & all required software & libraries

1.17.18. Software Requirements

Following are the software requirements for this research activity:

- Web Browser like Edge & Chrome
- Integrated Development Environment – Google Colab
- Integrated Development Environment – Visual Studio Code 1.103.2
- Latest version of Python - 3.13.6.
- Deep Learning libraries such as Scikit-learn, river,statsmodels
- NiceGUI Python UI Framework
- Python Libraries like - Numpy Pandas, Matplotlib, Seaborn etc.

- **CHAPTER 8: DATA MANAGEMENT PLAN**

All Data required for this research will be collected and stored in most ethical manner.

Artifact	Collection	Storage
Dataset (csv format)	We will be using an Opensource dataset for this research paper	github
Py Files	All Py artifacts and its contents will be a developed my me and without any plagiarism	github
Research Proposal	All artifacts and its contents will be a developed my me and without any plagiarism	github
Research Paper	All artifacts and its contents will be a developed my me and without any plagiarism	github

Table 14 - Data Management Plan

• REFERENCES

Ahmad Aminzadeh,Sasan Sattarpanah Karganroudi,Soheil Majidi (2024) A Machine Learning Implementation to Predictive Maintenance and Monitoring of Industrial Compressors.[online] Available at: [A Machine Learning Implementation to Predictive Maintenance and Monitoring of Industrial Compressors](#) [Accessed 2nd May. 2025].

Abhay Nambiar,Naveen Venkatesh S,Aravindh S.... (2024) [online] Prediction of air compressor faults with feature fusion and machine learning. Available at: [Prediction of air compressor faults with feature fusion and machine learning - ScienceDirect](#)[Accessed 15th May. 2025]

Akshita Chugh (2020) MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better? [online Image] Available at: [MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better? | by Akshita Chugh | Analytics Vidhya | Medium](#) [Accessed 10th May. 2025]

Daniel A, Bruno P., Carlos F (2025) [online]. Predictive Maintenance and Smart Sensors Aiming Sustainability: A Perspective From a Bibliometric Analysis. Available at: [\(PDF\) Predictive Maintenance and Smart Sensors Aiming Sustainability: A Perspective From a Bibliometric Analysis](#) [Accessed 15th May. 2025]

David Valdivieso López, Ignacio,Aguilera-Martos,.. (2023) [online]. Fusing anomaly detection with false positive mitigation methodology for predictive maintenance under multivariate time series. Available at: [Fusing anomaly detection with false positive mitigation methodology for predictive maintenance under multivariate time series \(2023\) | David Valdivieso López | 6 Citations](#) [Accessed 20th May. 2025]

Engineer Waqar (2020) Components of Rotary Screw Compressor [online] Available at: [What is a Screw Compressor | How does a Screw Compressor Work?](#) [Accessed 5th May. 2025]

Falsk Raja (2023) [online] AI for Predictive Maintenance in Industrial Systems. Available at: [\(PDF\) AI for Predictive Maintenance in Industrial Systems](#) [Accessed 20th May. 2025]

George Dimitrakopoulos; Pal Varga; Thomas Gutt...(2024) [online]. Industry 5.0: Research Areas and Challenges With Artificial Intelligence and Human Acceptance. . Available at: [Industry 5.0: Research Areas and Challenges With Artificial Intelligence and Human Acceptance | IEEE Journals & Magazine | IEEE Xplore](#) [Accessed 20th May. 2025]

João Barataa, Ina Kayserb (2023)) [online] Industry 5.0 – Past, Present, and Near Future. . Available at: [Industry 5.0 – Past, Present, and Near Future - ScienceDirect](#) [Accessed 15th May. 2025]

² Lei Y., Yang B., Jiang X., Jia F... (2020) [online] Applications of machine learning to machine fault diagnosis: A review and roadmap. Available at: [Applications of machine learning to machine fault diagnosis: A review and roadmap - ScienceDirect](#) [Accessed 220th May. 2025]

Mechstudies (2021) screw air compressor schematics [image online]. Available at: [screw-air-compressor-schematics.jpg \(720×400\)](#) [Accessed 5th May. 2025]

Ms.RM.Suganya, Pavithra, A.V.R Pooja. (2024) Predictive Maintenance for Industrial Equipment Using Random Forest Regressor [online] Available at: [IJSDR2404173.pdf](#) [Accessed 2nd May. 2025].

M. Sheik Syed Sulaiman (2023) [online] A Study of Random Forest Machine Learning based Predictive Maintenance in Industrial Maintenance. Available at: [a-study-of-random-forest-machine-learning-based-predictive-maintenance-in-industrial-maintenance-IJERTV12IS100042.pdf](#) [Accessed 2nd May. 2025]

Mounia Achouch,Mariya Dimitrova,Rizck Dhouib (2023) Predictive Maintenance and Fault Monitoring Enabled by Machine Learning: Experimental Analysis of a TA-48 Multistage Centrifugal Plant Compressor [online] Available at: [Predictive Maintenance and Fault Monitoring Enabled by Machine Learning: Experimental Analysis of a TA-48 Multistage Centrifugal Plant Compressor](#) [Accessed 2nd May. 2025]

Nikhil Technochem (2024) Screw Compressors [online image] Available from: [Importance of air compressors in industrial uses and types](#) [Accessed 27th May. 2025]

Nawal Daoudi, Smail Zaki, Mohamed Aboussaleh (2024) Predictive Maintenance System for Screw Compressors Using Machine Learning: A Comparative Study. [online] Available at: [\(PDF\) Predictive Maintenance System for Screw Compressors Using Machine Learning: A Comparative Study](#) [Accessed 2nd May. 2025].

Neuraldesigner (2024) Air Compressor Dataset [online] Available at: <https://www.neuraldesigner.com/wp-content/uploads/2023/10/aircompressor.csv> [Accessed 5th May. 2025]

Peter Poor, David Zenisek and Josef Basi (2019) Historical Overview of Maintenance Management Strategies: Development from Breakdown Maintenance to Predictive Maintenance in Accordance with Four Industrial Revolutions. [online] Available at: [\(PDF\) Historical Overview of Maintenance Management Strategies: Development from Breakdown Maintenance to Predictive Maintenance in Accordance with Four Industrial Revolutions](#) [Accessed 10th June. 2025].

Parthajit Bisal and Prithwiraj Jana (2025) Compressor health monitoring and predictive maintenance by anomaly detection using fuzzy logic and random forest model by machine

⁴ learning process through Matlab. [online] Available at: [Compressor health monitoring and predictive maintenance by anomaly detection using fuzzy logic and random forest model by machine learning process through Matlab](#) [Accessed 2nd May. 2025].

Punam J. Patil, Pankaj H. Zope (2025) [online]. Predictive Maintenance: Evolution, Purposes and Approaches for Industrial Systems - A Comprehensive Review. Available at: [Predictive-Maintenance-Evolution-Purposes-and-Approaches-for-Industrial-Systems-A-Comprehensive-Review.pdf](#) [Accessed 20th May. 2025]

⁵ Pooja Kamat and Rekha Sugandhi Zope (2024) [online]. Anomaly Detection for Predictive Maintenance in Industry 4.0-A survey. Available at: [\(PDF\) Anomaly Detection for Predictive Maintenance in Industry 4.0-A survey](#) [Accessed 20th May. 2025]

Philip Stahmann, Maximilian Nebel, Christian Janiesch (2025) [online] AI-based real-time anomaly detection in industrial engineering: A structured literature review, taxonomy, and research agenda. Available at: [AI-based real-time anomaly detection in industrial engineering: A structured literature review, taxonomy, and research agenda - ScienceDirect](#) [Accessed 20th May. 2025]

S. Arena,E. Florian,I. Zennaro (2022) [online] A novel decision support system for managing predictive maintenance strategies based on machine learning approaches. Available at: [A novel decision support system for managing predictive maintenance strategies based on machine learning approaches - ScienceDirect](#) [Accessed 20th May. 2025]

Shivansh Sabhadiya (2021) What is an Air Compressor? [online] Available at: [What Is Air Compressor?- Definition, Types And Working](#) [Accessed 5th May. 2025]

YUVRAJ JIVAN JADHAVI , DR. BHAGY(2024) Real-Time Anomaly Detection in Air Compressors Using Machine Learning [online] Available at: [IJIRT168680_PAPER.pdf](#) [Accessed 2nd May. 2025].

Doan Ngoc,Chi Nam, Chan Hian Leng and others (2024) A Method to Handle Concept Drift in Predicting Remaining Useful Life. [online] Available at: [A Method to Handle Concept Drift in Predicting Remaining Useful Life | SpringerLink](#) [Accessed 22nd August. 2025].

Jan Zenisek, Florian Holzinger and Michael Affenzeller (2019) Machine learning based concept drift detection for predictive maintenance Available at: [Machine learning based concept drift detection for predictive maintenance - ScienceDirect](#) [Accessed 22nd August. 2025].

FTMaintanance (2019) History of Maintenance: The Evolution of Industrial & Facility Maintenance Practices [online] Available at: [History of Maintenance: The Evolution of Industrial & Facility Maintenance Practices | FTMaintenance CMMS](#) [Accessed 5th May. 2025]

US Department of Energy(2010) Operations & Maintenance Best Practices [online] Available at: [Operations & Maintenance Best Practices Guide: Release 3.0](#) [Accessed 5th May. 2025]

Gian Antonio Susto Andrea Schirru Simone Pampuri and others(2015) Machine Learning for Predictive Maintenance: A Multiple Classifier Approach [online] Available at: [\(PDF\) Machine Learning for Predictive Maintenance: A Multiple Classifier Approach](#) [Accessed 5th May. 2025]

Nandini Nalawade Swati D. Jakkan and others(2024) Predictive Maintenance for Industrial Equipment using IIOT, AI and ML [online] Available at: [Predictive Maintenance for Industrial Equipment using IIOT, AI and ML](#) [Accessed 20th May. 2025]

Michael Nsor (2024) Predictive Maintenance Using Machine Learning for Engineering Systems Through Real-Time Sensor Data and Anomaly Detection Models. [online] Available at: [\(PDF\) Predictive Maintenance Using Machine Learning for Engineering Systems Through Real-Time Sensor Data and Anomaly Detection Models](#) [Accessed 20th May. 2025]

Sai Takawalea , Nutan Mehetreb , Neha Fadtarec , Swati Bagalid (2025) AI-Powered Predictive Maintenance Using Machine Learning for Industrial IoT Systems [online] Available at: [AI-Powered Predictive Maintenance Using Machine Learning for Industrial IoT Systems](#) [Accessed 20th Sept. 2025]

Mohammed Amine Moustakim1,Tamou Nasser1,Najib Elkamoun and Others(2025) Prediction of electric power and load forecasting using LSTM technique for EMS [online] Available at: [Prediction of electric power and load forecasting using LSTM technique for EMS](#)

• APPENDIX A : RESEARCH PLAN

1.17.19. Introduction

1.17.20. Research Project Plan

Research Project plan from topic selection to final completion is planned and documents as Project plan/schedule with Gant chart

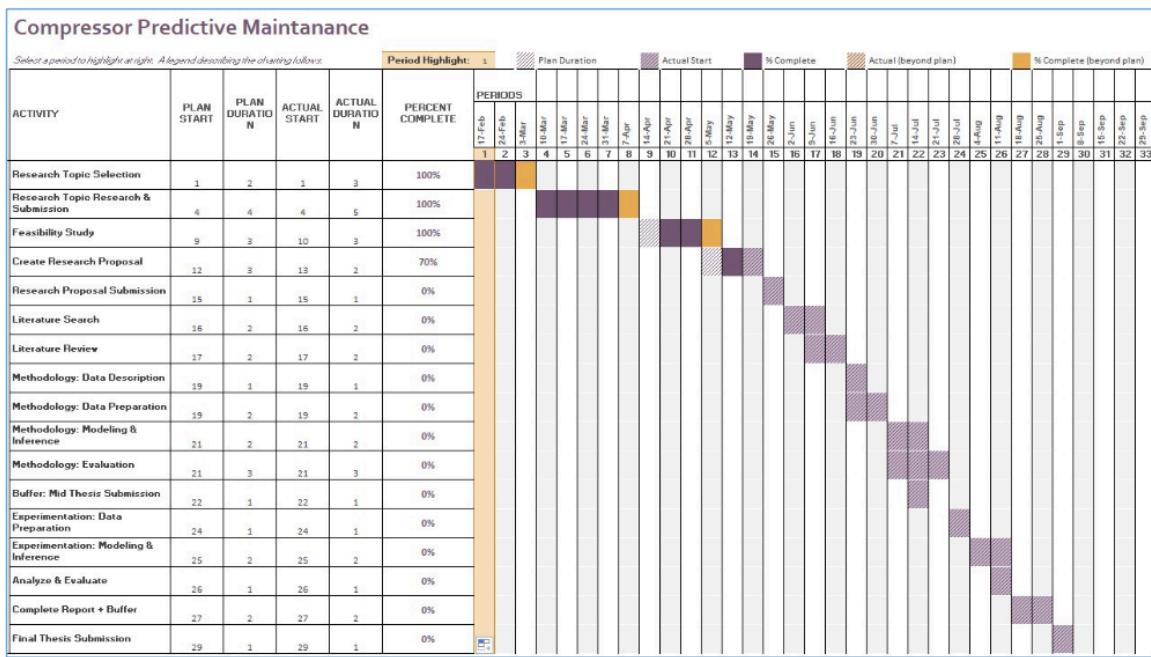


Figure 23 - Project Plan Gant Chart

1.17.21. Risk Mitigation and Contingency Plan

Risk Assessment is conducted and risks and their mitigation is added in below table (Table 4)

Risk	Mitigation
Any unplanned personal emergency hampering dedicated time for research	Have enough contingency in research plan to take care any such risks,

	Inform institution and take extension
Unavailability of required hardware or software	Use allopen-source software & hardware

Table 15 - Risks & Contingency

- APPENDIX B: RESEARCH PROPOSAL

- **APPENDIX B: ETHICS FORMS**

There is no Conflict of Interest to report.

Sarat Shankar_ResearchProposal_FinalReport_Compressor-Predictive Maintanance_V01-1.docx

ORIGINALITY REPORT



PRIMARY SOURCES

- | | | |
|---|---|----|
| 1 | ieomsociety.org
Internet Source | 2% |
| 2 | bibbase.org
Internet Source | 1% |
| 3 | Submitted to Liverpool John Moores University
Student Paper | 1% |
| 4 | www.mechanicaljournals.com
Internet Source | 1% |
| 5 | ijrpr.com
Internet Source | 1% |
| 6 | www.ijaer.com
Internet Source | 1% |
-

Exclude quotes Off
Exclude bibliography On

Exclude matches < 1%