

**Industrial Training**  
**on**  
**PREDICTIVE MAINTENANCE OF THERMAL POWER**  
**PLANT EQUIPMENTS**  
**SUBMITTED BY**

KSHITIJ KARN	CSE- C
210905121	kshitijkarn2003@gmail.com
21	9022355938

Under the Guidance of:

Mr. Pranit Kumar  
AGM  
MSX & DTG  
BHEL



Department of Computer Science and Engineering

August 2024

Figure 1

# Contents

<b>Acknowledgements</b>	<b>2</b>
<b>1 BHEL</b>	<b>3</b>
1.1 History of BHEL . . . . .	3
1.2 Products . . . . .	4
1.2.1 Transmission . . . . .	4
1.2.2 Rail Transpoetation . . . . .	4
1.2.3 Defence and Aerospace . . . . .	4
1.2.4 Water Management . . . . .	5
1.2.5 E-mobility . . . . .	5
1.2.6 Oil and Gas . . . . .	6
1.2.7 Renewables . . . . .	6
1.2.8 Battery Energy Storage Solutions . . . . .	6
1.2.9 RESEARCH & DEVELOPMENT . . . . .	7
1.3 Power . . . . .	8
1.3.1 Nuclear Power . . . . .	8
1.3.2 Hydro Power . . . . .	9
1.3.3 Thermal Power . . . . .	10
1.4 Thermal Power Plant . . . . .	11
<b>2 Remote diagnostics</b>	<b>13</b>
2.1 Steps in remote diagnostics system . . . . .	13
2.2 Benefits . . . . .	14
<b>3 Predictive Maintenance</b>	<b>23</b>
3.1 Key Components and Processes . . . . .	23
3.2 Task and Data description . . . . .	24
3.3 Exploratory Analysis . . . . .	25
3.3.1 Product ID . . . . .	25
3.3.2 Resampling with SMOTE . . . . .	27
3.3.3 PCA and Correlation Heatmap . . . . .	28

3.3.4	Metric . . . . .	29
3.4	Binary Task . . . . .	30
3.4.1	Preliminaries . . . . .	30
3.4.2	Feature Selection . . . . .	32
3.4.3	Models . . . . .	33
3.5	Multi-Class Task . . . . .	34
3.5.1	Models . . . . .	34
3.6	Conclusions . . . . .	36
3.6.1	Project Objectives . . . . .	36
3.6.2	Data Preprocessing . . . . .	37
3.6.3	Model Development and Evaluation . . . . .	37
3.6.4	Binary Classification . . . . .	37
3.6.5	Multiclass Classification . . . . .	38
3.6.6	Feature Importance . . . . .	38
3.6.7	Conclusion . . . . .	38
<b>4</b>	<b>Environmental and Societal Impact</b>	<b>39</b>
4.1	Environmental Impact . . . . .	39
4.2	Societal Impact . . . . .	39
<b>5</b>	<b>References</b>	<b>41</b>
<b>6</b>	<b>NBA/IET Mapping</b>	<b>42</b>



## भारत हेवी इलेक्ट्रिकल्स लिमिटेड

Bharat Heavy Electricals Limited

(भारत सरकार का उपकाम / A Government of India Undertaking)

CIN: L74899DL1964GOI004281

(पावर सेक्टर उत्तरी क्षेत्र / Power Sector Northern Region)

(भैल सदन, नवम तला/ BHEL Sadan, 9<sup>th</sup> Floor)

प्लॉट नंबर 25, सेक्टर-16ए, नोएडा, यूपी-201301 / Plot No. 25, Sector-16A, Noida, UP-201301

BHEL:PSNR:HR: 23 : 08 : 03

Dated : 17 July, 2024

### TO WHOM SO EVER IT MAY CONCERN

It is certified that Mr. Kshitij Karn, student of B.Tech. (Computer Science & Engineering, Manipal Institute of Technology, MAHE) has successfully completed his Summer training at BHEL PS-NR, Noida under the guidance of Mr Pranit Kumar, AGM, MSX & DTG on the topic of "**PREDICTIVE MAINTENANCE OF THERMAL POWER PLANT EQUIPMENT**" from 16<sup>th</sup> May to 16<sup>th</sup> July 2024 in partial fulfilment of his academic course.

We wish him success in all future endeavours.

(Monideepa Roy)  
Sr. Manager (HR)

पंजीकृत कागद : बीएचईएल हाउस, सिरी फोटो, नई दिल्ली - 110049 | फोन : 011-66337598 | ईमेल : contactus@bhel.in

Registered Office: BHEL HOUSE, Siri Fort, New Delhi - 110049 | Phone: 011-66337598 | E-mail: contactus@bhel.in  
[www.bhel.com](http://www.bhel.com) BHELOfficial BHEL\_India BHEL\_India bhel.india company/bhei

Figure 2: ITR-Certificate

# Acknowledgments

I would like to express my sincere gratitude and appreciation to the Training and Placement cell of MIT Manipal for providing me with this great opportunity to work as an intern at BHEL, a leading engineering and manufacturing enterprise. This internship was a valuable and rewarding experience that helped me gain practical skills and insights.

I would also like to thank my mentor Mr. Pranit Kumar, who guided and supported me throughout my internship. He was very helpful and generous in sharing his knowledge and expertise with me. He also gave me constructive feedback and suggestions on how to improve my work and skills. He was a great mentor and leader for me and the team.

I am grateful to BHEL Noida, for giving me this opportunity to work on such an interesting and impactful project. I hope that my work will contribute to the company's growth and success. I learned a lot from this internship and I enjoyed every moment of it. I am confident that this internship will be beneficial for my future academic and professional endeavors.

# **Chapter 1**

# **BHEL**

BHEL is India's largest engineering and manufacturing enterprise in the energy and infrastructure sectors. Established in 1964, it is a leading power equipment manufacturer globally and one of the earliest and leading contributors towards building an Aatmanirbhar Bharat.

The company serves its customers with a comprehensive portfolio of products, systems and services in the areas of power-thermal, hydro, gas, nuclear and solar PV; transmission; transportation; defence and aerospace; oil and gas and new areas like BESS and EV chargers.

## **1.1 History of BHEL**

BHEL, established in 1956, marked the beginning of India's heavy electrical equipment industry. Originally a manufacturing PSU with Soviet assistance, it merged with Heavy Electricals (India) Limited in 1974. By the 1980s, BHEL was at the forefront of thyristor technology and became a public company in 1991.

The company produces a range of electrical, electronic, and mechanical equipment for various sectors, with a significant portion of its revenue from power generation equipment. As of 2017, BHEL's equipment accounted for about 55% of India's installed power generation capacity. BHEL also supplies electric locomotives to Indian Railways and defense equipment. Exporting for over 40 years, BHEL's global presence spans 76 countries with over 9,000 MW of installed capacity in 21 countries, including Malaysia, Oman, and New Zealand.

## **1.2 Products**

The company undertakes projects and contracts in all modes including EPC, Supply, Supply & Supervision, Consortium partner, Contract Manufacturer, etc., as per customer requirement.

BHEL's spectrum of offerings includes one-stop solution for hydro, nuclear, solar, gas, thermal, railways, transmission as well as standalone products such as compressors, transformers, shunt-reactors, motors, pumps, heat exchangers, valves, oil-field equipment, after sales support etc.

### **1.2.1 Transmission**

BHEL is a leader in the field of power transmission in India with a wide range of transmission systems and products and having a proven track record across the globe.

- BHEL undertakes turnkey transmission projects from concept to commissioning on EPC basis which includes execution of EHV & UHV substations/ switchyards, both AIS and GIS types ranging from 33 kV to 765 kV, HVDC converter stations (up to  $\pm 800$  kV), and reactive power compensation schemes.
- BHEL is the first Transformer Manufacturer in India having successfully Short Circuit tested up to 500 MVA, 400 kV Auto Transformers.

### **1.2.2 Rail Transpoetation**

- BHEL is a leading designer and manufacturer of Rail Transportation systems like semi highspeed trains, Electric Locomotives, Diesel Electric Shunting Locomotives and Electrical Multiple Units.
- BHEL also manufactures critical equipment like Traction Converters/ Inverters, Motors, Transformers, Bogies, Train Control Management System (TCMS).

### **1.2.3 Defence and Aerospace**

- BHEL is a reliable supplier of equipment and services to Indian defence forces for almost three decades with dedicated engineering and manufacturing facilities



- Major products include super rapid gun mount (upgraded), strategic naval equipment, integrated platform management systems, thermos-pressed components, electrical machines, turret castings for T72 tanks, simulators, castings and forgings, etc.

#### 1.2.4 Water Management

BHEL offers complete water management solutions including desalination plants, water treatment plants (WTP), effluent treatment plants (ETP), sewage treatment plants (STP), tertiary treatment plants (TTP) and zero liquid discharge (ZLD) system for municipal segment, industries and power plant.

#### 1.2.5 E-mobility

- BHEL is committed to the world's e-mobility mission in a significant way, End-to-end integrated EV Charging infrastructure solutions including solar based chargers for cities & Highways
- First e-mobility friendly (Delhi-Chandigarh highway) in the country, equipped with 20 nos. BHEL-make solar PV based EV charging stations



### 1.2.6 Oil and Gas

- BHEL supplies complete onshore drilling rigs capable of drilling up to 9,000 metres, with ACSCR system or AC drives incorporating the latest state-of-the-art technology, and also mobile rigs, work-over rigs.
- BHEL also supplies onshore drilling rig equipment like draw works, rotary-table, travelling block, swivel, mast and substructure, mud systems, Artificial lift system (surface units of Sucker Rod Pumps) and rig electrics to leading oil and natural gas exploration companies of India.

### 1.2.7 Renewables

- BHEL is one of the few EPC players which manufactures almost all the major equipment of solar power plant viz. Solar Cells, Solar Modules, PCUs, HT Panels, SCADA, Power Transformers etc., with complete grid storage solutions
- BHEL has a dedicated industrial R&D Centre for developing high-efficiency silicon solar cells and process optimizations

### 1.2.8 Battery Energy Storage Solutions

- BHEL offers battery energy storage solutions to support the infirm power delivered by RE system and also standalone battery energy storage system connected with grid, with features like energy time shifting,

power smoothing, frequency regulation, anti-islanding and VAR support.

- Commissioned 1MWh battery energy storage system (BESS) with in-house developed PCS (power conditioning system) and EMS (energy management system) featuring three different types of battery chemistries – lithium ion battery, flow battery and advanced lead acid battery
- State-of-the-art facility for packaging and testing of Li-ion batteries for space applications. The company has assembled, tested and supplied batteries of 40 Ah - 180 Ah to ISRO for use in its satellites for more than two decades.

### **1.2.9 RESEARCH & DEVELOPMENT**

- R&D expenditure (>2.5% of Turnover) – one of the highest in Indian engineering field
- Centres of Excellence for Intelligent Machines & Robotics, Machine Dynamics, Compressors & Pumps, Nanotechnology, UHV Engineering, Simulators, Computational Fluid Dynamics, Surface Engineering, Permanent Magnet Machines, Advanced Transmission Systems, Power Electronics and Control Instrumentation.
- Five specialized institutes viz. Ceramic Technological Institute, Amorphous Silicon Solar Cell Plant, Pollution Control Research Institute, Welding Research Institute and Centre for Electric Transportation for carrying out advanced R&D in various engineering disciplines and identified technological areas.
- Filing patents/copyrights applications regularly – more than 4,800 filings.
- Strong engineering and R&D base for in-house development of indigenous technologies
- Technology collaboration agreements with leading global manufacturing and engineering companies.
- Focus on clean thermal technologies.

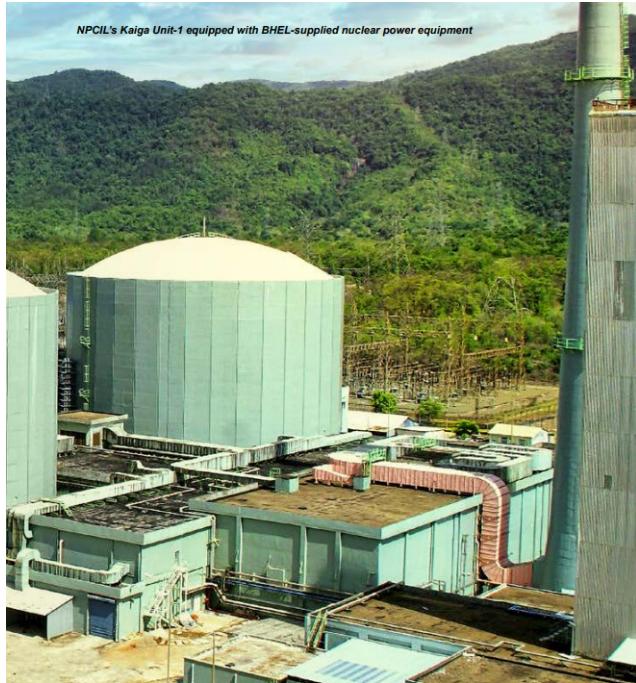
## **1.3 Power**

The power generation segment comprises of nuclear, hydro, thermal, gas and renewable power plant businesses. BHEL has been in this segment for over five decades, having commissioned its first thermal-based set in 1969. The company has proven turnkey capabilities for executing power projects from concept to commissioning.

- BHEL offers a wide variety of efficient supercritical sets up to 1,000 MW rating, including 660/ 700/ 800 MW. BHEL also offer lower rating thermal sets for utility and industrial applications.
- The company also offers state-of-theart emission control equipment for thermal-based plants for lower carbon footprint and compliance with the world standard emission norms
- The company manufactures a wide range of products for nuclear power plants viz. steam generators, reactor headers and end shields, besides nuclear turbine-generator sets ranging from 220 MWe to 700 MWe ratings POWER
- Customised hydro sets of Francis, Pelton and Kaplan types for different head-discharge combinations, are also engineered and manufactured by BHEL
- The company manufactures advancedclass large size gas turbines and matching generators. Additionally, the company supplies co-generation and combined cycle plants with higher plant efficiencies
- BHEL has proven expertise in plant performance improvement for BHEL and non-BHEL sets through renovation, modernization and uprating of a variety of power plant equipment including solution towards flexible operations

### **1.3.1 Nuclear Power**

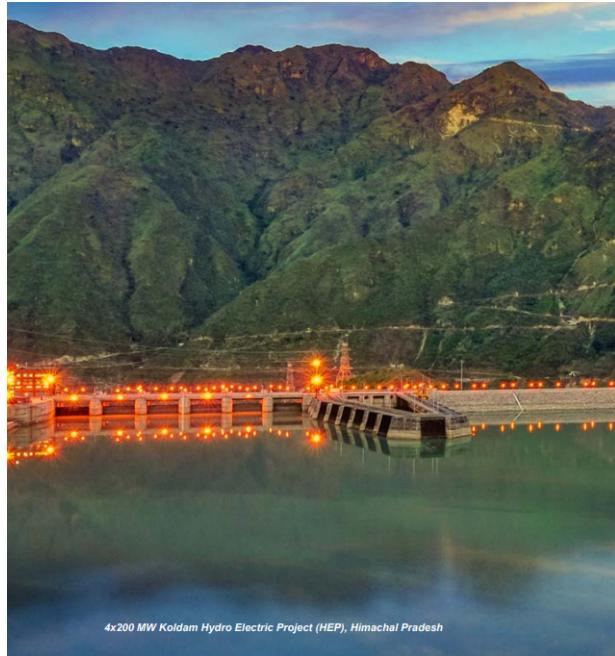
BHEL is the only Indian Company associated with all 3 stages of Indian Nuclear Power Program.



- BHEL has supplied and installed complete Turbine Island equipment for 12 out of 18 PHWRs that have been installed in the country, accounting for 74% of India's Indigenous installed capacity.
- BHEL is executing Turbine island equipment for 4x700 MWe PHWRs and has successfully synchronised India's highest rated first 700 MWe Turbine at Kakrapara Unit 3.

### 1.3.2 Hydro Power

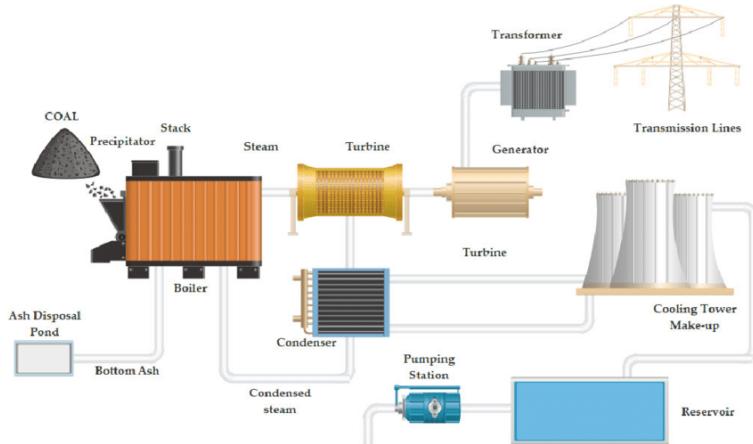
- BHEL is one of the leading players in the hydro power segment with a portfolio of over 500 hydroelectric sets, and a cumulative capacity of more than 31,000 MW globally.
- Hydro turbines in the range of 5 MW to 400 MW unit sizes of various impeller types namely Francis, Kaplan and Pelton along with matching generators are designed, engineered, manufactured and tested at BHEL's own manufacturing plants.



### 1.3.3 Thermal Power

- BHEL is capable of executing coal-based power projects on Engineering, Procurement & Construction (EPC) basis for subcritical & supercritical technologies up to 1,000 MW rating.
- BHEL is the pioneer in domestic manufacture and supply of Flue Gas Desulphurisation (FGD) System for SO<sub>x</sub> control and is executing a large number of orders for FGD systems for both, old and new plants.





## 1.4 Thermal Power Plant

Here is an overview of how thermal power plants work and their components

- **Fuel Combustion:** The primary fuel (coal, natural gas, or oil) is burned in a boiler to produce heat. In some thermal power plants, nuclear fuel (uranium or plutonium) is used to produce heat through nuclear fission.
- **Heat Transfer:** The heat generated from combustion or fission is used to convert water into steam in the boiler. This high-pressure steam is then directed to a steam turbine.
- **Turbine Operation:** The high-pressure steam drives the blades of the steam turbine, causing it to rotate. The turbine is connected to an electrical generator.
- **Electricity Generation:** As the turbine blades rotate, they turn the rotor of the generator, converting mechanical energy into electrical energy through electromagnetic induction.
- **Cooling and Condensation:** After passing through the turbine, the steam is condensed back into water in a condenser using cooling water from a nearby river, lake, or cooling tower. The condensed water is then recycled back to the boiler to be reheated.
- **Power Distribution:** The generated electrical power is transmitted through transformers to increase the voltage for efficient transmission over power lines to homes, businesses, and industries.

## Components of a Thermal power plant

- **Boiler:** Burns the fuel to produce heat, which converts water into steam.
- **Steam Turbine:** Converts thermal energy in steam into mechanical energy.
- **Generator:** Converts mechanical energy from the turbine into electrical energy.
- **Condenser:** Cools and condenses exhaust steam from the turbine back into water.
- **Cooling Tower:** Provides cooling water for the condenser.
- **Fuel Handling System:** Manages the storage, handling, and transportation of fuel.
- **Flue Gas Treatment System:** Reduces emissions and pollutants from the combustion process.
- **Control Room:** Monitors and controls the operation of the plant.

# Chapter 2

## Remote diagnostics

Remote diagnostics is a technology that allows for the monitoring, diagnosis, and troubleshooting of systems, equipment, or devices from a distance. This technology is increasingly used in various industries, including healthcare, automotive, information technology, and manufacturing.

### 2.1 Steps in remote diagnostics system

These are the steps that give an overview of how remote diagnostics works

- **Data Collection:** Sensors and monitoring devices collect data from the equipment or system. This data can include performance metrics, error codes, and environmental conditions.
- **Data Transmission:** The collected data is transmitted to a remote server or cloud-based platform via the internet or other communication networks.
- **Data Analysis:** Advanced software and algorithms analyze the data to detect anomalies, diagnose issues, and predict potential failures. This analysis can be real-time or scheduled.
- **Reporting and Alerts:** The system generates reports and sends alerts to technicians or engineers when issues are detected. This can include detailed diagnostic information and recommended actions.
- **Remote Access and Control:** In some cases, technicians can remotely access the equipment to perform diagnostics, updates, or repairs without needing to be physically present.

## 2.2 Benefits

- **Cost Savings:** Reduces the need for on-site visits by technicians, thereby saving travel and labor costs.
- **Improved Efficiency:** Speeds up the diagnosis and resolution of issues, minimizing downtime.
- **Enhanced Predictive Maintenance:** Allows for early detection of potential problems, preventing major failures and extending the lifespan of equipment.
- **Better Customer Service:** Provides quick and effective support to customers, enhancing their overall experience.
- **Data-Driven Insights:** Continuous data collection and analysis offer valuable insights for optimizing operations and improving product design.

## Equipment Exception Report



Customer  
**APGENCO**

Site  
**Vijayawada**

Unit  
**Vijayawada #7**

Date  
**13.10.2023**  
(00:00:00 to 23:59:59)

### Incidents (Last one week)

• Critical • Warning • Deviations

Total  
**92**

Approved/Pinned  
**0**

Added to Noise  
**0**

Without Feedback  
**92**

	13 OCT 2023	12 OCT 2023	11 OCT 2023	10 OCT 2023	09 OCT 2023	08 OCT 2023	07 OCT 2023
Incidents	<b>13</b> • 1 • 0 • 12	<b>21</b> • 2 • 0 • 19	<b>1</b> • 0 • 0 • 1	<b>0</b> • 0 • 0 • 0	<b>16</b> • 3 • 0 • 13	<b>13</b> • 2 • 1 • 10	<b>28</b> • 4 • 0 • 24
Approved/pinned	0	0	0	0	0	0	0
Added to noise	0	0	0	0	0	0	0
Open Incidents	149	144	144	144	142	141	137
Incidents without feedback	<b>13</b>	<b>21</b>	<b>1</b>	<b>0</b>	<b>16</b>	<b>13</b>	<b>28</b>

### Today's New Activity

26 new tags, 106 closed tags

Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Governing System 3-1-2	● Sec oil pr to ip VU_7MAX45CP011XQ01	Kg/cm2	<b>5.3</b>	6.1	<b>6.3</b>	<b>5.3</b>	<b>6.1</b>	132 hours in 7 days	
Mill-D	● Mill-d worm brg thr tmp VU_7HFC04CT122_XQ02	DegC	<b>41.4</b>	65.7	<b>75.4</b>	<b>44.3</b>	<b>63.5</b>	127 hours in 7 days	
Mill-D	● Mill-d low rad brg tmp VU_7HFC04CT120_XQ02	DegC	<b>43.5</b>	63.1	<b>75.0</b>	<b>41.8</b>	<b>62.8</b>	122 hours in 7 days	
Mill-D	● Mill-d up rad brg tmp VU_7HFC04CT119_XQ02	DegC	<b>43.1</b>	63.1	<b>75.2</b>	<b>43.7</b>	<b>62.4</b>	110 hours in 7 days	
Fdf-A	● Fdf-a mtr wng tmp-5 VUJ_7HLB10CT106_XQ02	DegC	<b>50.1</b>	55.3	<b>61.9</b>	<b>46.0</b>	<b>62.9</b>	40 hours in 7 days	

Figure 2.1: Equipment Exception Report

System Name: **BFP System**

Rootcauses and Recommended corrective actions:

Assets	Root Causes	Corrective Actions	Current Status	Duration
Tdbfp-A Discharge coupling Drain Oil Temperature High, Turbine rotor axial displacement high	Coupling Damage		Active	4 days in 5 days
Tdbfp-A TDBFP Lube oil header pressure low, Lube oil tank level low	Main oil tank level low	Top up the oil in MOT	Active	10 days in 9 days
Tdbfp-B Lube oil header pressure low, Gearbox journal bearing temperature high	Less lube oil supply	Increase LO header pressure by opening orifice, Rectify the NRVs	Inactive	9 days in 5 days

Figure 2.2: Equipment Exception Report

## Equipment Exception Report



Customer  
**APGENCO**

Site  
**Vijayawada**

Unit  
**Vijayawada #7**

Date  
**13.10.2023**  
(00:00:00 to 23:59:59)

Assets	Root Causes	Corrective Actions	Current Status	Duration
<b>Tdbfp-B</b> Lube Oil Header Pressure Low, BFP Journal bearing temperature high	Insufficient Lube Oil Pressure		Inactive	9 days in 5 days
<b>Tdbfp-A</b> Lube oil header pressure low, TDBFP Journal bearing temperature high	Insufficient Lube Oil Pressure		Inactive	30 minutes in 1 hours

Incidents									
Tdbfp		7		2		5		0	
7 Incidents, 19 Tags, 163 hours longest alarm state		Total	Critical	Warning	Deviations				
Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Tdbfp-A	Bfpdt-a lo tank lvl VIJ_7MTV10CL101XQ10	MM	1105.7	1119.3	1164.9	1127.9	1159.8	163 hours in 7 days	
Tdbfp-B	Bfpdt-b lo header press-2 VIJ_7MTV51CP122XQ10	-	2.5	2.7	3.4	2.7	3.0	142 hours in 7 days	
Tdbfp-B	Bfpdt-b lo header press-3 VIJ_7MTV51CP123XQ10	-	2.5	2.6	3.4	2.6	2.9	142 hours in 7 days	
Tdbfp-B	Bfpdt-b lo header press-1 VIJ_7MTV51CP121XQ10	Kg/CM2	2.5	2.6	3.4	2.6	2.9	142 hours in 7 days	
Tdbfp-A	Bfpdt-a lo header press-1 VIJ_7MTV50CP121XQ10	Kg/CM2	2.7	2.7	2.8	2.6	2.9	104 hours in 7 days	

+ 14 Other deviating tags for this asset:

VIJ\_7XD12CY031CXQ10 (Bfpdt-a axial disp) ; VIJ\_7MTV10CL102XQ10 (Bfpdt-a lo tank lvl) ; VIJ\_7MTV60CT124XQ10 (Bfpdt-a disc coupling drain temp) ; VIJ\_7MTV50CP122XQ10 (Bfpdt-a lo header press-2) ; VIJ\_7MTV50CP123XQ10 (Bfpdt-a lo header press-3) ; VIJ\_7XD12CY031AXQ10 (Bfpdt-a axial disp) ; VIJ\_7XD12CY031AXQ01 (Bfpdt-a axial disp upp) ; VIJ\_7XD12CY031BXQ10 (Bfpdt-a axial disp) ; VIJ\_7XD11CY011AXQ10 (Bfpdt-a diff expn) ; VIJ\_7LAC11CT105XQ10 (Bp jrn brg ml temp ndl) ; VIJ\_7KKD24CT001XQ10 (Bfpdt-b gear box ilp fb temp) ; VIJ\_7KAD22CY031AXQ02 (Bfpdt-a axial disp int) ; VIJ\_7LAC02CT104XQ10 (Jour brg nde mt temp) ; VIJ\_7PGB43CT105XQ10 (Mech s.cir de s.wtr of ltmp) ;

[View incidents on Pulse](#)

Figure 2.3: Equipment Exception Report

System Name: **Generator System**

**Exciter**

1 Incidents, 6 Tags, 20 hours longest alarm state

Incidents

<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>
Total	Critical	Warning	Deviations

Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Exciter	Hot air temp of exciter-2 VIJ_71MKC82CT02AXQ01	DegC	56.1	57.5	59.2	53.7	62.0	20 hours in 6 days	
Exciter	Hot air tempexciter-3 VIJ_72MKC82CT03AXQ01	DegC	55.1	56.6	58.2	52.5	61.4	19 hours in 5 days	
Exciter	Hot air temp exciter-1 VIJ_71MKC82CT01AXQ01	DegC	56.3	57.8	59.5	53.9	62.8	3 hours in 5 days	
Exciter	Hot air temp exciter-1 VIJ_72MKC82CT01AXQ01	DegC	56.3	57.8	59.5	53.9	62.5	3 hours in 5 days	
Exciter	Hot air temp of exciter-2 VIJ_72MKC82CT02AXQ01	DegC	56.1	57.5	59.1	53.6	62.3	3 hours in 5 days	

+ 1 Other deviating tags for this asset:

VIJ\_71MKC82CT03AXQ01 (Hot air tempexciter-3):

[View incidents on Pulse](#)

Incidents

<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>
Total	Critical	Warning	Deviations

**Gas System**

1 Incidents, 1 Tags, 19 minutes longest alarm state

Incidents

<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>
Total	Critical	Warning	Deviations

Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Gas System	% of h2 in air ( 100 - 90 % ) VIJ_7MKG25DQ003YQ21	%	97.7	98.0	98.1	97.9	98.2	19 minutes in 6 days	

[View incidents on Pulse](#)

Incidents

<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>
Total	Critical	Warning	Deviations

**Generator**

1 Incidents, 1 Tags, 1 hours longest alarm state

Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved

Report Generated on 14/10/2023 01:30:02 (IST Time)

4

Figure 2.4: Equipment Exception Report

System Name: **PCFB**

**Rootcauses and Recommended corrective actions:**

Assets	Root Causes	Corrective Actions	Current Status	Duration
<b>Fdf-B</b> Motor Input Current High, Motor Winding Temperature High	High input current	Check process parameters and driven equipment absorbed power vis a vis motor rating and consult driven equipment vendor, Check for overloading of Driven Equipment. Che...	Inactive	3 hours in 1 days
<b>Final Platen Sh</b> Superheater outlet header metal temp high, Steam temp at Platen O/L High	Steam temp at Platen O/L High	Refer Diagnosis of MS temp high	Inactive	1 hours in 10 hours
<b>Paf-B</b> Bearings damaged, Motor Winding Current High	Bearings damaged	Check Bearings for replacement	Inactive	3 hours in 4 days
<b>Rh Desuperheater</b> RH Spray Control Left and Right not in auto., Reheat Spray Flow High	RH spray Control Station (L&R) not in Auto	Keep RH Spray Control in Auto.	Inactive	439 days in 56 days
<b>Ltsh</b> Metal temperature MIN MAX difference high	Steam flow unbalance		Active	9 hours in 1 days
<b>Ltsh</b> SH tubes temperature above oxidation limit	Oxidation of tubes		Active	9 hours in 1 days
<b>Final Platen Sh</b> Main steam temperature high.	Superheater outlet left and right temperature high		Inactive	5 hours in 1 days
<b>Sh Attemperator</b> Main steam temperature high., spray control not in auto mode	SH Spray Control not in Auto	Keep SH Temp control in Auto	Inactive	482 days in 33 days
<b>Furnace</b> Main steam temperature high., Burner tilt in manual mode	Burner Tilt in Manual Mode	Put Burner Tilt in Auto	Inactive	881 days in 59 days
<b>Mill-E</b> Main steam temperature high., Top mills in operation	Top Mills in Operation	Switch to Bottom/Middle Mills	Inactive	37 days in 3 days

Figure 2.5: Equipment Exception Report

**Air Preheater**

1 Incidents, 1 Tags, 52 hours longest alarm state

		Total				Critical	Warning	Deviations	
Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Ah-A	Aph-a sb tmp VIJ_7HLD01CT101_XQ02	DegC	43.6	47.8	51.2	37.9	53.0	52 hours in 7 days	

[View incidents on Pulse](#)

**Boiler Drum**

1 Incidents, 1 Tags, 8 hours longest alarm state

		Total				Critical	Warning	Deviations	
Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Boiler Drum	Drum level-I VIJ_7HAD10CL102_XQ01	mm	758.6	811.8	850.1	765.0	877.5	8 hours in 6 days	

[View incidents on Pulse](#)

**Bus System**

1 Incidents, 4 Tags, 35 hours longest alarm state

		Total				Critical	Warning	Deviations	
Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Bus System	220kv bus-1 voltage r-y VIJ_7ACA00GX105_XQ06	KV	206.9	218.1	223.9	209.1	230.3	35 hours in 6 days	
Bus System	400kv bus-1 voltage r-y VIJ_7ACA00GX101_XQ06	KV	395.9	410.6	422.2	398.7	431.5	22 hours in 6 days	
Bus System	0bx bus voltage r-y ph VIJ_0BBC00GX101_XQ06	KV	10.6	11.0	11.3	10.7	11.6	18 hours in 6 days	
Bus System	220kv bus-2 voltage r-y VIJ_7ACA00GX107_XQ06	KV	211.7	220.3	226.1	213.9	232.1	17 hours in 6 days	

[View incidents on Pulse](#)

**Economizer**

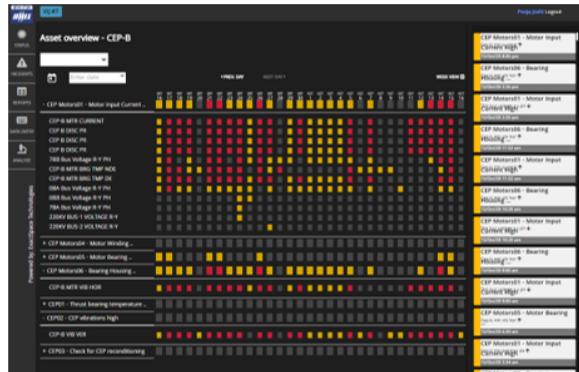
1 Incidents, 1 Tags, 39 minutes longest alarm state

		Total				Critical	Warning	Deviations	
Asset	Parameter	Measure Unit	Min	Mean	Max	Exp Min	Exp Max	Alarm Duration	Approved
Economizer	O2 in fg aft eco-I VIJ_7HHL00CF801_YJ23	%	2.0	4.0	7.7	1.9	6.9	39 minutes in 5 days	

[View incidents on Pulse](#)

Figure 2.6: Equipment Exception Report

- There are many red spots in CEP. The screen shot of Motor input current high and vibration high is attached below.



- Increasing CEP current with lower discharge pressure

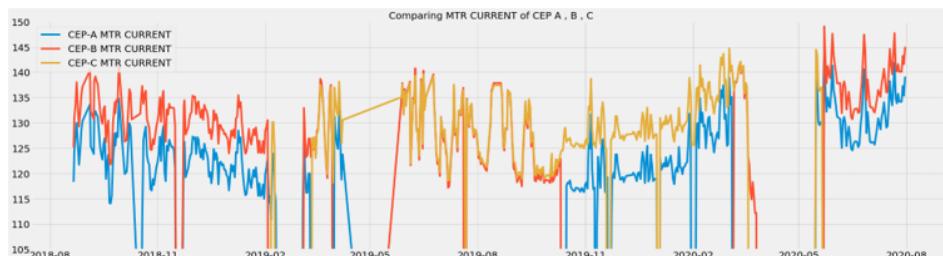


Figure 2.7: Equipment Exception Report

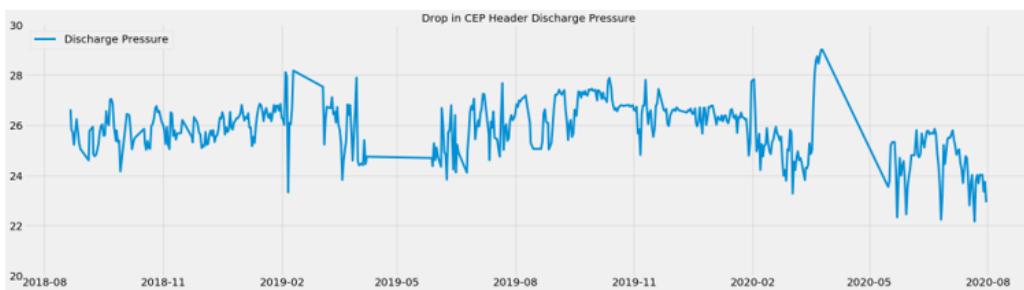


Figure 2.8: Equipment Exception Report

Recirculation Valve - No significant changes observed



Figure 2.9: Equipment Exception Report

34 - Condenser Wall Temperature High				
DETAIL		DATA		HISTORY
Month	Start Time	End Time	# of deviations	
Oct 2020	Oct 15th 9:00 am		34	<a href="#">View incident</a>
Oct 2020	Oct 15th 8:18 am	Oct 15th 8:33 am	4	<a href="#">View incident</a>
Oct 2020	Oct 15th 12:36 am	Oct 15th 1:03 am	6	<a href="#">View incident</a>
Oct 2020	Oct 14th 9:56 pm	Oct 14th 11:10 pm	5	<a href="#">View incident</a>
Oct 2020	Oct 14th 10:20 am	Oct 14th 8:53 pm	9	<a href="#">View incident</a>
Oct 2020	Oct 14th 9:21 am	Oct 14th 9:26 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 6:56 am	Oct 14th 7:03 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 5:53 am	Oct 14th 6:01 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 5:14 am	Oct 14th 5:19 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 2:35 am	Oct 14th 3:13 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 1:30 am	Oct 14th 1:58 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 12:58 am	Oct 14th 1:03 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 11:35 pm	Oct 14th 12:44 am	3	<a href="#">View incident</a>
Oct 2020	Oct 14th 8:12 pm	Oct 14th 8:23 pm	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 12:58 pm	Oct 14th 1:18 pm	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 9:58 am	Oct 14th 10:26 am	2	<a href="#">View incident</a>
Oct 2020	Oct 14th 9:26 am	Oct 14th 9:32 am	3	<a href="#">View incident</a>
			Comments	<a href="#">Comment</a>
Oct 2020	Oct 13th 8:27 am	Oct 13th 8:54 am	3	<a href="#">View incident</a>
Oct 2020	Oct 13th 2:22 am	Oct 13th 3:09 am	3	<a href="#">View incident</a>
Oct 2020	Oct 13th 2:22 am	Oct 13th 7:32 am	4	<a href="#">View incident</a>
Oct 2020	Oct 12th 8:20 pm	Oct 12th 11:52 pm	4	<a href="#">View incident</a>
Oct 2020	Oct 12th 7:14 pm	Oct 12th 7:28 pm	2	<a href="#">View incident</a>
Oct 2020	Oct 12th 4:34 pm	Oct 12th 5:20 pm	2	<a href="#">View incident</a>
Oct 2020	Oct 12th 10:16 pm	Oct 12th 2:55 pm	3	<a href="#">View incident</a>
Oct 2020	Oct 12th 3:46 am	Oct 12th 3:51 am	2	<a href="#">View incident</a>
Oct 2020	Oct 12th 2:07 am	Oct 12th 2:20 am	4	<a href="#">View incident</a>
Oct 2020	Oct 12th 12:13 am	Oct 12th 12:19 am	2	<a href="#">View incident</a>
Oct 2020	Oct 11th 8:17 pm	Oct 11th 11:00 pm	6	<a href="#">View incident</a>
Oct 2020	Oct 11th 12:02 pm	Oct 11th 2:14 pm	6	<a href="#">View incident</a>
Oct 2020	Oct 11th 10:45 am	Oct 11th 11:29 am	6	<a href="#">View incident</a>
Oct 2020	Oct 11th 6:24 am	Oct 11th 8:38 am	6	<a href="#">View incident</a>
Oct 2020	Oct 11th 1:39 am	Oct 11th 1:49 am	2	<a href="#">View incident</a>
Oct 2020	Oct 11th 12:32 am	Oct 11th 12:51 am	3	<a href="#">View incident</a>
Oct 2020	Oct 10th 5:21 pm	Oct 10th 8:49 pm	3	<a href="#">View incident</a>
Oct 2020	Oct 10th 2:34 pm	Oct 10th 4:47 pm	6	<a href="#">View incident</a>
Oct 2020	Oct 10th 12:31 pm	Oct 10th 12:53 pm	2	<a href="#">View incident</a>
Oct 2020	Oct 10th 11:23 am	Oct 10th 11:34 am	2	<a href="#">View incident</a>
Oct 2020	Oct 10th 9:38 am	Oct 10th 10:25 am	6	<a href="#">View incident</a>
Oct 2020	Oct 9th 7:51 am	Oct 10th 6:00 am	4	<a href="#">View incident</a>

Figure 2.10: Equipment Exception Report

- The incidents in LPBP system “LPBP Valve operation not Ok” is occurring repeatedly (refer the screen shot).

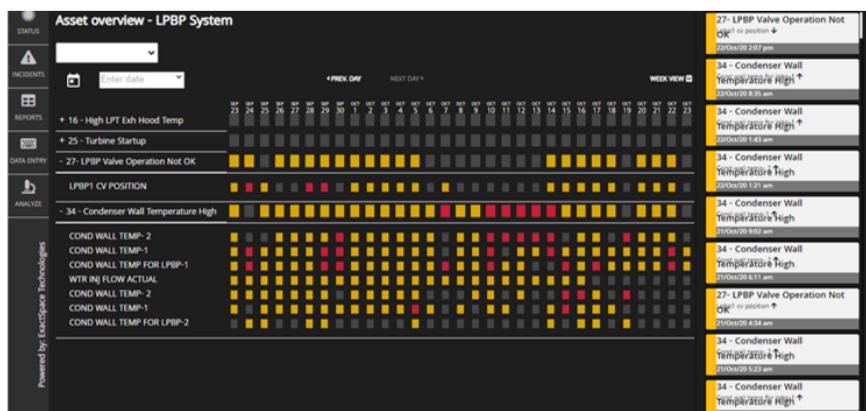


Figure 2.11: Equipment Exception Report

# Chapter 3

## Predictive Maintenance

Predictive maintenance leverages machine learning to analyze equipment sensor data and identify patterns indicative of potential failures before they occur. This allows manufacturers to schedule proactive maintenance and avoid costly unplanned downtime. This method is particularly beneficial in industrial settings, such as thermal power plants, where equipment reliability is critical to operational efficiency and safety.

### 3.1 Key Components and Processes

Implementing a predictive maintenance (PdM) program involves several key steps:

- **Identify assets for PdM:** First, it is essential to identify the equipment that would derive the most significant benefits from predictive maintenance. Assets that are critical to operations, expensive to repair or replace, or susceptible to unpredictable failures are prime candidates for this approach.
- **Collect and analyze data:** Once the target assets have been identified, the next step is to begin collecting condition monitoring data. This data may encompass readings such as vibration, temperature, pressure, voltage, and other relevant metrics over time. Analyzing this data can help detect patterns that signal impending equipment failures.
- **Choose monitoring techniques:** Select appropriate predictive maintenance techniques like vibration analysis, infrared thermography, ultrasonic analysis, oil analysis, etc. based on equipment type and failure modes. Then implement suitable monitoring hardware and software.

- **Develop predictive models:** Leverage machine learning algorithms to develop models that can predict failures from the condition monitoring data. Techniques like LSTM networks work well for timeseries data.
- **Deploy and iterate:** Initiate the process with a pilot deployment on non-critical equipment. Collect feedback, refine the models and processes, and subsequently scale the deployment to include all pertinent equipment. Continuously enhance the predictive maintenance (PdM) program through iterative improvements.

## 3.2 Task and Data description

Since real predictive maintenance datasets are generally difficult to obtain and in particular difficult to publish, the data provided by the UCI repository is a synthetic dataset that reflects real predictive maintenance encountered in industry to the best of their knowledge. The dataset consists of 10 000 data points stored as rows with 14 features in columns:

- **UID:** Unique identifier ranging from 1 to 10000.
- **Product ID:** Consisting of a letter L, M, or H for low (60% of all products), medium (30%) and high (10%) as product quality variants and a variant-specific serial number;
- **Air temperature [K]:** Generated using a random walk process later normalized to a standard deviation of 2 K around 300 K;
- **Rotational speed [rpm]:** Calculated from a power of 2860 W, overlaid with a normally distributed noise;
- **Torque [Nm]:** Torque values are normally distributed around 40 Nm with a standard deviation of 10 Nm and no negative values;
- **Tool wear [min]:** The quality variants H/M/L add 5/3/2 minutes of tool wear to the used tool in the process;
- **Tool wear failure (TWF):** The tool will be replaced or fail at a randomly selected tool wear time between 200 - 240 mins;
- **Heat dissipation failure (HDF):** Heat dissipation causes a process failure, if the difference between air- and process temperature is below 8.6 K and the tools rotational speed is below 1380 rpm;

- **Power failure (PWF):** The product of torque and rotational speed (in rad/s) equals the power required for the process. If this power is below 3500 W or above 9000 W, the process fails;
- **Overstrain failure (OSF):** If the product of tool wear and torque exceeds 11,000 minNm for the L product variant (12,000 M, 13,000 H), the process fails due to overstrain;
- **Random failures (RNF):** Each process has a chance of 0,1 % to fail regardless of its process parameters. If at least one of the above failure modes is true, the process fails and the 'machine failure' label is set to 1. It is therefore not transparent to the machine learning method, which of the failure modes has caused the process to fail.

### 3.3 Exploratory Analysis

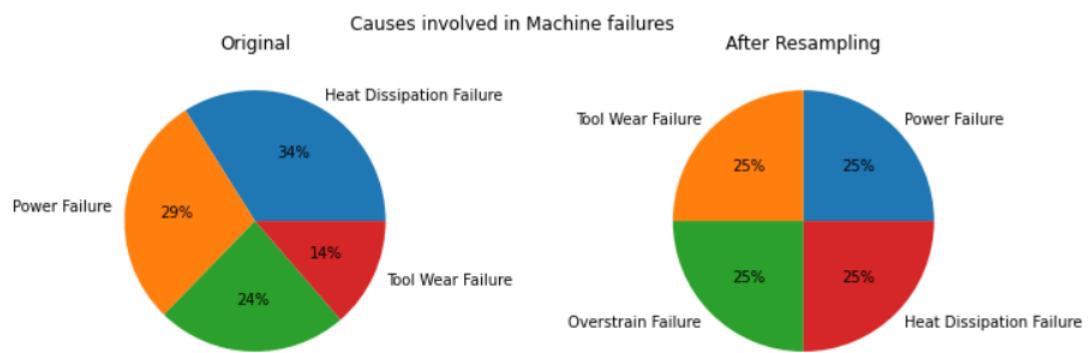
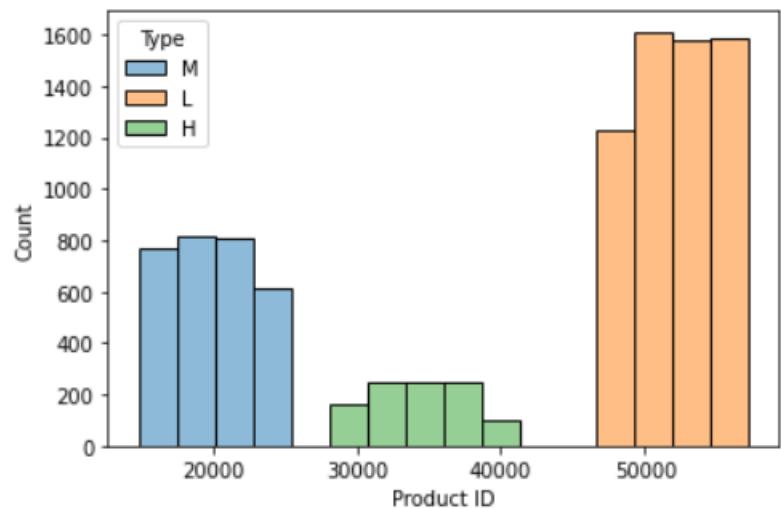
The data exploration starts by checking that each entry is unique and there are no duplicates. This is done by verifying that the number of unique ProductID corresponds to the number of observations. Then we print a report to look for missing values and check the data type for each column.

To sum up even more:

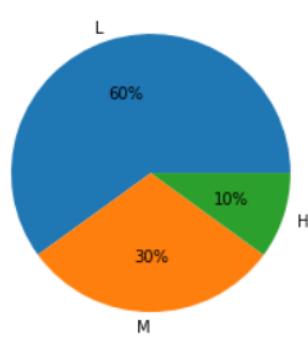
- There is no missing data;
- There are no duplicate values;
- Six columns are numerical features, including UDI.
- Three columns are categorical features, including ProductID.

#### 3.3.1 Product ID

Before addressing technical matters, we need to handle the two ID columns, as they might confuse our model. The UDI column is a copy of the dataframe index, and the Product ID, consisting of an initial letter (indicating machine type) followed by five numbers, seems to carry no additional information. The number sequences define three intervals based on machine type, confirming that Product ID doesn't add value beyond the Type feature. Therefore, it's reasonable to drop it.



Machine Type percentage



### 3.3.2 Resampling with SMOTE

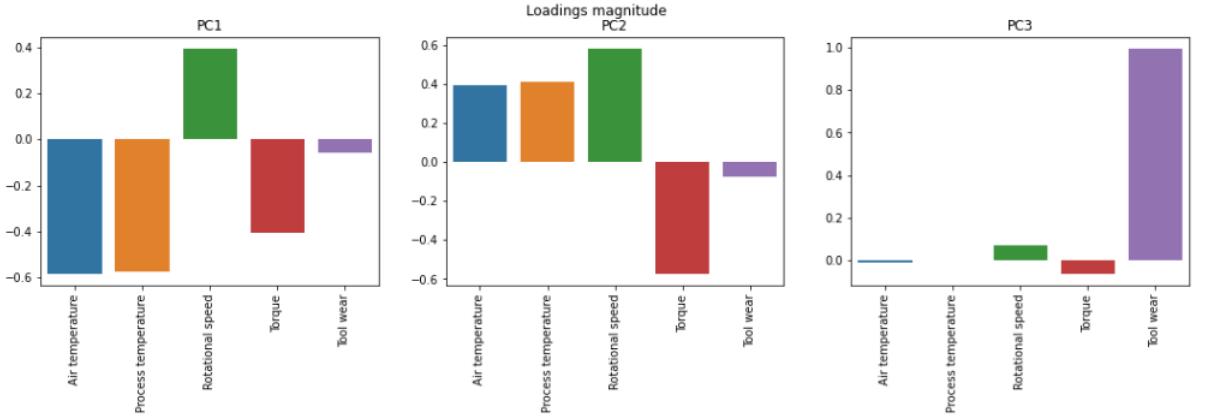
Class imbalance in machine learning can mislead models and skew interpretation. For instance, a model predicting no machine failures would be 97% accurate. To prevent this bias and achieve an 80/20 ratio between functioning and faulty machines, along with balanced failure causes, we use data augmentation.

Common techniques include:

- Under-sampling: Removing data points from the majority class.
- Over-sampling: Copying rows from the minority class.
- SMOTE (Synthetic Minority Oversampling Technique).

Under-sampling reduces dataset size, which is already limited. Over-sampling simply duplicates data. Instead, we use SMOTE to create synthetic samples by slightly shifting data points towards their neighbors. This generates new, realistic data points. SMOTE works by selecting a random minority sample, finding its  $k$  nearest neighbors, and creating new points along vectors between the sample and its neighbors.

Machine failure cases mainly involve low-quality machines, followed by medium and rarely high-quality ones. This imbalance is amplified when non-functioning machine observations are artificially increased. However, feature distributions remain similar across quality levels, except for the side peaks in Tool Wear, which aligns with the data description. This suggests that the higher failure rate in type L machines is likely due to their greater representation in the dataset, rather than a strong correlation with machine failure.



### 3.3.3 PCA and Correlation Heatmap

We run PCA to have a further way of displaying the data instead of making feature selection. Explained variance ratio per component:

- PC1 37.69
- PC2 36.81
- PC3 19.84
- PC4 3.08
- PC5 2.58

Since the first three components are enough to almost fully represent the variance of the data we will project them in a three dimensional space.

The bar plot of Principal Components weights makes easy to understand what they represent:

- PC1 is closely related to the two temperature data.
- PC2 can be identified with the machine power, which is the product of Rotational Speed and Torque.
- PC3 is identifiable with Tool Wear.

### 3.3.4 Metric

To evaluate the models we will use from a quantitative point of view, we resort to some metrics that summarize some characteristics of the classification results:

- Accuracy : expresses the fraction of instances that are classified correctly, it is the most intuitive metric that is usually used in classification tasks.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

- AUC : can be considered as a measure of the separation between True Positives and True Negatives, that is, the ability of the model to distinguish between classes. In detail, it represents the area below the ROC curve, given by the estimate of the True Positive Rate (Recall) for each possible value of the True Negative Rate).
- F1: reports the classification capacity of the model to Precision and Recall, giving both the same weight

$$F1 = 2 \frac{Precision * Recall}{Precision + Recall}$$

- Although generally effective, AUC can be optimistic in the case of highly unbalanced classes, as happens in the binary task, while the F1 score is more reliable in this kind of scenario. We consider this last metric particularly significant as it is able to mediate the cases in which the machines that are about to fail are classified as functioning (Recall) and the one in which functioning machines are classified as about to suffer a failure (Precision). To be more specific we will give more importance to Recall than Precision, by evaluating also an "adjusted" version of the F1 through a  $\beta$  parameter:

$$F_\beta = (1 + \beta^2) \frac{Precision * Recall}{\beta^2 Precision + Recall}$$

- With the choice  $\beta = 2$  (common in literature) a greater influence of the Recall is obtained. This choice is motivated by the fact that in order to optimize the costs for the maintenance of the machinery it is a good thing to limit the purchase of unnecessary replacement materials but it results far more important to avoid the possibility of having to replace a machinery after it is broken, since this second scenario generally has higher costs.

## 3.4 Binary Task

### 3.4.1 Preliminaries

The goal of this section is to find the best model for binary classification to predict Machine Failure. Classification algorithms, part of data mining, use supervised learning methods to make predictions. We start with labeled data to create a model, which is then used to classify new, unlabeled data. The dataset is typically divided into three groups:

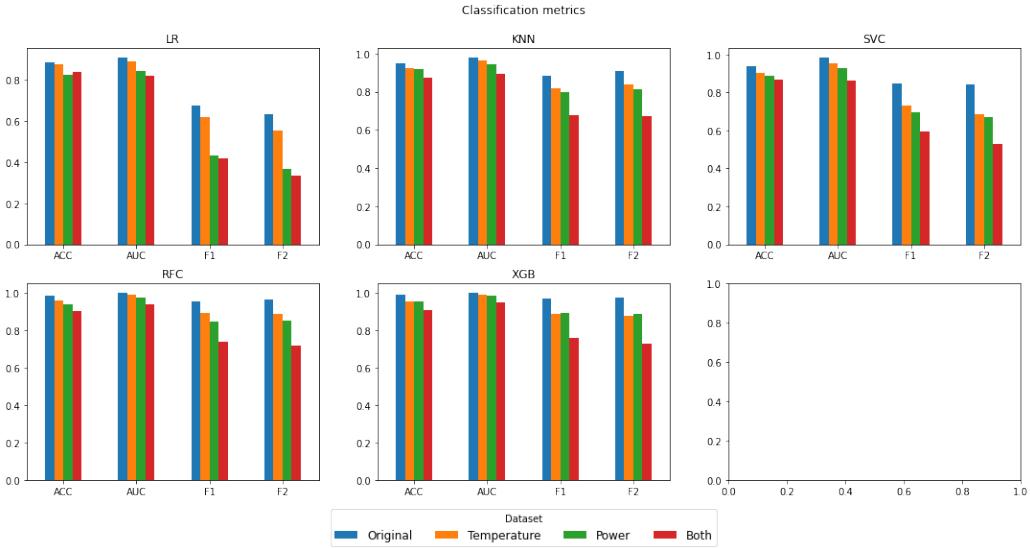
- **Training dataset:** Used to fit the model.
- **Validation dataset:** Used to evaluate model hyperparameters.
- **Test dataset:** Used to test the final model.

A data scientist must make this division at the start of a project. Common split ratios include:

- Training/Validation/Test: 80/10/10
- Training/Validation/Test: 70/15/15
- Training/Validation/Test: 60/20/20

In this project, we use an 80/10/10 split ratio because testing various strategies showed this to be the most effective. The classification techniques we implement are:

- **Logistic Regression:** Estimates the probability of a dependent variable based on independent variables. The dependent variable is the output we predict, while the independent variables are factors that could influence this output. Due to its simplicity and interpretability, we use Logistic Regression as a benchmark model, serving as a starting point for comparison.
- **K-Nearest Neighbors (K-NN):** Classifies data based on the distance between elements in the dataset. Data is assigned to a class if it is close enough to other data in that class. The parameter K represents the number of neighboring data points considered when assigning classes.
- **Support Vector Machine (SVM):** Aims to find a hyperplane in an N-dimensional space (where N is the number of features) that distinctly classifies data points while maximizing the margin distance between points of different classes.



- **Random Forest:** Uses ensemble learning by combining many classifiers to solve complex problems. It constructs multiple decision trees in parallel, with each tree having equal importance. The final output is the class selected by most trees, using a bagging technique.
- **XGBoost:** A gradient-boosted decision tree (GBDT) machine learning library. Unlike Random Forest, it uses boosting, which iteratively trains an ensemble of shallow decision trees. Each iteration uses the error residuals from the previous model to fit the next model, with the final prediction being a weighted sum of all tree predictions.

### 3.4.2 Feature Selection

Before training the models, we perform feature selection based on our correlation heatmap and exploratory data analysis. We noticed that "Process temperature" and "Air temperature" are positively correlated, while "Torque" and "Rotational speed" are negatively correlated. According to the dataset description, PWF failure occurs if the product of "Torque" and "Rotational speed" is within a certain range, and HDF failure occurs when the difference between "Air temperature" and "Process temperature" exceeds a certain value.

Rather than deleting these columns and losing important information, we combine them to create new features with physical meaning. We then compare the results of fitting the classification models (without tuning parameters) on the following datasets:

- The original dataset.
- A dataset with "Process temperature" and "Air temperature" replaced by their product.
- A dataset with "Torque" and "Rotational speed" replaced by their product.
- A dataset combining the above operations.

From the results, we observe that all the models applied to the entire dataset perform better than when they are applied to the ones created by reducing the number of features. The best performances and the modest number of features from which our dataset is composed encourage us to opt to avoid the feature selection step.

Validation scores:				
	KNN	SVC	RFC	XGB
ACC	0.959	0.966	0.977	0.987
AUC	0.954	0.987	0.997	0.999
F1	0.902	0.916	0.943	0.969
F2	0.928	0.931	0.954	0.970
Test scores:				
	KNN	SVC	RFC	XGB
ACC	0.966	0.973	0.972	0.983
AUC	0.954	0.992	0.997	0.998
F1	0.916	0.934	0.931	0.958
F2	0.927	0.941	0.945	0.956

### 3.4.3 Models

All selected models achieve similar results on the validation set, except for KNN, which performs slightly worse. Performance remains stable on the test set, indicating that overfitting was avoided. We evaluate the models using confusion matrices and metrics from the test set, which helps clarify their hierarchy. KNN has the worst performance, while XGB performs the best. SVC and RFC yield very similar results.

Regarding parameter tuning:

- Grid search focused on key parameters identified in the literature.
- Grid values were chosen based on literature and tests, balancing computational cost and search effectiveness.

Interestingly, RFC and XGB have opposite optimal parameters: RFC uses fewer estimators with deeper trees, while XGB uses more estimators with fewer splits. Despite XGB being the best quantitative classifier, it lacks the interpretability of RFC, which provides a clearer understanding of its workings.

## 3.5 Multi-Class Task

Next, we tackle predicting not only if a failure will occur but also the type of failure, making this a multiclass classification problem. This is feasible since we removed ambiguous observations during preprocessing, ensuring each sample belongs to one label.

For multiclass targets, we set the parameter "average=weighted" when calculating AUC, F1, and F2 scores to account for class imbalance (80% WORKING, 20% failing). We use Logistic Regression as our baseline model and seek models that perform better on these metrics. We adapt the models from the previous section for multiclass classification.

While many algorithms (e.g., K-NN, Random Forest, XGBoost) naturally support multiclass classification, some (e.g., Logistic Regression, SVM) are inherently binary. We convert these into multiclass classifiers using the "One-vs-Rest" approach, which trains a single classifier per class, treating that class as positive and all others as negative. This method is chosen for its computational efficiency.

### 3.5.1 Models

For each model we launch the Gridsearch for hyperparameter optimization, using as metric to evaluate the model the weighted average F2 score. Similarly to the binary case, the Gridsearch has been started on the parameters that, looking in the literature, are found to be preponderant for each specific model and the grid values to look for have been defined according to the literature and several tests carried out.

By comparing the results obtained, we see that K-NN is the model that performs the worst and its accuracy is a little lower than Logistic Regression's one. Despite this, we cannot exclude it a priori, as it still reaches high values for the metrics and, moreover, gives an immediate response. So, we can use it whenever we need to get an idea quickly about the situation and, then apply other models when we have more time.

All other models perform better than the benchmark and they obtain high values for the chosen metrics both for validation and test set. SVC and RFC's performances are very similar each other and XGB performs better than them. If we look at the training phase, SVC and RFC take the same time, while XGB takes more than four times as much as them. So, since, the improvement obtained with XGB is only 1.5%, one can choose which model

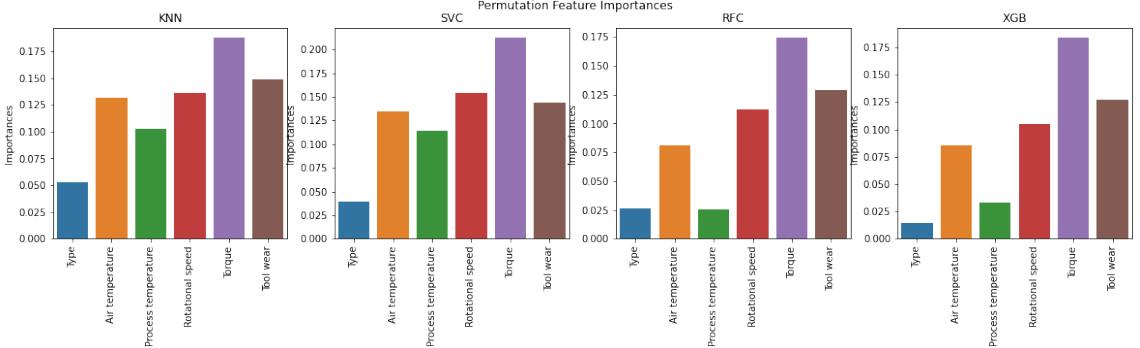
**Validation scores:**

	KNN	SVC	RFC	XGB
ACC	0.956	0.968	0.975	0.986
AUC	0.956	0.993	0.998	0.999
F1	0.957	0.969	0.975	0.986
F2	0.957	0.968	0.975	0.986

**Test scores:**

	KNN	SVC	RFC	XGB
ACC	0.966	0.973	0.972	0.989
AUC	0.956	0.995	0.997	0.999
F1	0.966	0.973	0.972	0.989
F2	0.966	0.973	0.972	0.989

he prefers according to his needs. While the best parameters for multiclass K-NN and SVC are the same as binary classification, for XGB and RFC the Gridsearch for the two types of task returns different parameters. Moreover, in the transition from binary to multiclass problem, the estimated training time remains the same for all models, except for XGB that triples it. In order to understand how features contribute to predictions, let's look at the Permutation Feature Importances for each model.



From previous barplots we see that the models give more importance to Torque, Tool wear and Rotational Speed while the Type contribution is very low. This is in accordance with the observations made in the exploration of the dataset in Section 1-2 and it is consistent with the Permutation Feature Importances of binary task. K-NN is the one who gives more importance to Type, but, different from binary case, here we see that for every model the Type contribution is almost zero.

So, we test the model on a new dataset, the old one from which we removed the column Type. For K-NN and SVC there is an insignificant improvement in the metrics' values, which were already very good. For RFC and XGB we do not see any change on metrics' values. Since the training time for the different models is approximately equal in both cases, we let users choose which dataset to use.

## 3.6 Conclusions

Based on the analyses and results obtained, several conclusive considerations can be made regarding this project.

### 3.6.1 Project Objectives

We aimed to address two primary tasks:

- **Predicting Machine Failure:** Determine whether a machine will fail or not.
- **Predicting Failure Type :** Identify the specific type of failure that will occur.

### 3.6.2 Data Preprocessing

Before developing the models, we undertook thorough data preprocessing to ensure model validity and optimize performance. Key steps included:

- **Removing Ambiguous Samples:** Eliminating data points that belonged to multiple classes to maintain the integrity of our classification tasks.
- **Label Encoding:** Converting categorical columns into a numerical format suitable for machine learning algorithms.
- **Scaling Features:** Using StandardScaler to standardize the dataset, ensuring that each feature contributes equally to the model's performance.
- **Outlier Analysis:** Initially identified outliers were later recognized as part of the natural variance, crucial for accurate classification.
- **Principal Component Analysis (PCA):** PCA revealed that most variance is explained by the first three components, which can be represented by:
  1. A combination of the two temperatures.
  2. Machine power (product of rotational speed and torque).
  3. Tool wear.

These features were found to be the most significant contributors to our predictions.

### 3.6.3 Model Development and Evaluation

We adapted various models for both binary and multiclass classification tasks. Specific techniques included:

- **Baseline Model:** Logistic Regression served as our starting point for comparisons.
- **Advanced Models:** We implemented K-Nearest Neighbors (K-NN), Random Forest, Support Vector Machine (SVM), and XGBoost.

### 3.6.4 Binary Classification

- **Best Model:** XGBoost consistently outperformed other models in accuracy and other performance metrics.

- **Worst Model:** KNN showed the lowest performance but had the fastest response time.

### 3.6.5 Multiclass Classification

- **Adaptation:** For binary-based algorithms like Logistic Regression and SVM, we employed the "One-vs-Rest" strategy, training individual classifiers for each class.
- **Performance:** XGBoost remained the top performer, while KNN was the least effective, though it maintained its advantage in speed.

### 3.6.6 Feature Importance

Contrary to initial expectations, our analysis showed that the machine's type does not significantly impact failure presence. The most influential features in predicting machine failures were:

- Combination of air and process temperatures.
- Machine power.
- Tool wear.

### 3.6.7 Conclusion

For both tasks, the chosen models performed well, with specific observations:

- **XGBoost:** Best overall accuracy and performance but slower, especially for multiclass classification.
- **KNN:** Fastest response time, making it suitable for applications where speed is crucial, though it sacrifices some accuracy.

This comprehensive approach ensures that the models not only perform well but also provide actionable insights tailored to specific business requirements.

# Chapter 4

## Environmental and Societal Impact

### 4.1 Environmental Impact

- **Reduction in Emissions:** By optimizing the maintenance of equipment, predictive maintenance reduces the chances of unexpected breakdowns, which can lead to excessive emissions of pollutants such as CO<sub>2</sub> and NO<sub>x</sub>.
- **Energy Efficiency:** Properly maintained equipment operates more efficiently, consuming less fuel and producing fewer emissions.
- **Extended Equipment Lifespan:** Predictive maintenance prolongs the operational life of critical components, reducing the need for frequent replacements and minimizing waste and the environmental impact associated with the production of new equipment.
- **Reduced Risk of Environmental Hazards:** Regular monitoring and timely interventions help prevent catastrophic failures, such as oil spills or leakages of hazardous materials, which can have severe environmental consequences.

### 4.2 Societal Impact

- **Community Well-being:** A more stable and reliable power supply promotes the well-being of local communities, supporting their daily activities, education, and healthcare systems by reducing power outages.

- **Increased Reliability and Safety:** Predictive maintenance enhances the reliability of power generation, ensuring consistent electricity supply to communities, which is critical for economic and social stability.

# Chapter 5

## References

- <https://www.bhel.com>
- <https://dataheadhunters.com/academy/how-to-use-python-for-predictive-maintenance-in-manufacturing>
- <https://www.learnpick.in/prime/documents/ppt/68/thermal-power-plant-description>
- <https://www.kaggle.com/code/gerardocappa/predictive-maintenance-final-project>

# Chapter 6

## NBA/IET Mapping

NBA - PROGRAM OUTCOMES (PO) & PROGRAM SPECIFIC OUTCOMES (PSO)

Engineering Graduates will be able to:

- PO1:Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- PO2:Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
- PO3:Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- PO4:Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- PO5:Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

- PO6:The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- PO7:Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
- PO8:Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
- PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
- PO10:Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
- PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
- PO12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.
- PSO1: Analyze and solve real world problems by applying a combination of hardware and software.
- PSO2: Formulate & build optimized solutions for systems level software & computationally intensive applications.
- PSO3: Design & model applications for various domains using standard software engineering practices.
- PSO4: Design & develop solutions for distributed processing & communication.

NBA CO PO Mapping

CSE 4298	CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CSE 4298.1	Understand the functioning of the industry	1	0	0	0	0	0	0	2	2	0	0	0	1	0	0	0
CSE 4298.2	Understand the requirements of real world applications	2	1	3	1	1	1	1	0	3	2	0	1	2	1	1	0
CSE 4298.3	Demonstrate skills to use modern engineering tools, software and equipment to analyze problems	0	0	2	2	3	1	1	0	3	2	0	1	2	1	1	0
CSE 4298.4	Demonstrate an ability to envisage and work on laboratory and multidisciplinary tasks	0	0	2	1	3	1	1	0	3	2	0	1	2	1	1	0
<b>CSE 4298 (Avg. correlation level)</b>		<b>0.75</b>	<b>0.25</b>	<b>1.75</b>	<b>1</b>	<b>1.75</b>	<b>0.75</b>	<b>0.75</b>	<b>0.5</b>	<b>2.75</b>	<b>1.5</b>	<b>0</b>	<b>0.75</b>	<b>1.75</b>	<b>0.75</b>	<b>0.75</b>	<b>0</b>

Figure 6.1: NBA CO PO Mapping

Course Code	Course Title	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CSE 4298	Industrial Training	0.75	0.25	1.75	1	1.75	0.75	0.75	0.5	2.75	1.5	0	0.75	1.75	0.75	0.75	0

Figure 6.2: NBA Articulation Matrix

### IET CLO – AHEP 4 Mapping

CSE 4298	CLO Statements	AHEP Los						
		C8	C9	C10	C12	C15	C16	C17
CSE 4298.1	Understand the functioning of the industry	2	1	0	1	1	2	2
CSE 4298.2	Understand the requirements of real world applications	0	2	1	1	3	3	2
CSE 4298.3	Demonstrate skills to use modern engineering tools, software and equipment to analyze problems	0	2	1	1	2	2	2
CSE 4298.4	Demonstrate an ability to envisage and work on laboratory and multidisciplinary tasks	0	1	1	0	0	1	1

Figure 6.3: IET CLO – AHEP 4 Mapping

AHEP 4 Statements - Map suitable statement with your ITR work	
C8	Identify and analyze ethical concerns and make reasoned ethical choices informed by professional codes of conduct
C9	Use a risk management process to identify, evaluate and mitigate risks (the effects of uncertainty) associated with a particular project or activity
C10	Adopt a holistic and proportionate approach to the mitigation of security risks
C12	Use practical laboratory and workshop skills to investigate complex problems
C15	Apply knowledge of engineering management principles, commercial context, project and change management, and relevant legal matters including intellectual property rights
C16	Function effectively as an individual, and as a member or leader of a team
C17	Communicate effectively on complex engineering matters with technical and non-technical audiences

Declaration: Through this Industrial Training, I have accomplished the above stated program articulation and IET learning outcomes.

Figure 6.4: AHEP 4 Statements - Map suitable statement with your ITR work

## 210905121\_Kshitij\_Karn\_ITR.docx

### ORIGINALITY REPORT

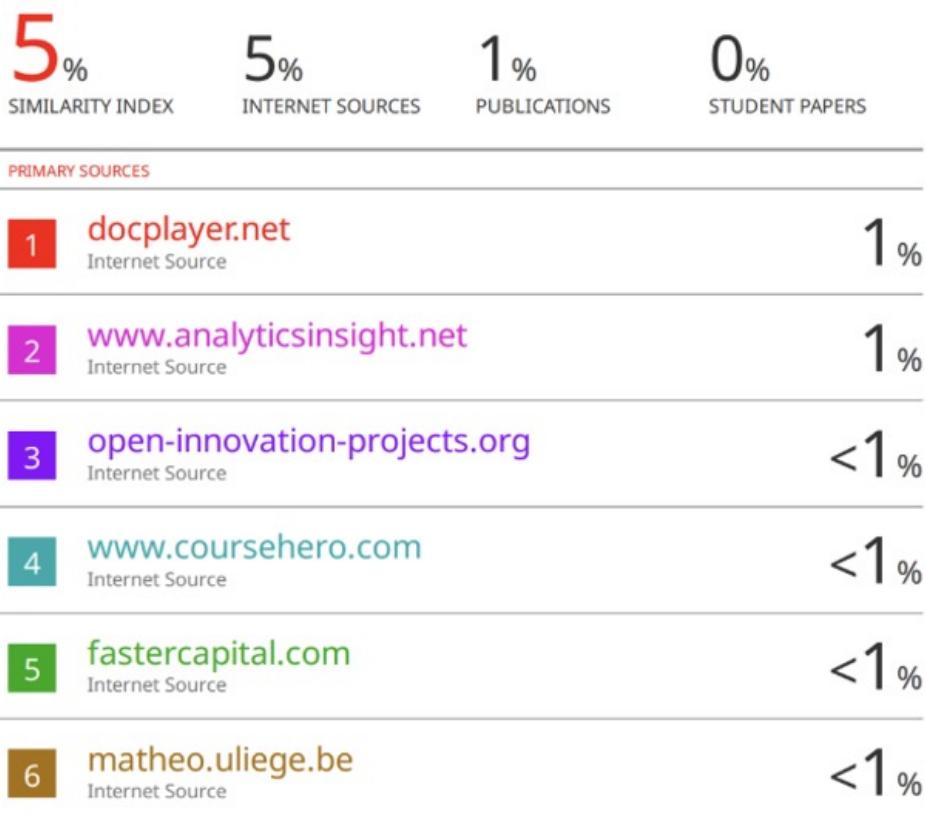


Figure 6.5: plagiarism report



Figure 6.6