JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING TECHNIQUES FOR THE IDENTIFICATION OF MACHINERY FAULTS THROUGH VIBRATION SIGNALS

A. Naga Saikumar[1], K. Venkata Rao[2], CH. Lakshmi Kanth[3]

[1]P.G Student, Department of Mechanical Engineering, Prasad V Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada, Andhra Pradesh, India

[2] Assistant Professor, Department of Mechanical Engineering, Prasad V Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada, Andhra Pradesh, India

[3]Assistant Professor, Department of Mechanical Engineering, Prasad V Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada, Andhra Pradesh, India

Abstract: For industrial systems to remain operationally efficient and minimize downtime, the diagnosis of mechanical problems is essential. Heuristic principles and manual inspection are frequently used in traditional fault detection techniques, which can be error-prone and time-consuming. New developments in machine learning (ML) provide viable substitutes by facilitating automatic and more precise fault diagnosis. The techniques for automatically identifying misalignment, unbalancing, looseness, and bearing wear in rotary machines are covered in this article. The defects in a rotating system are modelled using the finite element approach. In the steady-state operation, the modelled system then functions practically under various situations; the vibrational responses are computed numerically. To guarantee that equipment is operating well, regular monitoring is required. Condition monitoring is used to predict any problems by using several techniques mostly with vibration signals. The vibration signal technique is used in this study to predict defects in rotating machinery. A vibrometer is used to collect data from the rotating machine, which is then processed in Python using the Jupiter Notebook IDE. Based on the gathered data, predictive models like Support Vector Machine, Naive Bayes, and Logistic Regression are used to predict problems.

Keywords: Condition Monitoring, Classification algorithms, Fault Diagnosis, Data Preprocessing, Predictive analysis.

1.INTRODUCTION

LP N Saavedra, D E Ramirez [1] investigated that the variation in coupling stiffness upon rotation is what causes the vibration caused by shaft misalignment; the forcing frequencies produced are harmonics of the shaft's speed and are directly correlated with the frequency of the coupling stiffness variation. The effectiveness of traditional vibration analysis rules for identifying shaft misalignment in practical predictive maintenance is assessed.

Vaggeeram Hariharan and PSS. Srinivasan. [2] investigated that the rotor dynamic test device was used for experimental research in order to forecast the vibration spectrum for shaft misalignment. It is discovered that the vibration amplitudes caused by the shaft parallel misalignment are reduced by 85-89% when employing the newly designed flexible coupling.

Joseph P. Spagnol, Helen Wu, and Chunhui Yang [3] investigated that the weight-dominant breathing model was deemed inadequate for simulating the vibration of fractured rotors with high allowable residual unbalance due to the notable disparities between the two models' breathing mechanisms and vibration results.

F. Al-Badour M. Sunar L. Cheded [4] investigated that the vibration analysis is a key technique for monitoring the structural health; one new and potent tool in this field is time-frequency analysis, which includes the wavelet transform. Commonly used spectrum-based signal analysis methods, including the rapid Fourier transform, are effective in identifying a wide range of vibration-related issues in rotating machinery.

T.C. Tsai Y, Z. Wang [5] investigated that the verified through comparison with published experimental data that has already been published. By contrasting the fundamental mode forms of the shaft with and without a crack, the location of the crack can be estimated. Moreover, the difference in the natural frequency of the shaft with and without a crack can be used to determine the depth of the crack.

A Khadersab S Sivakumar [6] investigated that the vibration condition monitoring is a widely used maintenance approach that provides an accurate assessment of the rotating machinery's state of health. Furthermore, a precise evaluation of the bearing-related failure of rotating machinery will result from this experimental study.

Thomas G. Dietterich [7] here suggested categorizing algorithms according to the ways in which they learn. This classification divides algorithms into three primary categories: Learning algorithms under supervision: These algorithms are trained on data that has been labelled. Support vector machines, logistic regression, and decision trees are a few types of supervised learning algorithms. Algorithms for unsupervised learning: These are those that acquire knowledge from unlabeled data.

Asoke K. Nandi [8] has suggested various categories for the algorithms used in the field of signal processing and has made substantial contributions to it., supervised algorithms need labelled data, where each sample is linked to a predetermined result. The learning of a mapping between input and output data is the aim of supervised learning. Application areas for frequency-domain algorithms include audio signal processing, image processing, and voice recognition. These algorithms function on a signal's frequency spectrum.

Fatah, K. S., & Mahmood [9] suggested A response variable with two or more categories is linked to a collection of continuous or categorical predictor variables using a probability function, as explained by logistic regression analysis. This method's primary flaw is that it uses the Newton Raphson Method to estimate logistic parameters numerically via likelihood estimation.

T.M. Cover and J.A. Thomas [10] investigate that overarching objective is to comprehend the features of the data that influence naive Bayes' performance. With the use of Monte Carlo simulations, our method makes it possible to examine categorization accuracy methodically for a number of classes of randomly produced issues. [10]

Bennett, P. N [11] suggested that Naive Bayes can still be optimal if the dependences distribute uniformly across classes or if they cancel each other out, regardless of how strong the dependences are among attributes. We put up and establish sufficient and necessary requirements for naive Bayes to be optimal. We also look into whether naive Bayes is best under the Gaussian distribution. We introduce and demonstrate a sufficient condition for naive Bayes' optimality, wherein the attributes' reliance is observed. This shows that an attribute's dependence on another attribute may neutralize it.

S. B. Kotsiantis [12] investigated and explains the theory and applications of these decision trees. To build decision trees using these techniques, tree split algorithms are needed, and quantitative metrics are needed to train an efficient tree. As a result, the chapter discusses metrics including information gain, entropy, cross-entropy, and Gini impurity. [12]

2.DESCRIPTION OF EXPERIMENTAL SETUP AND ITS METHODS

This setup consists of induction motor, couplings, shafts, and bearings.

Couplings

Motor

Bearings

The experimental setup consists of a one horsepower AC induction motor running at 2880 rpm, driving a rotor shaft 1000 mm long with a 16 mm diameter and supported by a single ball bearing. The shaft has a 220 mm diameter disc at the bearing end. The motor's speed can be adjusted between 600 and 2880 rpm using a variable frequency drive (VFD). Vibration velocity at both the non-drive end (NDE) and drive end (DE) is measured using an FFT analyzer.

Fig.1. Photograph of Experimental setup

2.1 MACHINE LEARNING

A subset of artificial intelligence known as machine learning algorithms allows computers to learn from data and make judgments or predictions without the need for explicit programming.

These are a few of the most often used algorithms for machine learning:

Linear Regression: Based on one or more input variables, linear regression is a supervised learning technique that predicts a continuous output variable. The goal is to establish a linear relationship between the input and output variables.

Logistic Regression: Based on one or more input variables, logistic regression is a supervised learning technique that predicts a binary output variable (yes or no, true or false, etc.). It looks for a connection between the input variables and the likelihood that the output variable will be true.

Decision Trees: Suitable for both regression and classification applications, decision trees are supervised learning algorithms. Based on input variables, they construct a model of decisions and potential outcomes that resembles a tree.

Random Forests: Using several decision trees, random forests are a supervised learning approach that raises prediction accuracy. They construct decision trees using subsets of the input variables that they have randomly chosen.

Support Vector Machines (SVMs): For classification problems, SVMs are a supervised learning algorithm. They look for the most effective border—known as a hyperplane—between several data types.

Neural Networks: Modelling the functioning of the human brain, neural networks are supervised learning algorithms. They are made up of networked nodes, or neurons, that can learn from data and make predictions.

K-Nearest Neighbors (KNN): For classification and regression problems, KNN is a supervised learning technique. Based on the k-nearest neighbors in the training dataset, it classifies an input data point. **Xgboost:** It appears that you are referring to "XgBoost," a well-liked machine learning method that excels at processing structured data. The term "extreme Gradient Boosting"(XgBoost). Regression and classification problems make extensive use of it.

2.2 FLOW CHART

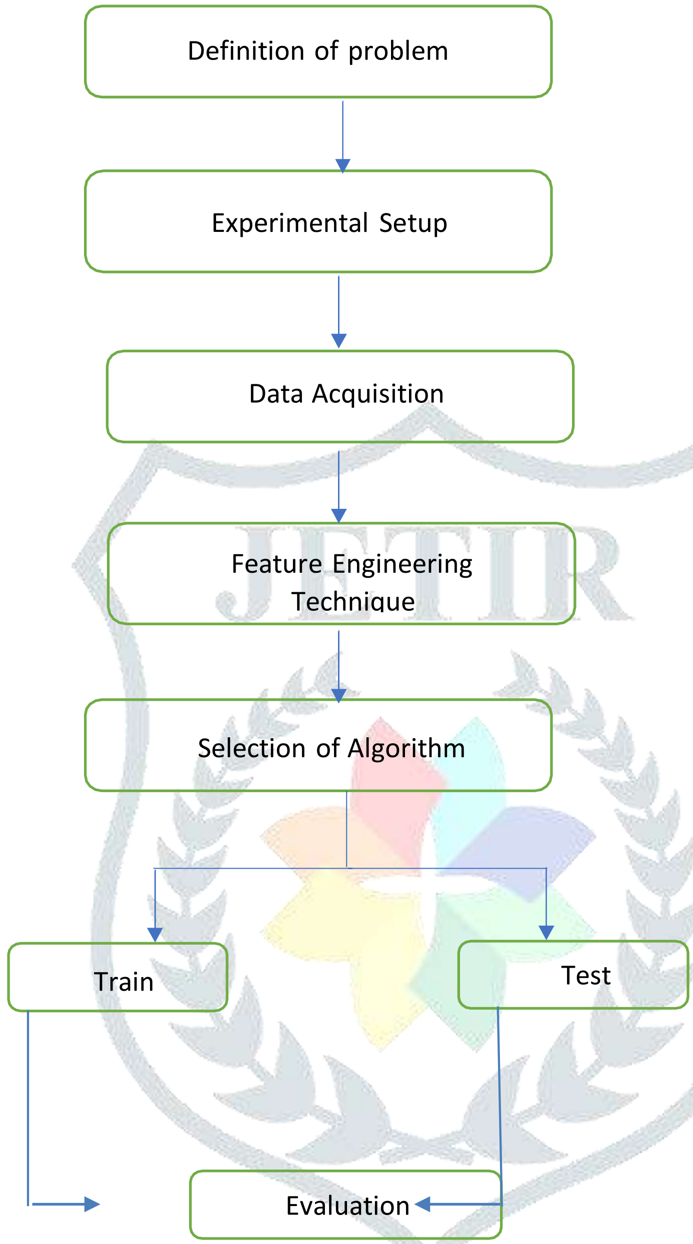


Fig 2.1 Process Flow Chart

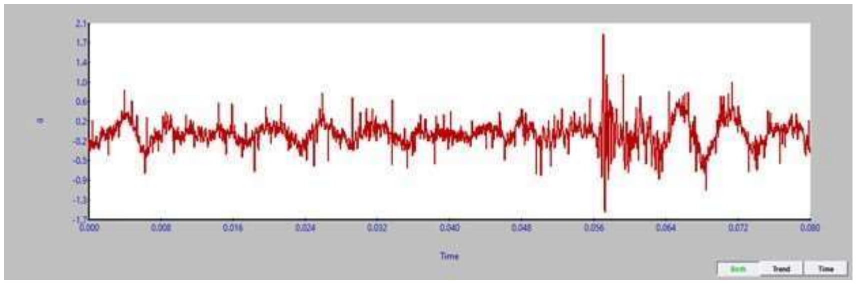
Prior to figuring out how to solve a problem, firstly put up an experiment to test several scenarios. Data acquisition is the process by which we obtain certain data after completing the necessary prerequisites. As we take data out of the experimental configuration under different circumstances. Next, we employ the feature engineering technique to determine which features are most crucial and helpful for our project. Once the key features are obtained, we choose an algorithm that is crucial in identifying the issue based on those features. Once the issue has been located, its causes, such as misalignment, looseness, or unbalance, are known to us. Once these causes are understood, it is simple to find a remedy.

2.3 VIBRATION GRAPHS

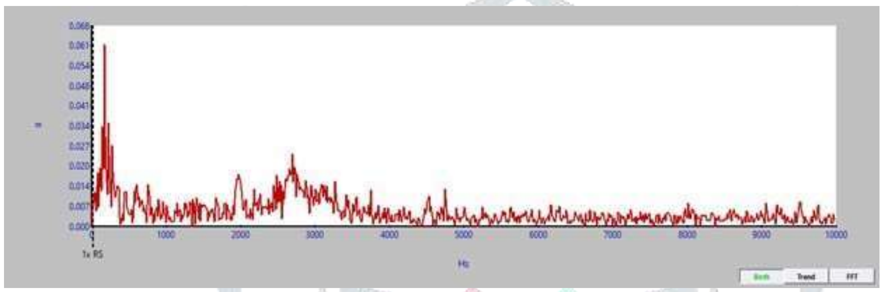
Taken waveform and spectrum vibrational data by using Wilcoxon FFT vibrometer at an 900 RPM at non drive end in radial direction.

Below are the graphs obtained from meta data software

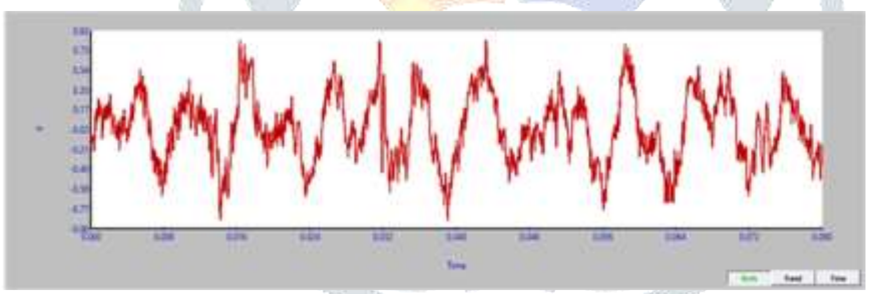
Vibration Graphs of 900 rpm Normal Condition-1 at Non-Drive End Waveform



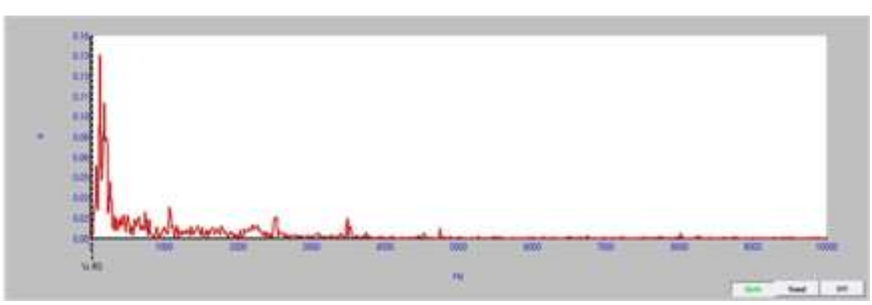
Vibration Graphs of 900 rpm Normal Condition-1 at Non-Drive End Spectrum



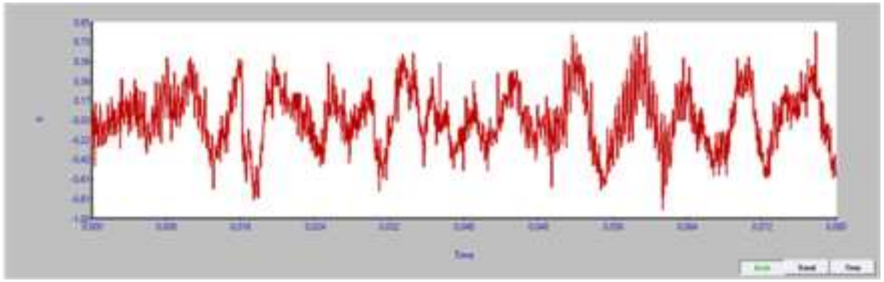
Vibration Graphs of 900 rpm Unbalance Condition-1 at Non-Drive End Waveform



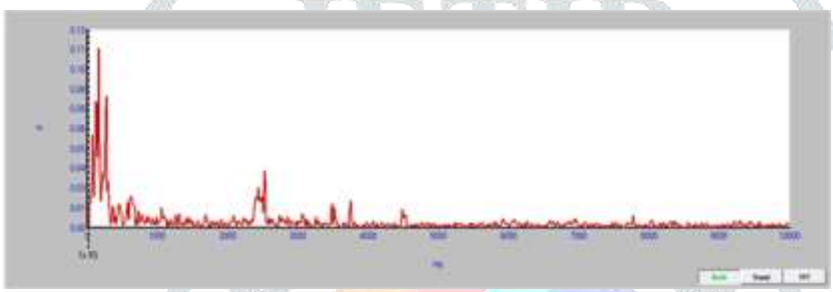
Vibration Graphs of 900 rpm Unbalance Condition-1 at Non-Drive End Spectrum



Vibration Graphs of 900 rpm Looseness Condition-1 at Non-Drive End Waveform



Vibration Graphs of 900 rpm Looseness Condition-1 at Non-Drive End Spectrum



2.4 Histograms

This graph shows that no data is missing; in the event that data is interrupted during the process of obtaining readings, the data will be indicated as dotted irregularities. Only the missing or complete data is displayed in this graph.

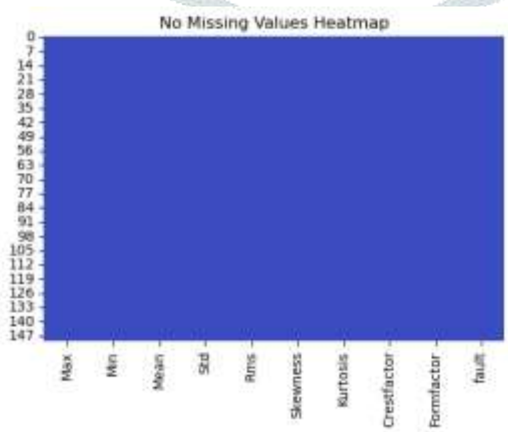


Fig 2.2 No missing Values Heatmap

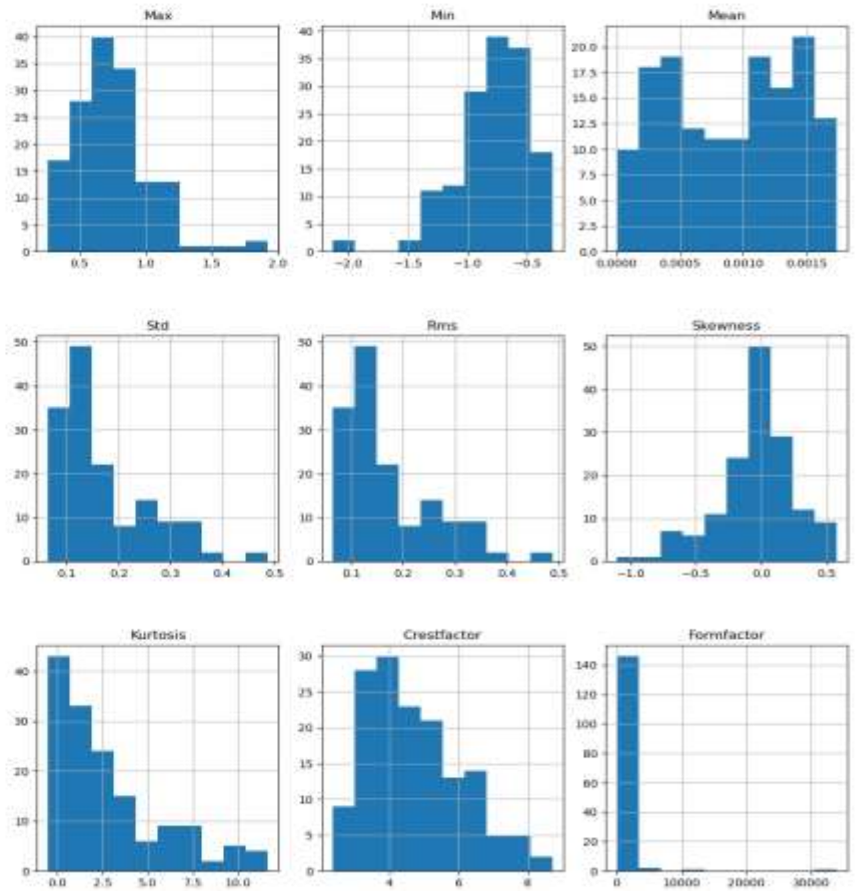


Fig 2.3 Density plots

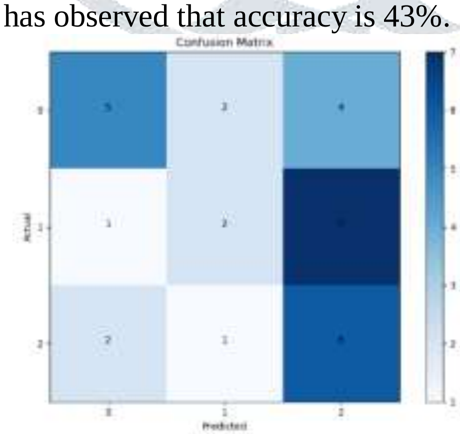
3 Results and Discussions

3.1 Confusion Matrix

One essential tool in the field of machine learning and classification problems is a confusion matrix. It's a table meant to assess a categorization model's performance. It gives insight into how well the model is doing in terms of several metrics including accuracy, precision, recall, and F1-score and enables the visualization of an algorithm's performance.

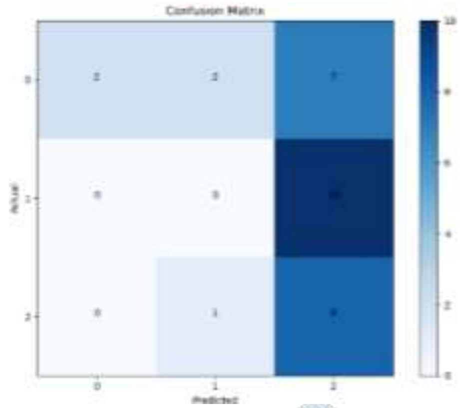
3.1.1 LOGISTIC REGRESSION CONFUSION MATRIX

From the below confusion matrix, it has observed that accuracy is 43%.



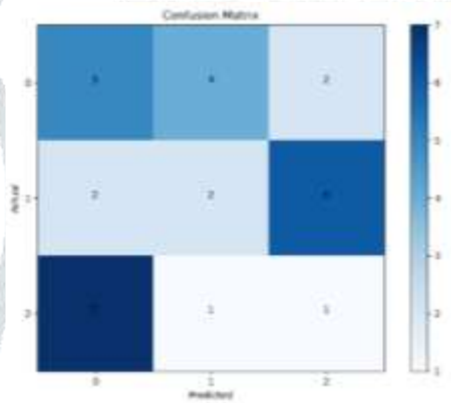
3.1.2 SUPPORT VECTOR MACHINE CONFUSION MATRIX

From the below confusion matrix, it has observed that accuracy is 33%



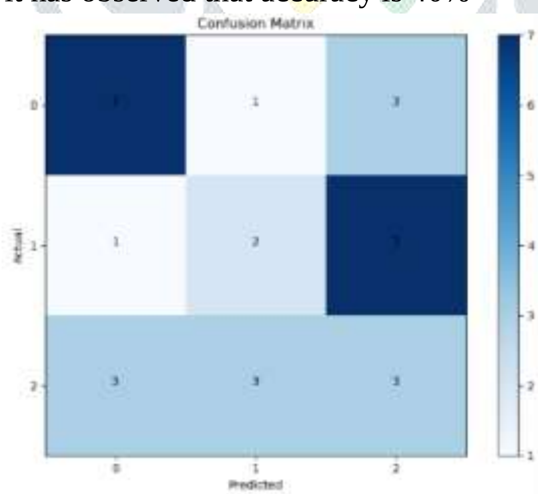
3.1.3 DECISION TREES CONFUSION MATRIX

From the below confusion matrix, it has observed that accuracy is 26%



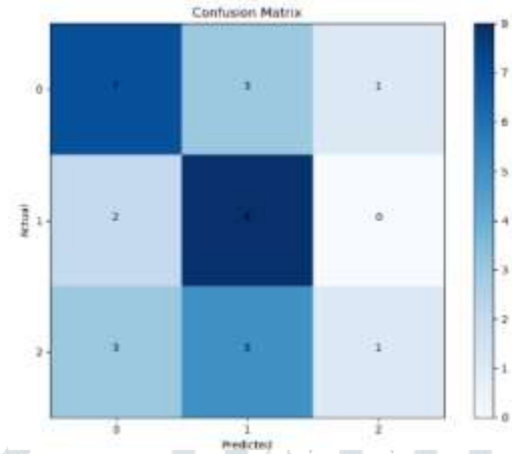
3.1.4 RANDON FOREST DECISION MATRIX

From the below confusion matrix, it has observed that accuracy is 40%



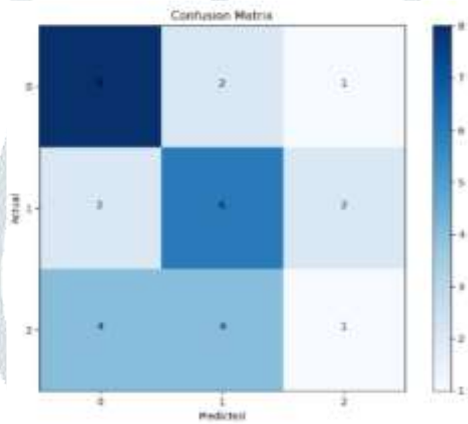
3.1.5 GAUSSIAN NB DECISION MATRIX

From the below confusion matrix, it has observed that accuracy is 53%



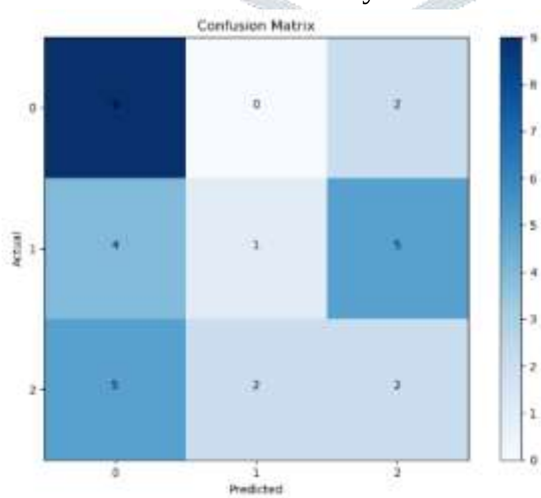
3.1.6 KNIGHBORS CLASSIFIER DECISION MATRIX

From the below confusion matrix, it has observed that accuracy is 50%



3.1.7 XG BOOST DECISION MATRIX

From the below confusion matrix it has observed that accuracy is 40%



3.2 Results Table

From the below result table, it is observed that the better results achieved for Logistic Regression, Gussian NB and KNighbore Classifier obtained with 43%,53% and 50% respectively. It was also noted that the Decision tree poor accuracy (26%), more research was necessary. Based on observation that the outlier values have impacted the precision and accuracy. Removing the outlier values by using different techniques, will get more precision and accuracy.

Algorithm	Precision	Recall	Accuracy
Logistic Regression	0.468	0.433	43%
Support Vector Machine	0.463	0.333	33%
Decision Tree	0.260	0.267	26%
Random Forest	0.414	0.4	40%
Gaussian NB	0.531	0.533	53%
KNighbors Classifier	0.45	0.5	50%
XGBoost	0.36	0.4	40%

Table 3.1 Results

CONCLUSIONS

Vibration analysis has been successfully investigated experimentally, and graphs created using the DATA MATE program validate the analytical shaft.

The following observations were made while working on this project:

- Waveform and spectrum forms show generalization vibration signals related to fault and machine conditions.
- Using Python code in the Jupiter IDE, statistical features were retrieved from the waveform under conditions at 900 rpm.
  - Minimum
  - RMS
  - SD
  - Max
  - Formfactor
  - Kurtosis
  - Crest factor
  - Mean
  - skewness
- Recognize the statistical characteristics present in both univariant and multivariant graphs.
- The linear classification algorithms, such as Naïve Bayes and logistic regression.
- Models were created for the non-linear algorithms XGBoost, Random Forest, Decision Tree, KNN, and Ensemble Algorithm.
- Achieved better results for Logistic Regression, Gussian NB and KNighbore Classifier obtained with

43%,53% and 50% respectively

- Reached the maximum accuracy of 53% using the Gussian NB algorithm.
- It was also noted that the Decision tree has poor accuracy (26%), more research was necessary.
- XGBoost, Random Forest, and Logistic Regression. It was advised to apply the XGBoost.
- By using vibration signals to identify machinery flaws, spot inspection of classification algorithms was carried out.

References:

[1] LP N Saavedra, D E Ramirez (Vibration analysis of rotors for the identification of shaft misalignment Part 1: Theoretical analysis, September 2004).

[2] Vaggeeram Hariharan and PSS. Srinivasan. (Vibration analysis of parallel misaligned shaft with ball bearing system 1 September 20).

[3] Joseph P. Spagnol, Helen Wu, and Chunhui Yang. (Vibration Analysis of a Cracked Rotor with an Unbalance Influenced Breathing Mechanism January 2018).

[4] F. Al-Badour M. Sunar L. Cheded. (Vibration analysis of rotating machinery usingtime- frequency analysis and wavelet techniques).

[5] T.C. Tsai Y, Z. Wang. (The vibration of a multi-crack rotor September 1997).

[6] A Khadersab S Sivakumar Dr. (Vibration Analysis Techniques for Rotating Machinery and its effect on Bearing Faults 2018).

[7] Thomas G. Dietterich (Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms October 1998)

[8] Asoke K. Nandi (Algorithms for automatic modulation recognition of communication signals April 1998)

[9] Fatah, K. S., & Mahmood, R. F. (2016). Parameter Estimation for Binary Logistic Regression Using Different Iterative Methods. Journal of Zankoy Sulaimani, Vol 19 No 2: 177-178.

[10] T.M. Cover and J.A. Thomas. Elements of information theory. New York: John Wiley & Sons, 1991.

[11] Bennett, P. N. 2000. Assessing the calibration of Naive Bayes' posterior estimates. In Technical Report No. CMUCS00-155.

[12] S. B. Kotsiantis. "Supervised machine learning: A review of classification techniques," Informatica 31, pp. 249–268, 2007.