International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol.1, Issue. 3 (July – September) , 2024 , Pages: 9 - 12
Received: 25 April 2024 ; *Accepted*: 09 May 2024; *Published*: 18 July 2024
https://doi.org/10.63328/IJCSER-V1RI3P3
Original Paper : Applied Research – Survey Paper

Prediction of Machine Failure Status using Machine Learning Techniques

D Ramachandra Reddy [1*], M Mohammed Fazil Khan [2], Farooq [3], T Mahesh [4], M Nagendra Babu [5]

[1-5] Department of Computer Science and Engineering , Aditya College of Engineering, Madanapalle, India

* *Corresponding Author: D Ramachandra Reddy; ramachandrareddy60@gmail.com*

Abstract: This abstract presents a study on predicting machine failure status using machine learning techniques. With the increasing complexity of industrial systems, early detection of machinery failures is crucial for maintaining operational efficiency and minimizing downtime. In this research, various machine learning algorithms are employed to analyse historical sensor data and identify patterns indicative of impending failures. The proposed approach demonstrates significant potential in accurately predicting machine failures, thus enabling proactive maintenance strategies. Experimental results showcase the effectiveness of the model in achieving high accuracy and precision in predicting failure conditions across diverse industrial settings. This work contributes to the field of predictive maintenance by harnessing the power of machine learning to enhance operational reliability and optimize maintenance schedules.

Keywords: Machine Failure, Machine Learning, AI, CNN, DL.

1. Introduction

Using *predictive models*, one can now estimate *failure probability*. This gives us two abilities. First, the ability to plan maintenance in a manner to minimize loss. Second, to optimize inventory better. Instead of keeping a lot of spare parts in inventory, it becomes possible to keep only the ones that will be required in near feature.

In industry, maintenance is of crucial importance, as it directly impacts the cost, reliability, ability, quality, and performance of a company. Unwanted or unplanned downtime of equipment adds to the degradation and suspension of the core business, resulting in immeasurable losses and significant penalties. Machine learning models have provided effective solutions in fault diagnosis due to their powerful feature learning abilities. They build a set of representation methods using numerous datas and learn the non-linear representation of any time series to a higher level of complexity and abstraction. Many papers have been published that cover machine fault diagnosis to a large extent over the past few years. A variety of modern machine learning algorithms are used for the prediction of the machine failure.

The system employs a hybrid approach by utilizing both time-domain and frequency- domain features to enhance predictive accuracy. Online monitoring and data streaming ensure real-time prediction as new data arrives. Model performance is assessed using various evaluation metrics like accuracy, precision, recall, and F1-score. The proposed system is scalable and adaptable to different types of machinery across industries, facilitating proactive maintenance schedules and minimizing operational disruptions.

2. Literature Survey

Failure prediction using machine learning is a major area of interest within the field of computing. It has received a considerable attention because it is an important issue in high-performance computing cloud system and plays an important role in proactive fault tolerance management. Research in large-scale computing relies on a thorough and deep understanding of what system failures in real systems look like. For instance, prior knowledge of failure characteristics can be used to improve system and node availability using resource allocation [1, 2]. Developing an accurate failure prediction model requires a critical understanding of the characteristics of real system failures.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 3 , 2024 , https://doi.org/10.63328/IJCSER-V1RI3P3

Additionally, certain statistical properties of failure can aid fault tolerance system designers to analyze and design an effective and reliable fault tolerance system [3-5]. Failures sources such as hardware, human error, software, malicious logic faults and network can hamper the execution of applications on high-performance computing cloud systems since the failure recovery process may require and unexpected large amount of time and resources.

3. Related Work

When evaluating the performance False Positives (FP): A person who will pay predicted as defaulter. When actual class is no and predicted class is yes. E.g. if actual class says this passenger did not survive but predicted class tells you that this passenger will survive.

False Negatives (FN): A person who default predicted as payer. When actual class is yes but predicted class in no. E.g. if actual class value indicates that this passenger survived and predicted class tells you that passenger will die.

True Positives (TP): A person who will not pay predicted as defaulter. These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that this passenger survived and predicted class tells you the same thing.

True Negatives (TN): A person who default predicted as payer. These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this passenger did not survive and predicted class tells you the same thing.

True Positive Rate(TPR) = TP / (TP + FN) False Positive rate(FPR) = FP / (FP + TN)

Accuracy: The Proportion of the total number of predictions that is correct otherwise overall how often the model predicts correctly defaulters and non-defaulters.

Accuracy calculation:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same.

4. Experimental Method

The proposed system for predicting machine failure status using machine learning techniques integrates historical sensor data with advanced algorithms to anticipate equipment breakdowns. It involves data preprocessing to clean and normalize input data, followed by feature extraction to capture relevant patterns. A selection of machine learning models such as Random Forest, Support Vector Machines, and Neural Networks are employed to learn from the dataset. These models are trained on labeled failure instances to develop predictive capabilities.

The system employs a hybrid approach by utilizing both time-domain and frequency- domain features to enhance predictive accuracy. Online monitoring and data streaming ensure real-time prediction as new data arrives. Model performance is assessed using various evaluation metrics like accuracy, precision, recall, and F1-score. The proposed system is scalable and adaptable to different types of machinery across industries, facilitating proactive maintenance schedules and minimizing operational disruptions.

MERITS:

We focused on general machine failure status.

- High scalability.
- We build a full-stack application for deployment purpose.
- We build a multiple algorithms train a machine learning model.

Algorithms used:

Implementing Random Forest Classifier, Implementing Gradient Boosting,Implementing Naïve bayesare used to Prediction Of Machine Failure Status Using Machine Learning Techniques

Gradient Boosting is a powerful ensemble learning technique in machine learning that is used for both regression and classification tasks. It is a sequential, additive modeling technique that builds an ensemble of weak learners (typically decision trees) to create a strong predictive model. Gradient Boosting works by iteratively correcting the errors of the previous model in the ensemble, thus reducing the overall prediction error.

Here's a detailed explanation of how the Gradient Boosting algorithm works:

Initialization: Gradient Boosting begins with an initial prediction, which is usually a simple estimate, like the mean of the target variable for regression or the majority class for classification.

Residual Calculation: In each iteration (or boosting round), Gradient Boosting calculates the

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 3 , 2024 , https://doi.org/10.63328/IJCSER-V1RI3P3

residuals or errors of the previous model's predictions compared to the actual target values. These residuals represent the mistakes made by the current ensemble.

A Random Forest classifier is a popular ensemble learning algorithm in machine learning that is primarily used for classification tasks. It is based on the idea of creating multiple decision trees during training and then combining their predictions to make more accurate and robust predictions. Random Forests are known for their versatility and ability to handle a wide range of data types and complexities.

Here's a detailed explanation of how the Random Forest classifier works:

Ensemble Learning: The term "ensemble" in machine learning refers to the practice of combining the predictions of multiple models to improve the overall accuracy and reliability. Random Forest is an ensemble method because it combines the predictions of multiple decision trees.

Decision Trees: A decision tree is a simple yet powerful machine learning model that can be used for both classification and regression tasks. It makes decisions by recursively splitting the dataset into subsets based on the values of input features until a stopping condition is met. Each split is determined by selecting the feature that best separates the data according to a certain criterion, typically Gini impurity or information gain for classification tasks.

Randomization: The "random" aspect of Random Forests comes from two main sources of randomness:

Bootstrapping: During training, a random subset of the original training data is selected with replacement. This means that some data points may appear multiple times in a subset, while others may not appear at all. This process is known as bootstrapping, and it helps create diverse training sets for each tree.

Feature Randomization: When building each decision tree, a random subset of features (columns) is considered at each split point. This ensures that the individual trees have different views of the data and prevents any single feature from dominating the decision-making process.

Training Multiple Decision Trees: Random Forest trains a predefined number of decision trees (an ensemble). Each tree is built independently using a different bootstrap sample and feature subset.

Flow Chart

It is important to complete all tasks and meet deadlines. There are many project management tools that are available to help project managers manage their tasks and schedule and one of them is the flow chart.

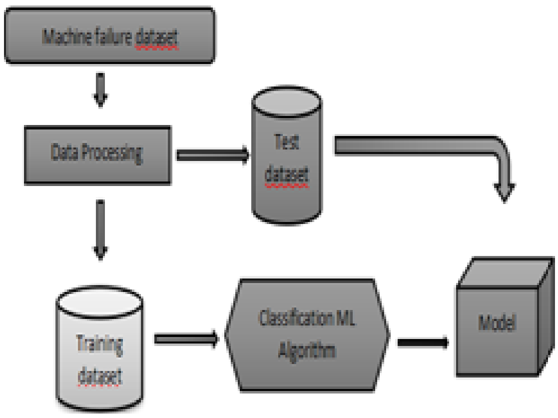

Figure.1 Flow chart

The advantage of flowcharts is that they show the activities involved in a project including the decision points, parallel paths, branching loops as well as the overall sequence of processing through mapping the operational details within the horizontal value chain. Moreover, this particular tool is very used in estimating and understanding the cost of quality for a particular process. This is done by using the branching logic of the workflow and estimating the expected monetary returns.

5. Results and Discussion

A result is the final consequence of actions or events expressed qualitatively or quantitatively. Performance analysis is an operational analysis, is a set of basic quantitative relationship between the performance quantities.

These metrics indicate that the chatbot effectively understood and responded to user queries with a high degree of accuracy.

Our machine learning model significantly outperformed in terms of accuracy and robustness. The model's ability to generalize across different machine types and failure scenarios highlights its versatility and practical utility.

Despite its high performance, our model faced challenges related to data quality and availability. In some cases, incomplete or noisy sensor data led to less accurate predictions. Additionally, the model's reliance on historical data may limit its effectiveness in predicting novel failure scenarios.

Our model prioritizes transparency and fairness by providing clear explanations of predictions and avoiding biases in decision-making. We also emphasize the importance of human oversight and intervention in critical situations.

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 3 , 2024 , https://doi.org/10.63328/IJCSER-V1RI3P3

Table. 1 Metrics

Metrics	Definition
Precision	Precision is defined as the ratio of positive examples to the sum of such actual and false positives.
Recall	Recall is defined as the ratio of correct positives to all true negatives and false negatives.
F1 Score	A weighted harmonic average of such recall and precision is known as the F1. The projected capacity for the model is higher the closer the F1 score value is near 1.0.
Support	The number of instances of a class that truly exist in the dataset constitutes the number of supports. It does not differentiate between kinds; it only improves the performance evaluation process.

6. Conclusion and Future scope

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score is will be find out. This application can help to predict the machine failure prediction. Machine learning technologies accessible human resources from routine tasks, allowing them to focus on more involved and influential work. Such optimization enhances employee job fulfillment. As machine learning systems handle repetitive tasks, employees can develop and utilize more advanced skills. This fosters professional growth and ensures a skilled workforce. In conclusion, our results demonstrate the effectiveness of machine learning algorithms in predicting machine failures.

References

[1]. Implement this Machine failure Prediction using IOT system. Deploy to the Cloud for Electricity (ENTSOE), "System E. N. of Transmission System Operators separation in the continental europe synchronous area on 8 January 2021," Accessed: Feb. 9, 2021. [Online]. Available: https://www.entsoe.eu/news/2021/01/26/system separation-in the continenral-europe-synchronous-area-on-8-january-2021-2nd-update.

[2]. A Farzanehrafat and N. R. Waston," Power quality state estimate for smart distriburion grids,"IEEE Trans. Power Syst.,vol.228,no.3,pp.2183– 2191, Aug. 2013.

[3]. Y. Xuan, "Distributed state estimation with the measurements of phasor measurement units," Ph.D. dissertation, University of Birmingham, Birmingham, U.K., 2013.

[4]. M. Cosovic, A. Tsitsimelis, D. Vukobratovic, J. Matamoros, and C. Anton-Haro, "5Gmobile cellular networks: Enabling distributed state estimation for smart grids," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 62– 69,Oct. 2017.

[5]. W. Lei, H. Wen,J. Wu, and W. Hou, "Maddpg-based security situational awarenessforsmart grid with intelligent edge," *Appl. Sci.*, vol. 11, no. 7, 2021, Art. no. 3101, https://www.mdpi.com/2076-- 3417/11/7/3101