Predicting Asset Maintenance Failure Using Supervised Machine Learning Techniques

Gregory Opara, Johnwendy Nwaukwa, Felix Uloko, Clinton Oborindo

Abstract— **Maintenance activities can be broadly divided into three major categories and are corrective, preventive, and predictive maintenance. Our research focused on condition monitoring which is a form of predictive maintenance for brake pad failure for heavy-duty vehicles asset. The failure of a machine can stop production and cause a huge number of losses of money and people, moreover, it may take several months to order a new one. At the same time, excessive maintenance actions may slow production. Existing works of literature on predicting maintenance were studied in this research. Different machine learning techniques have been used for predicting maintenance, and to the best of our knowledge, Neural Network was only used for the prediction of the brake pad failure. Neural Network makes accurate prediction if the dataset is very large and also consume a lot of computational power. However, due to the fact that the problem is a classification problem, it is a necessity to carry out performance check of the best supervised model for the dataset downloaded from GitHub. Gaussian Naïve bayes, Decision tree and K-Nearest Neighbour are used to check for accuracy of our dataset. The dataset was divided into training and testing data where the training data has larger rows than the testing data. We then compared the performances of the selected supervised algorithm. Python is the preferred language used in this research. For our result, we showed that Decision tree performed well more than Gaussian Naïve bayes and K-Nearest Neighbour with an accuracy of 95%.**

Index Terms— **Gaussian Naïve bayes, Neural Network, K-Nearest Neighbor, and Decision tree.**

I. INTRODUCTION

Technology has been integrated into asset management for decades with a variety of uses [1]. Given the evolution of technology broadly in society, it is not surprising that there continue to be new technologies and use-cases in asset management [2].

The introduction of Artificial Intelligence (AI) technology in asset management is being heralded as an opportunity to streamline the creation of more targeted and bespoke outcomes for organizations [3], but the reality is there is still a considerable disparity in the interpretation of what AI can deliver versus its use in practice [4].

Assets are sometimes described as tangible because they generally have some physical existence [5].In most cases, assets are usually expected to bring a long-term profit to the company as a result of their function [6]. At the same time,

Gregory Opara, Computer Science, Veritas University, Abuja, Nigeria
Johnwendy Nwaukwa, Computer Science, Veritas University, Abuja, Nigeria
Felix Uloko, Computer Science, Veritas University, Abuja, Nigeria
Clinton Oborindo, Computer Science, Petroleum Training Institute, Delta, Nigeria

they have a finite lifespan, because their usefulness usually declines over the years as a result of wear and usage [7].

To get the most out of these assets over their lifespan, companies must be able to accurately track the depreciation value these assets bring, as well as accurately manage preventive maintenance [6].

Predicting Maintenance is one of the early adopters of enormous information investigation and AI and IoT (Internet of Things) since it is so easy to consider and execute for that utilization case. Computing when a machine needs support is an issue that fits conveniently into a prescient calculation. This is because machine wear is a function of time and usage. With the development of manufacturing technology and mass production, the methodology of maintenance scheduling and management has become an important topic in the industry [8]. Maintenance activities can be broadly divided into three major categories and are Corrective, Preventive, and Predictive Maintenance [9]. Corrective maintenance refers to repair work after an equipment outage occurs [9]. Preventive maintenance is based on certain periodic intervals to prevent equipment failure before it occurs [9]. Predictive maintenance relies on engineering tools and statistical analysis to process the data and analyze the health condition of the equipment [10]. In this research, we will focus on condition monitoring which is also a form of predictive maintenance which relies on equipment sensors data to predict asset failure.

XYZ development organization specializes in leasing weighty IOT prepared trucks that send information from vehicles utilizing a phone or satellite sign either as a stream. Trucks are fitted with sensors and GPS trackers that action heat, vibration, distance voyaged, speed, and so forth. These sensors are appended to the motor, brakes, transmission, etc. XYZ organization study this information to work their vehicles in the most secure and least expense way conceivable.

II. STATEMENT OF THE PROBLEM

The failure of a machine can stop production and cause a huge number of losses of money and people. Moreover, it may take several months to order a new one. At the same time, excessive maintenance actions may slow production [11].

Existing works of literature on predicting maintenance were studied in this research. Different machine learning techniques have been used for predicting maintenance [13],[15],[16],[17], and to the best of our knowledge, Neural Network was only used for the prediction of the brake pad failure [23]. Neural Network make accurate prediction if the dataset is very large and also consume a lot of computational power. However, due to the fact that the problem is a classification problem, it is a necessity to carry out performance check of the best supervised model.

Given attributes which are speed measure KM, heat measure H, vibration measure V and Pressure P which represents X, the target variable (or output), Y, can take only discrete values for a given set of features (or inputs) X. The ideal goal is to predict brake failure for the output Y, which is a classification problem.

III. RESEARCH OBJECTIVES

The objectives for this research are
   I. To specify the requirements model for the predictive model
  II. To design the predictive model.
 III. To predict supervised machine learning algorithm performance for our brake pad dataset.

IV. RELATED WORKS

Studies related to predicting maintenance are examined to identify a gap concerning the concept of predictive maintenance within the literature. We reviewed some of the important kinds of literature on predictive maintenance for this research. We were able to collect works of literature from various authors that used different machine learning approaches for predicting maintenance failure.
[11] presented a general methodology to improve risk assessment in the specific workshops of semiconductor manufacturers. They were concerned, with the problem of equipment failures and drifts. They observed that failures are generally, concerned with delay, during the product metrology phase. To improve the reactivity of the control system, they proposed a predictive approach based on the Bayesian technique.
[12] presented an approach that used timed hybrid automata of the machine's normal behavior for predictive maintenance of industrial plants. This hybrid model reduces discrete and continuous signals (e.g., energy data) to individual states, which refer to either the present condition of the machines.
[13] reviewed different techniques of maintenance, ArtificialNeuralNetworks (ANNs), and their various applications in fault risk assessment and an early fault detection analysis. The predictive maintenance is focused on production facilities supplying in long supplier chains of the automotive industry to ensure reliable and continuous production and on-time deliveries.
[14] explored the use of Autoregressive Integrated Moving Average (ARIMA) forecasting on the time series data collected from various sensors from a Slitting Machine, to predict the possible failures and quality defects, thus improving the overall manufacturing process.
[15] reviewed Machine Learning as a complementary approach to maintenance planning by analyzing significant data sets of individual machine performance and environment variables, identifying failure signatures and profiles, and providing an actionable prediction of failure for individual parts.
[16] selected simple vibration data collected from an exhaust fan, and have fit different unsupervised learning algorithms such as PCA T statistic, Hierarchical clustering, K-Means, Fuzzy C-Means clustering, and model-based clustering to test its accuracy, performance, and robustness. In the end, they proposed a methodology to benchmark different algorithms and choosing the final model.

[17] presented an overview of condition-based predictive maintenance solutions that aim to avoid unexpected and unplanned failures during the manufacturing and operational process based on advanced data analytics.
[10] predicted when equipment failure could occur (based on certain factors), followed by preventing the failure through regularly scheduled and corrective maintenance.
[18] presented a machine learning-based approach for detecting drifting behavior – so-called concept drifts – in continuous data streams. This maintenance strategy identified wear and tear, and consequent malfunctioning by analyzing condition monitoring data, recorded by sensor-equipped machinery, in real-time
[19] showed how open-source software enables engineers to develop predictive maintenance applications with basic programming knowledge. These types of applications can be widely used in the industrial field to inform about possible equipment malfunction helping reduce possible damages.
[20] presented a Conventional Neural Network (CNN) framework to tackle assets predictive maintenance problems and a method to transform 1-dimensional (1-D) data into an image-like representation (2-D). They showed that data transformation steps were important to make the use of CNN feasible. They obtained two datasets from fans, with a distinct electrical pattern, from a building at Western University.
[21] developed a model to perform a predictive maintenance algorithm that can advise and alert for future system failures by monitoring and comparing with its current and historical parameters.
[22] aimed to apply a predictive maintenance strategy with advanced technologies to overcome limitations of the unforeseen future condition of MEP components and repair in advance to extend the lifetime of facilities.
[23] used Neural Network for the prediction of brake pad failure. Neural Network make accurate prediction if the dataset is very large and also consume a lot of computational power. However, due to the fact that the problem is a classification problem, it is a necessity to carry out performance check of the best supervised model.

V. RESEARCH METHODOLOGY

The Operations analysis approach is used in this work, which is facilitated by predicting supervised machine learning algorithm for the brake pad dataset. In the operations study, problems are decomposed into fundamental components and then solved using logical and mathematical reasoning in different stages. In this research, we are dealing with a classification problem. There are different supervised machine learning techniques that can be used to solve classification problem. The idea of this research is to use Gaussian Naïve bayes, Decision tree and K-Nearest Neighbour for our brake pad dataset. Check for the algorithm that meets the best accuracy for prediction of brake pad failure.

VI. RESEARCH STRATEGY

The study surveyed the brake pad dataset to identify the attributes and factors that are essential in predicting brake pad failure. The research conducted experiments with different machine learning algorithms which are Gaussian Naïve bayes, Decision tree and K-Nearest Neighbour for knowing the best accurate model for our dataset of the research outcome.

VII. DATASET OVERVIEW

The dataset consists of the brake pad history of 15 records. The details capture in the data include km, heat, vibration, and the worn. The variables are essential in helping check for model accuracy for the research dataset.

VIII. DATASET COLLECTION

The research employed a survey in collecting secondary data. Several case studies and research papers on asset predictive maintenance were studied. The dataset that was used for this research work was collected from GitHub. The theoretically studied concepts were practically applied to transform the business logic technologically. These involved the implementation of a machine learning algorithm to predict brake pad failure.

IX. METHODOLOGICAL ASSUMPTION

The methodological assumption for formulating the machine learning algorithm focuses on finding an accurate prediction for the brake pad failure. The models employed include the Gaussian Naïve bayes, Decision tree and K-Nearest Neighbour algorithm. The dataset for this research id divided into two, which are the training data and the testing data. The machine learning model will be built using Python and then use to predict the training data. We will check for accuracy and use the model to predict the testing data.

X. THEORETICAL FRAMEWORK

The fundamental concern of the technical framework consists of data processing and data selection. The research employs the Gaussian Naïve bayes, K-Nearest Neighbour and Decision tree for classification. Given attributes which are speed measure KM, heat measure H, vibration measure V which represents X, the target variable (or output), Y, can take only discrete values for a given set of features (or inputs) X. The ideal goal is to predict brake failure for the output Y, which is a classification problem. According to [23], he made prediction of brake pad failure using Artificial Neural Network, in this research, we will enhance on his research by building a supervised machine learning model for prediction of brake pad failure based on the known dataset downloaded from GitHub.

XI. GAUSSIAN NAÏVE BAYES

Gaussian Naïve Bayes, is form of naïve bayes that deals with continuous data, it is derived from the bayes theorem. It is used in classification problems, and it assumes that features follow a normal distribution. Given a feature vector $X=(x1,x2,...,xn)$ and a class variable Ck, Bayes Theorem states that:

$P(Ck|X)=P(X|Ck)P(Ck)/(P(X))$, for k=1,2,...,K

We call $P(Ck/X)$ the posterior probability, $P(X/Ck)$ the likelihood, $P(Ck)$ the prior probability of a class, and $P(X)$ the prior probability of the predictor. We're interested in calculating the posterior probability from the likelihood and prior probabilities (Nikita, 2020).In terms of machine learning, we mean to say that the features provided to us are independent and do not affect each other, and this does not happen in real life. The features depend on the occurrence or value of another, which is simply ignored by the Naive Bayes

classifier and is hence given the term, **"NAIVE"** (Nikita, 2020)**.**Thus, by conditional independence, we have:

$$P(y|x_1,...,x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)} \quad (1)$$

XII. DECISION TREE

A decision tree is an efficient algorithm for describing a way to traverse a dataset while also defining a tree-like path to the expected outcomes. This branching in a tree is based on control statements or values, and the data points lie on either side of the splitting node, depending on the value of a specific feature [24].The structure of a decision tree can be defined by a root node, which is the most important splitting feature. The internal nodes are tests on an attribute.

Selection measures

Information Gain: This attribute provides us the splitting attribute in terms of the amount of information required to further describe the tree. Information gain minimizes the information needed to classify the data points into respective partitions and reflects the least randomness or "impurity" in these partitions (Nikita, 2020).

$$Info(D) = -\sum_{i}^{m} p_i \log_2(p_i), \quad (2)$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j). \quad (3)$$

pi is the probability that an arbitrary tuple in dataset D belongs to class Ci and is estimated by|Ci,D|/|D|.

Info (D) is simply the mean amount of information needed to identify the class/category of a data point in D. A log function to base 2 is used, since in most cases, information is encoded in bits.

Info (D) is also known as the entropy of the dataset D.
The information gain can be calculated as follows;

$$Gain(A) = Info(D) - Info_A(D). \quad (4)$$

Gain Ratio: The information gain measure is biased toward tests with many outcomes. Thus, it prefers selecting attributes that have a large number of values. Gain ratio is an attempt to improve this problem.

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right). \quad (5)$$

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}. \quad (6)$$

Gini Index: The Gini index is calculated in the following manner:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2, \quad (7)$$

Where pi is the probability that a tuple in D belongs to class Ci and is estimated by |Ci,D|/|D|. The sum is computed over m classes.

XIII.   K-NEAREST NEIGHBOR

K-Nearest Neighbor (KNN) is a supervised classification algorithm that is based on predicting data by finding the similarities to the underlying data. KNN is most widely used for classification problems, but can also be used to solve regression problems.   The concept of finding nearest neighbors may be defined as "the process of finding the closest point to the input point from the given data set" [25]. The algorithm stores all the available test data and classifies new test data by majority votes of its K neighbors. When implementing KNN, the first step is to transform data points into their mathematical values (vectors). The algorithm works by finding the distance between the mathematical values of these points. It computes the distance between each data point and the test data and then finds the probability of the points being similar to the test data. Classification is based on which points share the highest probabilities. The distance function can be; Euclidean, Minkowski or the Hamming distance.

Euclidean Distance Function

Euclidean distance can simply be defined as the shortest between the 2 points irrespective of the dimensions. The most common way to find the distance between is the Euclidean distance. According to the Euclidean distance formula, the distance between two points in the plane with coordinates (x, y) and (a, b) is given by;

dist $((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}$

Minkowski Distance Function

Minkowski distance is a distance measurement between two points in the normed vector space (N-dimensional real space) and is a generalization of the Euclidean distance.

Consider two points P1 and P2:

P1: (X1, X2,……Xn)

P2: (X1, X2,……Xn)

Then, the Minkowski distance between P1 and P2 is given as: When p=2, Minkowski distance is same as Euclidean distances

Hamming Distance

Hamming distance is a metric for comparing two binary data strings. While comparing two binary strings of equal length, Hamming distance is the number of bit positions in which the two bits are different.

The Hamming distance between two strings, a and b is denoted as d (a, b).

In order to calculate the Hamming distance between two strings, and, we perform their XOR operation, $(a \oplus b)$, and then count the total number of 1s in the resultant string.

XIV.   CONCEPTUAL DESIGN

During the system analysis, the analysis of system data is very important [26]. Analysis of data is made up of more than one level at the beginning (first level) and different ideas are used at each level. At the first level, we develop a conceptual system design.
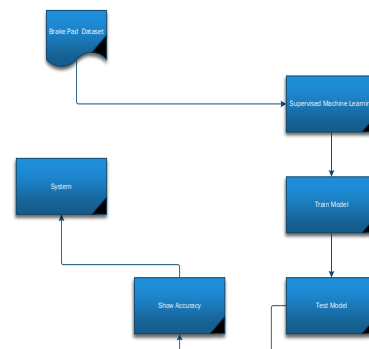

Figure 1. Conceptual Design of the system

XV.   RESULT

In this phase, we will be more detailed. In this research, we are working on a classification problem. Here we will use the K-Nearest Neighbor, Decision tree and Gaussian Naïve bayes to test for accuracy for our dataset. We use python 3 programming language to test for accuracy which comes with many helpful analytics libraries installed.

A. Model design phase

In this phase, the researcher will build the K-Nearest Neighbor, Decision tree, and Gaussian Naïve bayes model. The dataset is made up of attributes which is a property of an instance that may be used to determine its classification. In our dataset, the attributes are the km, heat, and vibration which is also known as features. Our output variable for our brake pad dataset is in binary which worn should be either 1 or 0. In the model design, the research aim to achieve the following steps to successfully build the research model.

- ➢ Loading our dataset.
- ➢ Show how data are distributed
- ➢ Divide the brake pad dataset into training and testing data. We will make our training dataset larger then than the testing dataset.
- ➢ Use the K-Nearest Neighbor, Decision tree, and Gaussian Naïve bayes for our classification problem.
- ➢ Passing the testing data to the trained algorithm to predict the outcome.
- ➢ Checking the accuracy of each of the selected algorithm and recommending the best algorithm for deployment

1. Loading the dataset

```
In [26]: dataset = pandas.read_csv("brakepad.csv")

In [27]: print(dataset.head(15))

       worn       km  heat        z
    0     1  20000.0   240    2.720
    1     0   5000.0    98  -57.706
    2     1  50000.0   140  122.420
    3     0   8000.0   260  -45.220
    4     1  23790.0   225   17.835
    5     1  24644.0   245   21.311
    6     1  29934.0   195   42.321
    7     0  14045.0   153  -21.361
    8     0   8000.0   222  -45.334
    9     0   9855.0   149  -38.133
   10     1  24633.0   271   21.345
   11     1  20753.0   209    5.639

In [28]: print(dataset.shape)

    (12, 4)
```

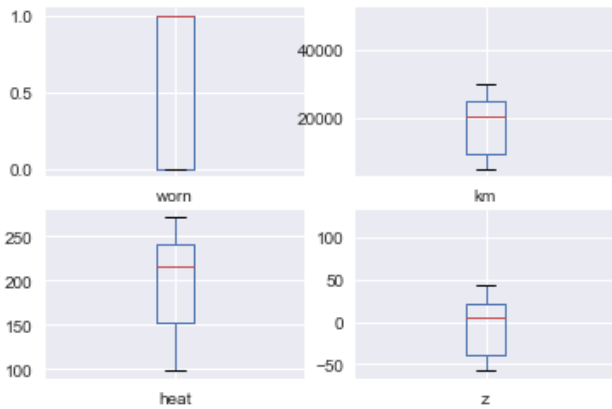Figure 2: Brake Dataset

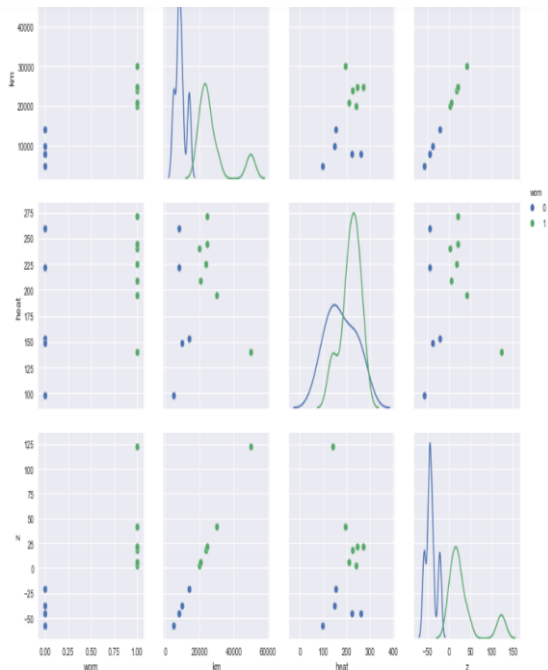*B.   Data Visualization*

Figure 2: Data Visualization
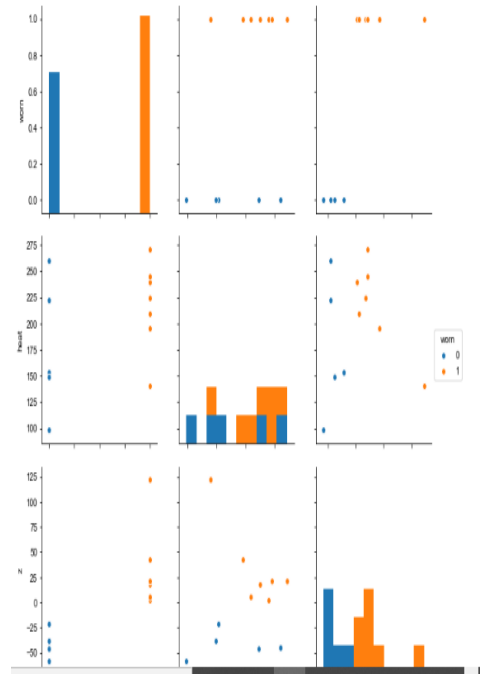

Figure 4: Data Visualization
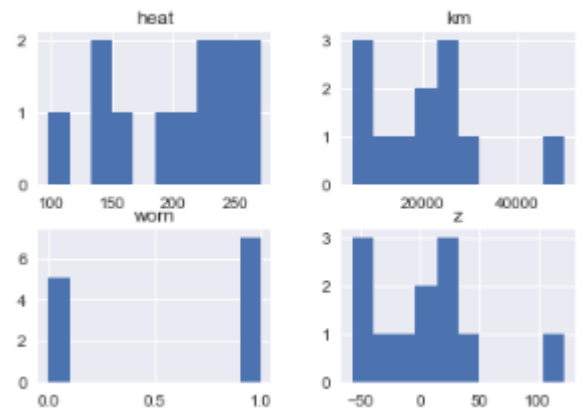

Figure 3: Data Visualization

Figure 5: Data Visualization

C.   Accuracy of the testing model

We can see that the Decision tree classifier performed very well with a high-performance accuracy of 95% for our small dataset for this research work.

```
In [75]: predResult=brake[['km','heat','vibration','labelbrake']]
```

```
In [76]: train_p,test_p=train_test_split(predResult,test_size=0.3,random_state=0)
         train_x_p=train_p[['km','heat','vibration']]
         train_y_p=train_p.labelbrake
         test_x_p=test_p[['km','heat','vibration']]
         test_y_p=test_p.labelbrake
```

```
In [28]:
```

```
In [77]: model=DecisionTreeClassifier()
         model.fit(train_x_p,train_y_p)
         prediction=model.predict(test_x_p)
         print('The accuracy of the Decision Tree is:',metrics.accuracy_score(prediction,test_y_p))

         The accuracy of the Decision Tree is: 0.9555555555555556
```

```
In [78]: model=KNeighborsClassifier(n_neighbors=3)
         model.fit(train_x_p,train_y_p)
         prediction=model.predict(test_x_p)
         print('The accuracy of the KNN Tree is:',metrics.accuracy_score(prediction,test_y_p))

         The accuracy of the KNN Tree is: 0.9111111111111111
```

```
In [79]: model=GaussianNB()
         model.fit(train_x_p,train_y_p)
         prediction=model.predict(test_x_p)
         print('The accuracy of the Gaussian Naive Bayes is:',metrics.accuracy_score(prediction,test_y_p))

         The accuracy of the Decision Tree is: 0.8888888888888888
```

XVI. CONCLUSION

The dataset that was used for this research work was downloaded from GitHub. The dataset was divided into training dataset and the testing dataset. Python 3 was used to build the machine learning models for prediction. Gaussian Naïve bayes, Decision tree and K-Nearest Neighbor model was built. The training dataset was tested for accuracy based on the built models. From our result, Gaussian Naïve bayes model has an accuracy of 91.11%, Decision tree model has an accuracy of 91.11%, and K-Nearest Model has an accuracy of 93.33%. We then tested the training model on our testing data which Gaussian Naïve bayes model has an accuracy of 88.88%, Decision tree has an accuracy of 95.55% and the K-Nearest Neighbor has an accuracy of 91.11%. From the result, we showed that the Decision tree algorithm model is the best based on the accuracy we get for our research dataset. In the future, we intend making our predicting model more user friendly. We intend migrating it to our SQL Server 2019 database where we can import our python model to the SQL Transact code and store the code in a SQL procedure so that we can build a scalable application by calling the deployed model anywhere on any kind of application.

REFERENCES

[1] Oshios , E. I., Angela , M. O., Moses , E., and Albert , T. W. (2020). Development of asset management model using real-time equipment monitoring (RTEM): case study of an industrial company. *cogent bussiness and management*, 1-10.

[2] Barbara , N., Daniel , M., Sherry , M., & Rachel , B. (2014). *Artificial intelligence and machine learning in asset maintenace.* blackrock.

[3] Gavaghan. (2021). *The Impact of Artificial Intelligence on Jobs and Work.* Retrieved from otago: https://www.otago.ac.nz/caipp/otago828396.pdf

[4] Rebecca, H. (2019). *The Future of Asset Management: AI in Asset Management*. Retrieved from liquidnet: https://www.liquidnet.com/expert-insights/ai-in-asset-management

[5] Investopedia. (2019). *Fixed Asset vs. Current Asset: What's the Difference?* Retrieved from investopedia: https://www.investopedia.com/ask/answers/040915/what-difference-between-fixed-assets-and-current-assets.asp

[6] Institute (2020). *Types of Assets*. Retrieved january 15, 2021, from corporatefinanceinstitute: https://corporatefinanceinstitute.com/resources/knowledge/accounting/types-of-assets/

[7] Team, I. (2019). *Fixed Asset vs. Current Asset: What's the Difference?* Retrieved January 2021, from investopedia.com: https://www.investopedia.com/ask/answers/040915/what-difference-between-fixed-assets-and-current-assets.asp

[8] Holstein, W. K. (2021). *Production management*. Retrieved from britannica: https://www.britannica.com/technology/production-managemen

[9] Gackowiec, P. (2019, september). *General overview of maintenance strategies – concepts and approaches*. Retrieved from researchgate: https://roadtoreliability.com/types-of-maintenance

[10] Noria, A. (2019). *Predictive Maintenance Explained*. Retrieved from reliableplant: https://www.reliableplant.com/Read/12495/preventive-predictive-maintenance

[11] Mohammed , F. B., Eric, Z., and Frédéric, D. (2013). Towards Bayesian network methodology for predicting the equipment health factor of complex semiconductor systems. *International Journal of Production Research*, 4597-4617 .

[12] Kroll, D., and Nigggemann, O. (2014). System modeling based on machine learning for anomaly detection and predictive maintenance in industrial plants. *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA* (pp. 1-7). Barcelona, Spain: IEEE.

[13] Jiri , K., Kamil , K., Pavel, B., Ondrej, K., and Daniel, J. (2016). Application of Artificial Neural Networks in Condition Based Predictive Maintenance. *springer*, 56-76.

[14] Kanawaday, A., and Sane, A. (2017). Machine learning for predictive maintenance of industrial machines using IoT sensor data. *8th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 87-90). Beijing, China: IEEE.

[15] Cline, B., Niculescu, B., and Decke. (2017). Predictive maintenance applications for machine learning. *2017 Annual Reliability and Maintainability Symposium (RAMS)* (pp. 1-7). Orlando, FL, USA: IEEE.

[16] Amruthnath ,N., and Gupta,T. (2018). A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 355-361). singapore: IEEE.

[17] Nazmus , S., and Thorsten, W. (2018). Challenges and Opportunities of Condition-based Predictive maintanace : a review. *elsevier*.

[18] Jan, Z., Florian, H., and Michael, A. (2019). Machine learning based concept drift detection for predictive maintenance. *Chinese Journal of Aeronautics*.

[19] Liulys, K. (2019). Machine Learning Application in Predictive Maintenance. *2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)* (pp. 1-4). Vilnius, Lithuania: IEEE.

[20] Willamos, S., & Miriam , C. (2019). Assets Predictive Maintenance Using Convolutional Neural Networks. *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. Toyama, Japan: IEEE.

[21] Vibha, K. , Hridi , P., Manasvi , V. B., and Nikita , P. (2020). Development of Predictive Maintenance System of Motors. *International Journal of Advanced Science and Technology*, 2344-2351.

[22] Jack, C. P., Cheng, Weiwei , C., Keyu , C., and Qian , W. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Chinese Journal of Aeronautics*.

[23] Walker, R. (2021). *Brake Failure Prediction*. Retrieved 2021, from Spark ML and Hydrosphere Mist Example: Preventive Maintenance: https://github.com/werowe/mist_preventive_maintenance_ml#5

[24] Nikita, S. (2020). *Understanding the Mathematics Behind decision tree*. Retrieved July 2021, from HeartBeat: https://heartbeat.fritz.ai/understanding-the-mathematics-behind-decision-trees-22d86d55906

[25] Pandey, A. (2020). *The Math Behind KNN*. Retrieved from ai.plainenglish: https://io/the-math-behind-knn-7883aa8e314c

[26] Dhiraj, M. (2021). *Conceptual System Design*. Retrieved January 2021, from Technology: http://latesttechnomanias.blogspot.com/2011/08/conceptual-system-design.html