Predictive Maintenance for

Single Stage Air Compressors Using Advanced Machine Learning

Sarat Shankar

Student ID**: 1175510**

[Sharat.sk04@gmail.com](mailto:Sharat.sk04@gmail.com), +919916200532

Research Interim Report for

Masters in Data Science

Liverpool John Moores University & upGrad

July 2025

**TABLE OF CONTENT**

[Abstract 3](#_Toc203199885)

[List of Tables 4](#_Toc203199886)

[List of Figures 4](#_Toc203199887)

[List of Abbreviations 4](#_Toc203199888)

[1. Introduction 5](#_Toc203199889)

[1.1. Background of the Study 5](#_Toc203199890)

[1.2. Problem Statement 7](#_Toc203199891)

[1.3. Aim and Objectives 8](#_Toc203199892)

[1.4. Research Questions 8](#_Toc203199893)

[1.5. Scope of the Study 8](#_Toc203199894)

[1.6. Significance of the Study 8](#_Toc203199895)

[1.7. Structure of the Study 9](#_Toc203199896)

[2. Related Research 10](#_Toc203199897)

[2.1. Introduction 10](#_Toc203199898)

[2.2. Types of Maintenance 10](#_Toc203199899)

[2.3. Technology & Predictive Maintenance 10](#_Toc203199900)

[2.4. Summary of Related Research Work: 11](#_Toc203199901)

[2.5. Findings from the Review: 13](#_Toc203199902)

[3. Research Methodology 14](#_Toc203199903)

[3.1.1. Gather Dataset 14](#_Toc203199904)

[3.1.2. EDA (Data Processing & Feature Selection) 17](#_Toc203199905)

[3.1.2.1. Data Processing: 17](#_Toc203199906)

[3.1.2.2. Data Analysis: 17](#_Toc203199907)

[3.1.2.3. Feature Selection: 18](#_Toc203199908)

[3.1.3. Model Selection 18](#_Toc203199909)

[3.1.4. Model Evaluation 19](#_Toc203199910)

[3.1.5. Incremental Training Concept 20](#_Toc203199911)

[3.1.6. Visualization & Monitoring 21](#_Toc203199912)

[4. Resource Requirements 22](#_Toc203199913)

[4.1. Hardware Requirements 22](#_Toc203199914)

[4.2. Software Requirements 22](#_Toc203199915)

[5. Data Management Plan 22](#_Toc203199916)

[REFRENCES 23](#_Toc203199917)

[APPEDIX A : RESEARCH PLAN 25](#_Toc203199918)

[Research Project Plan 25](#_Toc203199919)

[Risk Mitigation and Contingency Plan 25](#_Toc203199920)

# Abstract

Predictive Maintenance using ML algorithms has become a very critical operation in industry where vast amounts of data generated by different equipment are analyzed to predict potential failures. By leveraging historical and real-time data, along with the power of Machine Learning, this approach enables maintenance activities to be precisely timed, ensuring equipment is serviced only when necessary.

This research paper presents a review on Predictive Maintenance and anomaly detection of an industrial Single stage screw compressor using ML algorithms to analyze historical operational data and predict failures and monitor and visualize real-time equipment data to detect anomalies for an industrial Single Stage Compressor. The study will focus on having an end-to-end solution that can be scaled and implemented across the shop floor. Predictive Maintenance of Compressor prevents unplanned downtime, prevents costly capital expenditure, and improves the overall equipment efficiency.

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.

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

In this research we will predict key parameters like temperature, airflow etc. using regression Random Forest models. Anomalies will be detected and flagged based on their deviations between the actual values and predicted values, which will be visualized over the simple user interface.

The results of research will help us prove that advanced machine learning techniques can be applied to single stage twin screw 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 4](#_Toc203121183)

[Table 2 - List of Dataset Parameters 14](#_Toc203121184)

[Table 3 - Data Management Plan 20](#_Toc203121185)

[Table 4 - Risks & Contingency 23](#_Toc203121186)

# List of Figures

[Figure 1 - Siingle Stage Screw Compressor Working Diagram 5](#_Toc203121298)

[Figure 2 - Single Stage Screw Compressor Block Diagram 6](#_Toc203121299)

[Figure 3 - Architecture Diagram for proposed system 12](#_Toc203121300)

[Figure 4 - Architecture diagram on Proposed Research Structure 12](#_Toc203121301)

[Figure 5 - Mean Squared Error Formula 17](#_Toc203121302)

[Figure 6 - Root mean Squared Error Formula 17](#_Toc203121303)

[Figure 7 - Mean Absolute Error Formula 17](#_Toc203121304)

[Figure 8 - R Squared Formula 17](#_Toc203121305)

[Figure 9 - Incremental Training Concept 18](#_Toc203121306)

[Figure 10 - Project Plan Gant Chart 23](#_Toc203121307)

# 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

# Introduction

# 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.

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](https://www.theengineeringchoice.com/author/shivansh/),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.

A diagram of a mechanical part

AI-generated content may be incorrect.

Figure 1 - Siingle Stage Screw Compressor Working Diagram

Above figure (Figure 1) shows the working principal of screw 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)

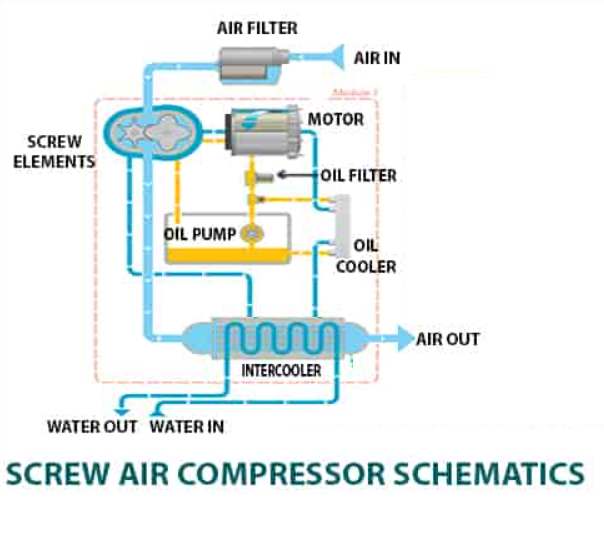


Figure 2 - Single Stage Screw Compressor Block Diagram

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

* **Air In**

Air is taken into the 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 compressor.

* **Screw Element**

Air inlet in compressor compresses the air. There are 2 rotors/screws in Screw 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 compressor, this is used to power and rotate the male and female screw elements.

* **Oil Filters**
* In compressor one of the important operations is lubrication, there are multiple types of lubricant used in 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 form screw element.

* **Bearings**

Bearing is another mechanical element in compressor, these are ani 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 reduces the losses caused by friction.
* It helps free rotation of Screw in compressor.

# 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

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

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.

# Aim and Objectives

**Aim:** of this research is topredict key parameters and detect anomalies for Single Stage Screw Air Compressor using advanced machine learning algorithm

**Objectives:**

* To conduct a comprehensive review of available literature regarding Anomaly Detection and Predictive Maintenance in Industrial Compressors.
* To predict critical features of Industrial Single stage Screw Compressor for predictive Maintenance.
* 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.

# Research Questions

This thesis tries to answer the following questions:

* Can a Random Forest Classifier be effective in predicting key parameters of an Air compressor?
* How can model help with real-time monitoring of 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 parameter will be helpful in fault detection
* Can the research conducted on the Compressor be easily scaled to other industrial equipment on a plant floor (theoretical analysis)?

# 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](#_Research_Plan).
* Open-Source models and software will be used to conduct all types of experiments.
* We will be using Google-Colab which is a publically available GPU for all experimentations.
* We will be only using automated metrics for evaluation of models.
* 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.
* This research will be using open-source dataset (refer next section) and may not be having all features which is generally available from compressor.

# Significance of the Study

In this study I am trying to address issue of unplanned downtime for industrial single stage air screw 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 shows 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 manufacturers 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 compressor

# Structure of the Study

The remainder of the paper is structures in this way:

**Section2:** Provides brief background on the problem statement and related Literature Reviews and research work carried out the area of Predictive Maintenance and Anomaly Detection.

**Section3:** 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,

**Section 4:** This section details resource utilized for this research including hardware, software and others.

**Section 5:** This section covers Data Management Plan and describes how data required for this research are collected and stored in most ethical manner.

**REFRENCES**: This section provides citations and references

**Appendix A:** This section details resource utilized for this research including hardware, software and others.

# Related Research

# Introduction

Even though Predictive maintenance is not a new field and there is a lot of work going on in this field from many years, by with technological advancements has boasted this area with very high potential and unexplored opportunities where system 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.

# Types of Maintenance

**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

**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

**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.

# Technology & Predictive Maintenance

Authors (João Barataa and Ina Kayserb,2023) talks about Industry 5.0 as a vison 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 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.

Pratanjit and Prithwiraj talk about the importance of anomaly detection mechanisms in real-time in research (Parthajit Bisal and Prithwiraj Jana,2025), In this paper author also talks about an interesting way to maintain the health status of a compressor using a scoring mechanism.

Authors (Yuvraj Jivan Jadavi and Dr. Bhagya,2024) mention similar approaches in his papers and talks 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 point 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

# Summary of Related Research Work:

|  |  |  |  |
| --- | --- | --- | --- |
| **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, and research agenda | (Philip Stahmann Et al,2025) | Structured literature review, systematically decomposed implementation options of  real-time anomaly detection | -- |
| 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. |
| 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 |
| 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 Prithwiraj Jana,2025), | Maintain the health status of a 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 Industrial Compressors | (Ahmad Et al, 2024) | Focuses on comparison of computational needs vs performance | Simple Linear Regression |
| 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 Maintanance | 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 |

# Findings from the Review:

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

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.

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.

# Research Methodology

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

A diagram of a model building

AI-generated content may be incorrect.

Figure 3 - Architecture Diagram for proposed system

Here, for this research, we are employing multiple Random Forest Classification models to focus on detecting anomaly using classic algorithms.

A Random Forest classifier is trained on the pre-processed dataset to detect anomaly is from the batch of Live data accumulated from Sensors (In this research we will reuse available dataset)

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

A diagram of a computer component

AI-generated content may be incorrect.

Figure 4 - Architecture diagram on Proposed Research Structure

# Gather Dataset

Data for any industrial compressor is collected from multiple sensors available with the compressor unit. These parameters are captured and stored in a time series fashion. For this research we will be using open-source data available in .csv for an industrial Screw Compressor

Dataset used is from (Neuraldesigner,2024) available at [https://www.neuraldesigner.com/wp-content/uploads/2023/10/aircompressor.csv](https://www.neuraldesigner.com/wp-content/uploads/2023/10/aircompressor.csv%20%20%20%20%20)

Data Preparation techniques like handling missing values, Noise Reduction, Outlier detection, Feature Scaling etc., will be used.

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.

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 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 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 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). | In a 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 compressor housing |
| Water Pump, Inlet Pressure | Numerical | bar | Water Pump’s Inlet pressure. |
| Water Pump, Inlet Pressure | Numerical | bar | Water Pump’s Outlet pressure. |
| Vibration Acceleration | Numerical | g (acceleration due to gravity) | 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 |
| 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 radiatoe 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 Failurewhere “Not Stable” means AC Motor fault and “Stable” means no fault |

Table 2 - List of Dataset Parameters

# EDA (Data Processing & Feature Selection)

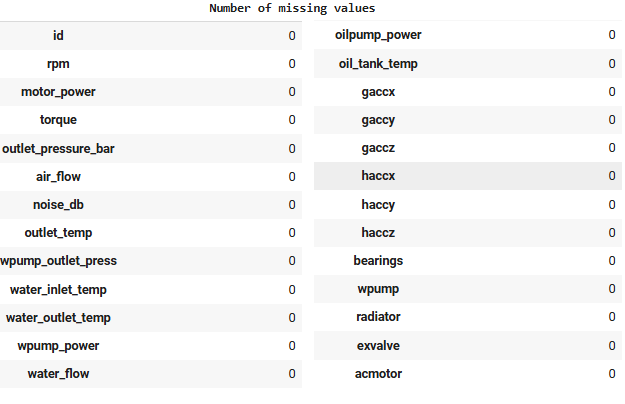
# Data Processing:

Every dataset entry represents data collected at a defined period or when any major event occurs in compressor, these dataset with multiple parameters gathered at against time provide a comprehensive view of operational and health status of compressor over time.

The following steps were taken to clean, transform, and prepare data before performing deeper analysis

* **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 | This has a single label for all records, not helpful  in classification |

* + There are no missing values.
  + 
* **Data Transformation**- Normalization/Standardization of data is done for the required parameters. We will create new derived parameters like differential air pressure.
  + **Timestamp is added to each record of dataset indicating the data collected in a defined frequency**. This will be used in the time series analysis of key features.
* **Treat Outliers**: All outliers are treated, unusual values are identified and treated that may skew analysis and take appropriate action to handle outliers.

# Data Analysis:

In this phase, 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.

* + **Multivariate Analysis**: Examining interactions among three or more variables. The tools used here will be Pair plots, heatmaps etc.
  + **Correlation Analysis**: Measures the strength and direction of relationships between numeric variables.
  + **Trend and Pattern Detection**: Time series plots, moving averages, seasonality analysis (for time-based data)

# 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.

The feature Selection process will be conducted to identify the most relevant predictors for the model. We are using Principal Component Analysis (PCA)to reduce the number of features in a dataset while keeping the most important information

# Model Selection

Based on research and Literature Review, we are employing multiple Random Forest classification models to predict failures and detect anomalies

In case of anomaly detection RF model can detect anomalies by identification of outliers and anomalies by evaluation of datapoints on individual decision trees classification

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

**Classification Model: Random Forest**

Random Forest is one of the most popular machine learning algorithms used for Predictive maintenance, it is an ensemble method that combines multiple trees to make predictions and solve complex problems and increase the performance of the model. Both classification and regression problems can be handled by random forest. RF extends bagging method as it utilizes both feature randomness and bagging to create an uncorrelated forest of decision trees.

Here we are using the concept of Incremental training concept with Random Forest classification model to periodically re-train and update an already deployed machine learning model with new/additional labelled data

For Predicting failures based on Anomaly Detection we will be using Random Forest Classification model.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Key benefits if RF is that with higher no of trees in forest leads to better accuracy and reduces the risk of overfitting.

**Time Series Forecasting Model: LSTM**

LSTM networks are an extension of recurrent neural networks ([RNNs](https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/)).

In this paper we are researching on simple approach for Univariate time series forecasting to forecast key parameters of compressors for Fault detection and analysis

Here single series of observations and a model is required to learn from the series of past observations to predict the next value in the sequence. For this purpose, we will be using Vanila LSTM Model

# Model Evaluation

The models will be evaluated using:

**MSE - Mean Squared Error:** It measures the average of the squares of the errors comparing the actual and the predicted values. Lower mean-squared Error values indicate better performance.

Formula for MSE in given below in figure (Figure 5)

A mathematical equation with numbers and symbols

AI-generated content may be incorrect.

Figure 6 - Mean Squared Error Formula

**RMSE - Root Mean Squared Error** is calculated as the square root of MSE and provides an error metric in the same units as the target variable, making it easier to interpret. Lower Root Mean Squared Error values reflect better predictive accuracy.

Formula for RMSE in given below in figure (Figure 5)

A math equations with numbers and symbols

AI-generated content may be incorrect.

Figure 7 - Root mean Squared Error Formula

**MAE - Mean Absolute Error**: It is statistical measure which represents the average absolute difference between the predicted and the actual values it quantifies how much, on average, the predicted values deviate from the true values

Formula for MAE in given below in figure (Figure 5)

A mathematical equation with numbers and symbols

AI-generated content may be incorrect.

Figure 8 - Mean Absolute Error Formula

**R²- R Squared**: It is statistical measure which calculates the proportion of variance in the dependent variable that can be explained by the independent variables. It shows how well model fits the data and a value which is closer to 1 fits better.

Formula for R Squared in given below in figure (Figure 5)

A mathematical equation with numbers and symbols

AI-generated content may be incorrect.

Figure 9 - R Squared Formula

All above figures in this section is taken for reference from ([Akshita Chugh](https://medium.com/@akshita-chugh024?source=post_page---byline--cd0326a5697e---------------------------------------),2020)

# Incremental Training Concept

Here Incremental training concept with Random Forest refers to periodically re-train and update an already deployed machine learning model with new/additional labelled data , this retraining will create new RF model which includes new labelled dataset., this will enhance performance of model in comparison to of-line RF model.

When the model should be re-trained depends on many factors, in many cases Data scientists might periodically retrain models (e.g., weekly, monthly)m in others when monitoring systems detect a significant drop in performance.

* We will be generating initial RF model based on the available historical dataset , we will call this model as Base Model.
* All Live dataset received from sensors(in this research we will reuse available dataset) will be passed to model to classify/predict any anomaly in received set of data.
* A minimum batch size of (x) records is used for classification.
* All new data received from live data is stored and with option provided to label data.
* Feedback/label is provided to received live data within a defined period of time(t)..
* Periodically every week a new model is created including Historical + new labelled live dataset.
* Above steps are repeated for every week.

For this research paper, we will be using smaller time period to retrain model with new data.

A diagram of a process

AI-generated content may be incorrect.

Figure 10 - Incremental Training Concept

# Visualization & Monitoring

Anomaly Detection: The predicted values are compared against actual values, and anomalies are flagged if deviations exceed predefined thresholds

Also, this research will focus on the minimum required user interface to visualize & monitor critical parameters & anomalies.

We will be having the following as part of the visualization:

* **Condition Monitoring**: Visualization of real-time and historical data for analysis help prevent failures and downtime, following key parameters are plotted for analysis
  + - **Temperature** – Time series trend with Upper and Lower Control Limits.
    - **Air Flow** – Time series trend with Upper and Lower Control Limits.
    - **Compressor Efficiency** – Efficiency is calculated and plotted on time series chart *formula: outlet Air Pressure/Motor Power*
* **Univariate Forecasting**: Forecast of predicted values for key selected features, we are going to use LSTM models to univariate time series forecasting for the key parameters below:
  + - **Power Consumption**: historical Power consumption is plotted with a forecast of power consumption for next 1 week.
    - **Air Flow:** historical Power consumption is plotted with a forecast of airflow for next 1 week.
* **Anomaly Detection**: Anomaly detection visualization will list the classified anomalies predicted by Random Forest model predicted over timeseries plot. Types of anomalies supported by current data set.
  + - Bearing Fault
    - Water Pump Fault
    - Radiator Fault
    - Ex Valve Fault

# Resource Requirements

# Hardware Requirements

Following are the Hardware requirements for this research activity:

* A laptop or Desktop with good internet connection & all required softwares & libraries

# Software Requirements

Following are the software requirements for this research activity:

* Web Browser like Edge & Chrome
* Integrated Development Environment – Google Colab
* Latest version of Python.
* Deep Learning libraries such as Scikit-learn, PyTorch ,TensorFlow,
* Python Libraries like - Numpy Pandas, Matplotlib, Seaborn etc.

# 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 3 - Data Management Plan

# REFRENCES

1. 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](https://www.mdpi.com/1424-8220/25/4/1006) [Accessed 2nd May. 2025].
2. Abhay Nambiar,Naveen Venkatesh S,Aravinth 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](https://www.sciencedirect.com/science/article/pii/S0950705124011535#b3)[Accessed 15th May. 2025]
3. [Akshita Chugh](https://medium.com/@akshita-chugh024?source=post_page---byline--cd0326a5697e---------------------------------------) (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](https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e) [Accessed 10th May. 2025]
4. 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](https://www.researchgate.net/publication/388852338_Predictive_Maintenance_and_Smart_Sensors_Aiming_Sustainability_A_Perspective_From_a_Bibliometric_Analysis) [Accessed 15th May. 2025]
5. [David Valdivieso López,](https://scispace.com/authors/david-valdivieso-lopez-40c3k4k3q7) [Ignacio,Aguilera-Martos,](https://scispace.com/authors/ignacio-aguilera-martos-4p20bp3mho)[..](https://scispace.com/authors/marta-garcia-barzana-29xz32tjn1) (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](https://scispace.com/papers/fusing-anomaly-detection-with-false-positive-mitigation-2es3jzdlui) **[**Accessed 20th May. 2025]
6. Engineer Waqar (2020) Components of Rotary Screw Compressor [online] Available at: [What is a Screw Compressor | How does a Screw Compressor Work?](https://mechanicalboost.com/screw-compressor/) [Accessed 5th May. 2025]
7. Falsk Raja (2023) [online] AI for Predictive Maintenance in Industrial Systems. Available at: [(PDF) AI for Predictive Maintenance in Industrial Systems](https://www.researchgate.net/publication/375722960_AI_for_Predictive_Maintenance_in_Industrial_Systems) [Accessed 20th May. 2025]
8. [George Dimitrakopoulos](https://ieeexplore.ieee.org/author/37411162400); [Pal Varga](https://ieeexplore.ieee.org/author/37547907300); [Thomas Gutt](https://ieeexplore.ieee.org/author/37328742800)…(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](https://ieeexplore.ieee.org/document/10508399) [Accessed 20th May. 2025]
9. 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](https://www.sciencedirect.com/science/article/pii/S1877050923003605) [Accessed 15th May. 2025]
10. 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](https://www.sciencedirect.com/science/article/abs/pii/S0888327019308088) [Accessed 220th May. 2025]
11. Mechstudies (2021) screw air compressor schematics [image online]. Available at: [screw-air-compressor-schematics.jpg (720×400)](https://www.mechstudies.com/wp-content/uploads/2021/10/screw-air-compressor-schematics.jpg) [Accessed 5th May. 2025]
12. Ms.RM.Suganya, Pavithra, A.V.R Pooja. ( 2024) Predictive Maintenance for Industrial Equipment Using Random Forest Regressor [online] Available at: [IJSDR2404173.pdf](https://www.ijsdr.org/papers/IJSDR2404173.pdf) [Accessed 2nd May. 2025].
13. 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](https://www.ijert.org/research/a-study-of-random-forest-machine-learning-based-predictive-maintenance-in-industrial-maintenance-IJERTV12IS100042.pdf) [Accessed 2nd May. 2025]
14. 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](https://www.mdpi.com/2076-3417/13/3/1790#:~:text=This%20paper%20aimed%20to%20apply%20machine%20learning%20algorithms,approach%20through%20the%20adoption%20of%20Industry%204.0%20approaches.) [Accessed 2nd May. 2025]
15. Nikhil Technochem (2024) Screw Compressors [online image] Available from: [Importance of air compressors in industrial uses and types](https://www.nikhiltechnochem.com/importance-of-industrial-air-compressors/) [Accessed 27th May. 2025]
16. Nawal Daoudi, [Smail Zaki](https://www.researchgate.net/profile/Smail-Zaki?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19), [Mohamed Aboussaleh](https://www.researchgate.net/profile/Mohamed-Aboussaleh?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19) (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](https://www.researchgate.net/publication/385309992_Predictive_Maintenance_System_for_Screw_Compressors_Using_Machine_Learning_A_Comparative_Study) [Accessed 2nd May. 2025].
17. Neuraldesigner (2024) Air Compressor Dataset [online] Available at: [https://www.neuraldesigner.com/wp-content/uploads/2023/10/aircompressor.csv](https://www.neuraldesigner.com/wp-content/uploads/2023/10/aircompressor.csv%20%20%20%20%20)[Accessed 5th May. 2025]
18. Parthajit Bisal and Prithwiraj Jana (2025) Compressor health monitoring and predictive maintenance by anomaly detection using fuzzy logic and random forest model by machine 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](https://www.mechanicaljournals.com/ijmtme/article/48/5-2-13-707.pdf) [Accessed 2nd May. 2025].
19. 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](file:///C:\Users\sshankar\Downloads\Predictive-Maintenance-Evolution-Purposes-and-Approaches-for-Industrial-Systems-A-Comprehensive-Review.pdf) [Accessed 20th May. 2025]
20. 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](https://www.researchgate.net/publication/341735017_Anomaly_Detection_for_Predictive_Maintenance_in_Industry_40-A_survey) [Accessed 20th May. 2025]
21. 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](https://www.sciencedirect.com/science/article/pii/S0360835225003821) [Accessed 20th May. 2025]
22. 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](https://www.sciencedirect.com/science/article/abs/pii/S0925753521003726) [Accessed 20th May. 2025]
23. [Shivansh Sabhadiya](https://www.theengineeringchoice.com/author/shivansh/) (2021) What is an Air Compressor? [online] Available at: [What Is Air Compressor?- Definition, Types And Working](https://www.theengineeringchoice.com/what-is-air-compressor/) [Accessed 5th May. 2025]
24. YUVRAJ JIVAN JADHAV1 , DR. BHAGY( 2024) Real-Time Anomaly Detection in Air Compressors Using Machine Learning [online] Available at: [IJIRT168680\_PAPER.pdf](https://ijirt.org/publishedpaper/IJIRT168680_PAPER.pdf) [Accessed 2nd May. 2025].

# APPEDIX A : RESEARCH PLAN

## 

## Research Project Plan

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

A screenshot of a graph

AI-generated content may be incorrect.

Figure 11 - Project Plan Gant Chart

## 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 4 - Risks & Contingency