**Machine Failure Prediction**



A Project Report in partial fulfillment of the degree

# Bachelor of Technology

in

# Computer Science & Engineering

## By

**2103A51138** PAKA SATHVIKA

**2103A51139** PAMERA SHIRISHA

**2103A51140** PEDDAPELLY MRUNALINI

**Under the Guidance of**

# D. Ramesh

**Submitted to**



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

## S R UNIVERSITY, ANANTHASAGAR, WARANGAL



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CERTIFICATE**

This is to certify that the Project Report entitled “MACHINE FAILURE PREDICTION” is a record of Bonafide work carried out by **P.Sathvika, P.Shirisha, P.Mrunalini** bearing Roll No(s) **2103A51138, 2103A51139, 2103A51140** during the academic year 2022-2023 in partial fulfillment of the award of the degree of ***Bachelor of Technology*** in **Computer Science Engineering** by the SR UNIVERSITY, WARANGAL.

# Supervisor Head of the Department

Mr. D. Ramesh Dr. M. Sheshikala

SR University SR University

**External Examiner**

**ACKNOWLEDGEMENT**

We express our thanks to course coordinator Mr. D. Ramesh, Asst. prof. for guiding us from the beginning through the end of the course project. We express our gratitude to head of the department CS&AI, Dr. M. Sheshikala, Associate Professor for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, Dr C. V. Guru Rao, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thank to all teaching and non-teaching staff of the department for their suggestions and timely support.

# ABSTRACT

Machine failure prediction is a critical task in many industries, ranging from manufacturing and transportation to energy and health care. Unplanned downtime due to equipment failure can result in significant financial losses and safety hazards, It is important to develop accurate prediction models to prevent such failure.

Machine failure prediction involves analyzing historical data from machine to identify Patterns and trends that can indicate potential failure. This data typically includes sensor readings, maintenance logs, and operational data, and is analyzed using statistical models, machine learning algorithms, and artificial neural networks.

By accurately predicting when a machine is likely to fail, companies can schedule preventative maintenance and avoid unexpected downtime, which can result in significant cost savings and increased productivity. In addition, machine failure prediction can also help companies identify underlying issues with their equipment and make improvements to prevent future failures.

Overall, machine failure prediction is a critical aspect of equipment maintenance and management, enabling companies to optimize their operations and improve overall efficiency.

# Table of Contents

**Chapter No. Title Page No.**

**1 INTRODUCTION 1**

1.1 Problem Statement 1

1.2 Existing System 1 1

1.3 Proposed System 2 1

1.4 Objectives 2 1

**2 LITERATURE SURVEY 2-4**

**3** **DATA PRE-PROCESSING 5-10**

3.1 Dataset Description 6

3.2 Pre-processing through Standard scaler 6

3.3 Data Visualization 7-10

**4** **METHODOLOGY 11-12**

4.1 logistic regression 11

4.2 K-Nearest Neighbor 11

4.3 Naïve bayes 12

4.4 Decision tree 12

4.5 support vector machine 12

**5 RESULTS 13-15**

**6 CONCLUSION 16**

**7 REFERENCES 16**

1. **INTRODUCTION**

Our project is about machine failure prediction. Although machine failure is a very general issue that can occur in any machine, predicting the failure and taking steps to prevent such failure is most important for any machine or software application. In the present situation, when we are entirely dependent on machines and computers, machine failure affects the entire lifecycle to a great extent. Machine failure also leads to substantial business losses if we talk about organizations. In that case, we can easily avoid such failures and prevent the entire system from breaking down by using the machine failure prediction application.

**1.1.PROBLEM STATEMENT**

To develop a model which helps to predict the failure of a machine with low error rate and a high precision of accuracy. The model will not tell the failure of a machine which will done in the future. But, it helps the companies and organizations identify potential problems with their equipment and take prevention measures to avoid costly downtime and repairs.

**1.2.EXISTING SYSTEM**

Firstly, we have collected the dataset of our project from the online source. The dataset represents the failure of machine. The dataset includes all the information about Air temperature, Process temperature, Torque and Rational. The existing systems which use various machine learning algorithms and techniques to analyze data and predict when a machine is likely to fail.

**1.3.PROPOSED SYSTEM**

A proposed system for machine failure prediction can be developed by the techniques of Data collection, Feature engineering, Model selection, Model training, Model evaluation, Development and monitoring. By this techniques it improves the processes, reduce downtime and increases lifespan of the equipment.

**1.4.OBJECTIVES**

The main objective of the project is to predict the failure of a machine. It helps or useful more to the companies which improves the equipment process and lifespan, which has the low error rate and high precision of accuracy.

# 2. LITERATURE SURVEY

There is a significant amount of literature available on machine failure prediction, with many studies and research papers exploring various techniques and algorithms for predicting machine failures. The area of Artificial Intelligence has been the suitable criteria to carryout predictions on the datasets by feature extraction and data pre-processing.

In a 2018 study published in the Journal of Manufacturing systems, researches used machine learning techniques to predict the remaining useful life (RUL) of a CNC machine tool. They used vibration and acoustic signals to train an SVM model, achieving an accuracy rate of 93.33%.

In a 2019 study published in the International Journal of advanced Manufacturing technology used deep learning techniques to predict the RUL of a bearing. The researches used a convolutional neural network (CNN) to analyze vibration data and achieved an accuracy rate of 99.2%.

Recently, In a 2021 study published in the journal of Intelligent Manufacturing used a hybrid model combining SVM and deep learning techniques to predict the RUL of a gearbox. The researchers achieved an accuracy rate of 95.6% by using a combination of acoustic and vibration data.

Overall, these studies and others like them demonstrate the potential for machine learning and deep learning techniques to accurately predict machine failures and improve maintenance processes. As more data becomes available and these techniques continue to evolve, machine failure prediction is likely to become an increasingly important tool for organizations looking to optimize their equipment performance and reduce costs.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SI NO DATE OF AUTHORS NAME METHADOLOGY ACCURACY  PUBLICATION | | | | | | |
| 1 | 24 July,  2017 | Zhilong Wang | Failure prediction using machine learning and time series in optical network | SVM |  | 95 |
| 2 | 21 March, 2019. | Bashi Mohammud | Failure prediction using machine learning in a virtualized HPC system and application | SVM, Random forest, KNN, classification and regression trees. |  | 90 |
| 3 | 1 June, 2015 | Seokho Kang | An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction | SVM |  | 80 |
| 4 | 11 May, 2020 | [Jiechao Gao](https://ieeexplore.ieee.org/author/37088198074) | Task Failure Prediction in Cloud Data Centers Using Deep Learning | Bidirectional Long Short Term Memory |  | 93 and 87 |
| 5 | 2013 | Bingpeng Zhu | Proactive drive failure prediction for large scale storage systems | SVM |  | 95 |
| 6 | 29th Jan,  2021 | Prasanta Kumar | Heart failure prediction using machine learning techniques | SVM, naïve bayes, logistic regression ,dission tree, KNN |  | 85.5 |
| 7 | 15 July, 2023 | Luigi Di simone | LSTM-based failure prediction for railway rolling stock equipment | Long short term memory |  | 99 |
| 8 | December 2012 | Thanyalak chalemarrewong | Failure Prediction of Data Centers Using Time Series and Fault Tree Analysis | ARMA  (Author regressive moving average) and fault tree anaylsis |  | 97 |

# 3.DATA PREPROCESSING

**3.1Dataset description**

This is the data set of machine failure prediction which consists of 10 columns and 10000 rows.

In the data set the dependent variable is target.

In the data set there are five independent variables there are:

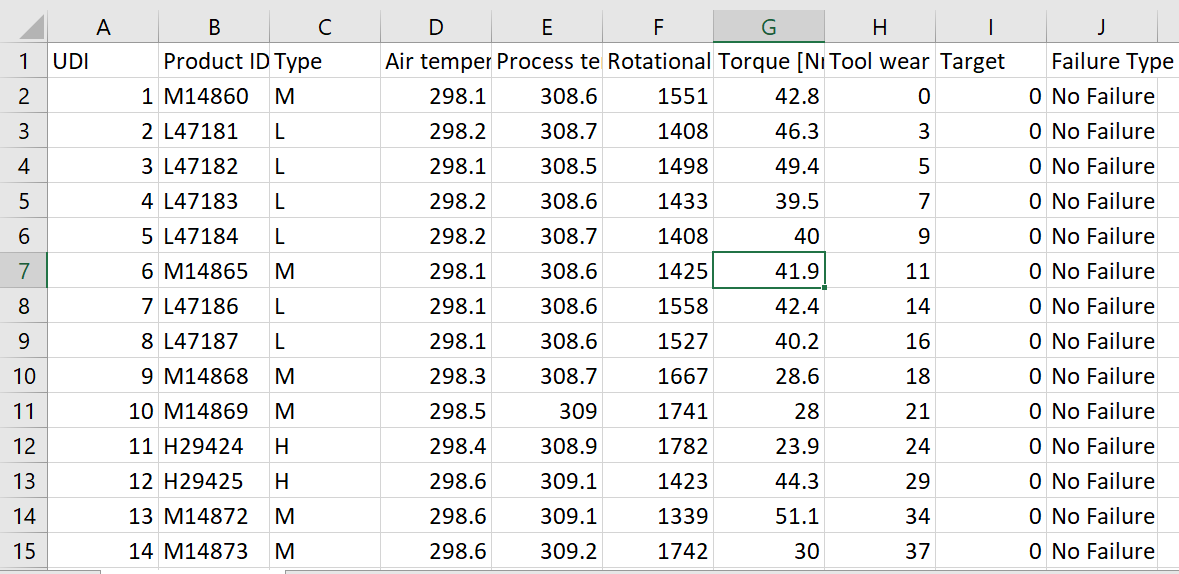
1.Air temperature

2.Process temperature

3.Rational speed

4.Torque

5.Tool wear

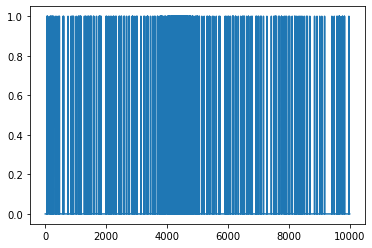


**3.2 Pre-processing through standardscaler**

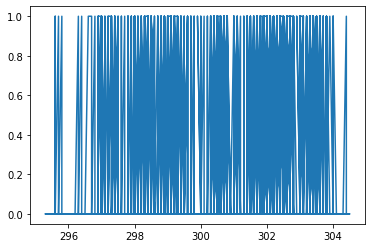
Our dataset got overfit, so we used SMOTE(synthetic minority oversampling technique) technique inorder to balance the dataset.

**3.3 Data visualization**

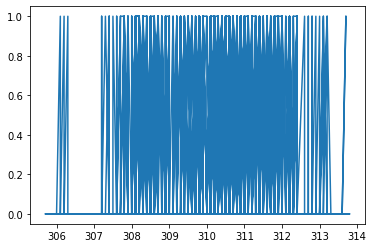
We visualized our data set through graphs. We plotted graphs between the dependent variable to independent variable. Where the dependent variable (target) is on the y-axis and the independent variables (Air temperature, Process temperature, Rational speed, Torque, Tool wear) is on the x-axis.



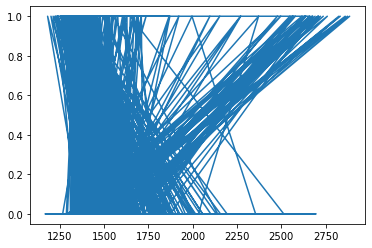
**UDI VS TARGET**



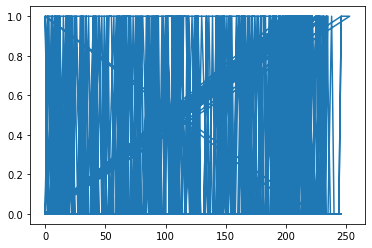
**AIR TEMPERATURE VS TARGET**



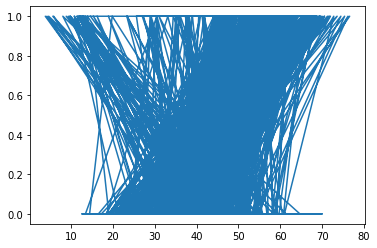
**PROCESS TEMPERATURE VS TARGET**



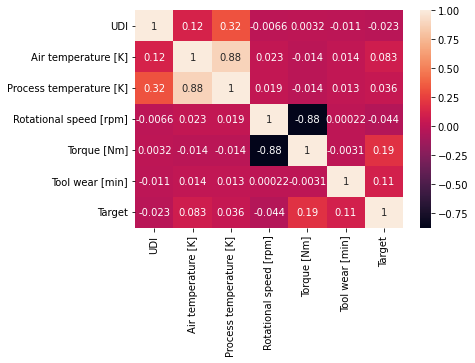
**RATIONAL SPEED VS TARGET**

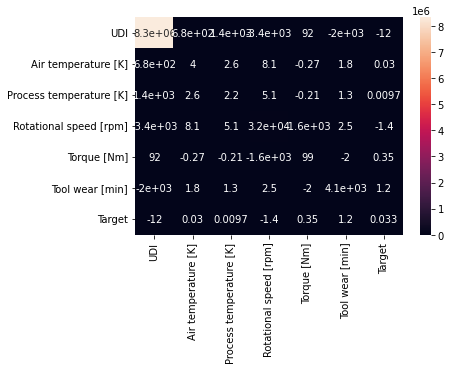
****

**TOOL WEAR VS TARGET**



**TORQUE VS TARGET**

** CORELATION MATRIX**



**COVARIANCE MATRIX**

# 4. METHODOLOGY:

After Data pre-processing and data visualization the next step is to apply the models on the dataset. Our dataset comes under supervised learning as it contains the labeled data (target variables, feature variables). First the dataset is splitted into training set and testing set. Then the model is trained on training set and then tested on testing set.

**4.1logistic regression algorithm:**

Logistic regression is a machine learning algorithm which comes under supervised learning. It is a parametric method, where an equation is formed to solve. The equation returns continues values. These continues values should to converted to categorical values.so, we use a activation function called “sigmoid”.by using log error function we calculate the error.

* from sklearn.linear\_model import LogisticRegression
* lr=LogisticRegression()
* mm=lr.fit(x\_resem\_train,y\_resem\_train)

**4.2K-Nearest Neighbor algorithm:**

K-Nearest Neighbor algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression. This algorithm is non parametric. This is also called as lazy learning algorithm. This algorithm works by first selecting the k value which is an integer value and less than the number of rows. When a new data point is given, KNN finds the nearest neighbors to that data point based on the distance using various methods like Euclidean distance or Manhattan distance. And assigns the data point to that class.

* from sklearn.neighbors import KNeighborsClassifier
* classifier=KNeighborsClassifier(n\_neighbors=5,metric='minkowski',p=2)
* classifier.fit(x\_resem\_train,y\_resem\_train

**4.3Naive Bayes algorithm:**

# Naive Bayes algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression. This algorithm is non parametric. This algorithm works based on the bayes theorem. Naive Bayes algorithm is a probabilistic classifier. It predicts the probability of an object. And also it does not require much training data.

* from sklearn.naive\_bayes import GaussianNB
* gnb=GaussianNB()
* gnb.fit(x\_resem\_train,y\_resem\_train)

# 4.4Desicion Tree algorithm:

# Decision tree algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression problems. This algorithm is also known as ID3 algorithm. This algorithm is non parametric method. It forms a tree from the given dataset. It has two nodes decision nodes and leaf nodes. Decision nodes are used for taking decisions and leaf nodes are the output of that decisions. The attribute selection happens by entropy and information Gini.

* from sklearn.tree import DecisionTreeClassifier
* classifier=DecisionTreeClassifier(criterion='entropy',random\_state=0)
* mm=classifier.fit(x\_resem\_train,y\_resem\_train)

# 4. 5support vector machine algorithm:

# Support vector machine algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression problems. SVM works by constructing a hyperplane or a line that separates the different classes of data points. SVM has support vectors. The distance between positive hyperplane and negative hyperplane is called margin.

* from sklearn.svm import SVC
* svm\_model=SVC(kernel='linear')
* svm\_model.fit(x\_resem\_train,y\_resem\_train)

**5.RESULTS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **MACHINE LEARNING MODEL** | **ACCURACY** |
| 1 | Logistic regression | 0.8235 |
| **2** | K-Nearest Neighbor | 0.9327 |
| **3** | Naive Bayes algorithm | 0.8314 |
| **4** | Decision Tree | 0.9617 |
| **5** | support vector machine | 0.8243 |

# Confusion matrix:

# 1.logistic regression:

# 

# 2.K-Nearest Neighbor

# 

# 3. Naive Bayes:

# 

# 4.Desicion Tree:

# 

# 5.support vector machine:

# 

# 6. CONCLUSION:

# Finally, machine failure prediction is the best way to ensure the failure of a machine. Predicting the failure of a machine within less time can help the users to save their time, efforts, and money in a great extent. It also improves the overall efficiency. Decision tree model gives high accuracy for this machine failure prediction dataset. KNN and decision tree confusion matrix diagonal are best among the all models.

**7.References:**

1. <https://www.sciencedirect.com/science/article/abs/pii/S0278612521001849>
2. <https://opg.optica.org/oe/fulltext.cfm?uri=oe-25-16-18553&id=370117>
3. <https://www.sciencedirect.com/science/article/abs/pii/S0951832021006694>
4. <https://www.sciencedirect.com/science/article/abs/pii/S0164121212001732>
5. <https://www.sciencedirect.com/science/article/abs/pii/S0957417415000573>
6. <https://ieeexplore.ieee.org/abstract/document/9090992>
7. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3759562>
8. https://www.sciencedirect.com/science/article/abs/pii/S095006181631068