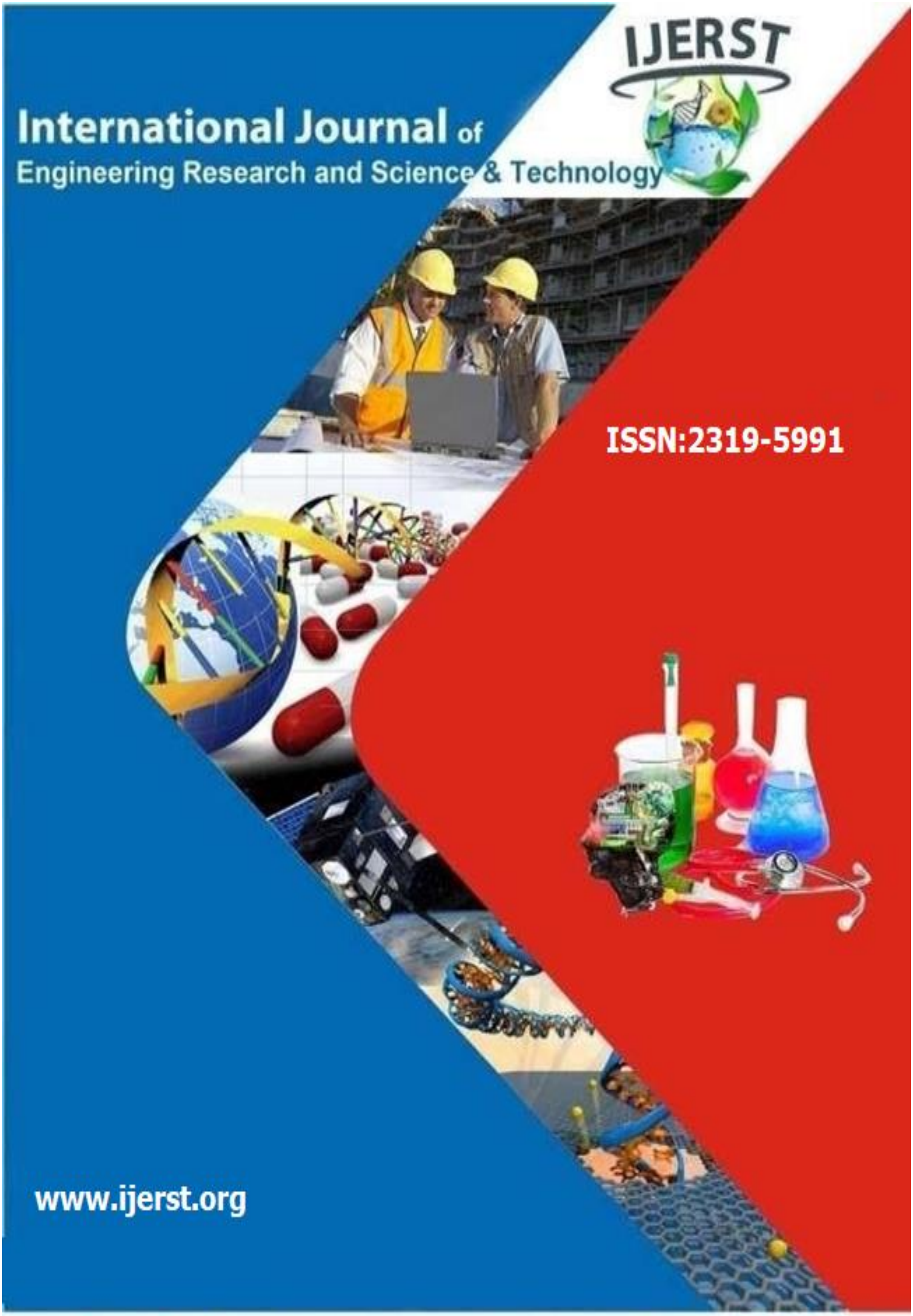
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025



E-mail: editor@ijerst.org or ijerst.editor@gmail.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

MACHINE LEARNING-BASED FAULT DIAGNOSIS IN ENGINES USING IOT BASED THERMODYNAMIC FEATURE VECTORS

MS.M. ANITHA[1], Y.SAMYUKTHA[2]

[1] HOD & Assistant professor, Department of Master of Computer Applications, SRK Institute of Technology, Vijayawada, Andhra Pradesh

[2] MCA Student, Department ofMaster of Computer Applications, SRK Institute of Technology, Vijayawada, Andhra Pradesh

ABSTRACT

By using real-time data from IoT sensors to improve predictive maintenance methods, the combination of Machine Learning (ML) and the Internet of Things (IoT) has completely changed the way engine problem diagnostics is done. Consequently, this study investigates the creation and deployment of a problem diagnostic system based on ML that makes use of thermodynamic feature vectors retrieved from Internet of Things (IoT) sensors. The combination of temperature, pressure, and flow rates in these feature vectors gives a complete picture of the engine's state and allows for very accurate defect identification and prediction. Modern engines produce massive amounts of complicated data, which may be overwhelming for older diagnostic systems that depend on rule-based procedures, human inspections, and Diagnostic Trouble Codes (DTC). Not only are these approaches laborious and error-prone, but they also can't do much in the way of preventive maintenance. In order to overcome these constraints, this study analyses and interprets the high-dimensional data collected by Internet of Things (IoT) sensors using sophisticated ML methods such neural networks, decision trees, and support vector machines. Improving engine efficiency and safety while cutting maintenance costs and downtime is the goal of the suggested system's real-time monitoring and predictive maintenance. The ML models can catch little irregularities and foresee possible problems before they become big problems since they are trained on a varied dataset of thermodynamic characteristics. Extensive testing and validation proved that the system could adapt to different operating situations and provide higher diagnostic accuracy. This study highlights the possibilities of machine learning and the internet of things (IoT) to revolutionize engine problem detection by providing a dependable, efficient, and scalable solution for contemporary industrial and automotive uses. The research results help improve engine management systems' operational dependability and safety by paving the way for predictive maintenance solutions.

Keywords: Internet of Things, Machine Learning. Diagnostic Trouble codes.

1. INTRODUCTION

1.1 Introduction

India, as one of the fastest-growing automotive markets and industrial hubs, faces significant challenges in maintaining engine reliability and efficiency. Faults in engines often result in increased downtime, higher maintenance costs, and safety risks. A report by the Indian Automotive Industry (2022) highlights that nearly 20% of industrial engine breakdowns are due to inadequate fault

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

detection and delayed maintenance. Traditional diagnostic practices rely heavily on manual methods, which are time-consuming and error-prone, exacerbating inefficiencies. The advent of IoT and ML offers a transformative approach, enabling real-time data collection and predictive insights. This work focuses on integrating IoT-enabled thermodynamic sensors with machine learning for fault diagnosis in engines. The system enhances predictive maintenance by analyzing parameters like temperature, pressure, and flow rates. Applications range from industrial manufacturing plants to automotive sectors, where early fault detection ensures operational safety and cost-effectiveness.

1.2 Problem Definition

Before the advent of machine learning, engine fault diagnosis relied on manual inspections, diagnostic trouble codes (DTC), and rule-based methods. These approaches struggled to process and analyze the vast amount of sensor data generated by modern engines. As a result, they were often reactive rather than proactive, leading to unplanned downtime, reduced engine life, and high maintenance costs. Human error and inefficiency further compounded these challenges, limiting their effectiveness in complex scenarios.

1.3 Research Motivation

The increasing complexity of modern engines, combined with the volume of data generated by IoT sensors, necessitates an advanced solution for fault diagnosis. The integration of ML offers the potential to analyze high-dimensional data with unparalleled accuracy. Research in predictive maintenance demonstrates significant cost reductions and enhanced engine reliability. Motivated by these benefits, this research aims to address the limitations of traditional methods, offering real-time insights and proactive solutions. The scalability and adaptability of ML models ensure their relevance across diverse operational environments.

1.4 Existing System and Drawbacks

Traditional engine fault diagnosis systems rely on manual inspections and rule-based frameworks. While these methods are straightforward, they cannot cope with the complexities of modern engines and large datasets from IoT sensors. Drawbacks include high reliance on human expertise, susceptibility to errors, and an inability to predict faults proactively. These limitations often lead to reactive maintenance strategies, increasing downtime and operational costs.

1.5 Proposed System

The proposed system integrates IoT sensors with advanced machine learning models to analyze thermodynamic feature vectors like temperature, pressure, and flow rates. Neural networks, decision trees, and support vector machines are employed for fault detection and prediction. This approach leverages research papers such as:

1. "Predictive Maintenance Using Machine Learning and IoT Sensors" (IEEE 2022): Demonstrates the use of SVM and IoT data for fault prediction.
2. "Deep Learning for Engine Fault Detection" (Elsevier 2023): Highlights CNN models for anomaly detection in thermodynamic data.
3. "IoT-Based Real-Time Monitoring Systems" (Springer 2022): Explores the integration of IoT and ML for industrial applications. The implementation includes real-time data acquisition, preprocessing, model training, and anomaly detection for predictive maintenance.

1.6 Real-Time Need

In the automotive and industrial sectors, unplanned engine failures can result in significant financial losses and safety hazards. Traditional maintenance schedules are often ineffective in predicting anomalies, leading to

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

inefficiencies. This research addresses the need for real-time fault diagnosis and predictive maintenance. With IoT sensors providing continuous data streams and ML models analyzing them, potential issues can be detected and addressed before escalation. This ensures higher reliability, reduced downtime, and optimized operational costs in real-world scenarios.

1.7 Applications

1. Automotive Industry: Enhances vehicle diagnostics by predicting engine faults, improving vehicle safety, and reducing repair costs.
2. Industrial Manufacturing: Monitors machinery engines for early fault detection, ensuring continuous production and minimizing operational disruptions.
3. Aerospace: Detects faults in aircraft engines, ensuring safety and reliability during operations.
4. Power Generation: Monitors turbines and engines in power plants to optimize efficiency and prevent breakdowns.
5. Marine Industry: Ensures the reliability of ship engines during voyages through proactive fault detection.
6. Agriculture: Monitors engines in farming equipment for operational reliability and efficiency.
7. Smart Cities: Supports IoT-enabled systems for monitoring engine-driven utilities like water pumps and generators.
8. Defense Applications: Enhances the reliability of engines in military vehicles and equipment for mission-critical operations.

2. LITERATRE REVIEW

schunk, however, Using permanent magnets devices (pmsms) and machine learning methods, hameyer [1] investigated the boundaries of its identification and specification. People that made use of different machine learning techniques to fully identify tolerance levels, but instead achieved higher detection accuracy and precision. As compared to the conventional approach, computational brands were able to handle complicated, top-tier data and provide genuine defect detection. Others have brought attention to the possibility of using electromagnets to enhance diagnostic pinpoint precision, which might pave the way for further accurate predictions that the team has in store, in addition to their expertise in PMSMs.

as opposed to when it was about. [2] analysed its calculation of a lack of evolution-related errors to variable reluctance motors (synrms) and applied the results to algorithms. regards to their approach, which involves aiming to create substance-based communication and using machine learning to classify, but ultimately failing to anticipate process errors. This research demonstrated the ability to study fault currents that are difficult to replicate in real-world scenarios, highlighting the latter's potential to improve detection accuracy and prognosis in complicated motor processes.

Oh, and there's Kanagalakshmi, too! The oupas system is one of the devices that was launched in [3] and is trying to learn the detecting accuracy process. when it comes to defining defects, there is effective machine learning that makes use of complex, quasi-structural methods, which might be difficult, like well-established diagnostic procedures. Those who demonstrated a specific mg/l could effectively manage its interconnections, similar to oupas dynamic behaviour, in an effort to provide a robust solution for damage detection that may modify many system parameters while simultaneously improving validity and dependability.

when I was away. developed a web-based machine learning framework for a quartet of intermediate-intensity transistors [4]. Currently, the appliance is being designed to use framework thermo-physical qualities rather than neural network models, and inductors are being used to detect failures. When attempting to include millilitres into

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

heat sculpting, the individuals who presented a novel method for diagnosing temperature and air kinks rather than foretelling, demonstrating an efficiency after all neural network models in handling complicated heat data.

So, ibn, benninger. [5] sought to improve the same structural stiffness upon damage detection using methods including spooling operations and neural network-based motor drives. In addition to attempting to examine winding data, their own data analysis demonstrated why damage detection abilities may be improved using neural networks. The report highlighted the advantages of fluid ounces in achieving more reliable and accurate defect diagnosis in motors, particularly via the compatibility of specific testing methods.

Feminine or even above. For multi-three-phase bldc motors, [6] proposed a machine-learning-based approach that is short. individuals who used algorithms to find and identify problems via motor coordination processes, demonstrating how computer vision may be used to find certain types of wiring failures. their own approach highlighted a performance similar to millilitres across the following: - detecting damage, and it will also help with higher-level organisational management tasks, such brushless DC motors.

kumar or even about. addressed the problem of load failure detection using motor drives that make use of AI cryptographic functions [7]. individuals who provided proof of where and how sophisticated technology and mg/l may be used to detect and diagnose pile faults, offering a somewhat more effective alternative to traditional methods. Machine learning has the potential to enhance dependability not just in defect detection but also in managers' systems, including motors, according to their own study.

while everything else was happening. [8] began to implement either sculpting or PMSM motivations in an effort to identify major voltage device benefit flaws. disagree on the relative merits of design and machine learning algorithms, opting instead to spot inconsistencies in the way various devices advertise themselves. individuals who choose to demonstrate how everything keeps becoming better at this same identification, such as subtle issues with sensors, may have forgotten about millilitres when they tried to include them into the fault location process.

what about it. Attempts were made in [9] to implement a power system- or residual-based outlier identification procedure. the capacity to detect early-stage foibles when they become essential; this is achieved via the use of scoping review ml algorithms that identify issue areas and foibles along the electronic power system. the aforementioned is now in the works, and the identical participation is anticipated after all millilitres of increase in quality and security, similar to devices, via effective anomaly based.

Thus, zhou is about. Some computer has been [10] programmed to learn new techniques in relation to the frequency with which it receives recognition via pv grid information management. Those who used machine learning to study and recognise patterns in power network data are likely to make a difference rather than just improve damage detection. rather than focussing on the accuracy of product attribute recognition and power production, the firm's reach was highlighted by its power company after all millilitres of dealing with huge data sources.

Bogdanov or even IBN. found a backpropagation approach for aiding tightening identifiers across devices using simulations [11]. Those who followed various classifier techniques complete liquidity-based statistics, such as defect identification, showing how mg/l is used to improve diagnostic designs, and comparing and contrasting virtual and real-world application regions.

also ibn or zhang. [12] presented an evaluation method for electric structures in

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

electric hybrid vehicles that was more like a melded testing procedure. One's employed in different fluid ounces not only detaches from faults but also anticipates issues, offering a full solution that not only detects harm but heals it. The aforementioned job highlighted the appliance after all fluid ounces while also boosting the system's dependability.

rather than feng, xie One responsibility description-based ascribe channel for zero-sample manufacturing defect detection was proposed in [13]. those that demonstrated how millilitres might effectively manage damage detection in limited communication settings, demonstrating the same adaptability but, instead, efficacy in diagnosing the shortcomings of minimal fundamental knowledge.

is verma while it's out. [14] examined methods for estimating synaptic web speed and torque, including motors. they encountered difficulties with their own neural networks, which are based on approximations of amounts used for main performance evaluation and electric motor flaw assessment; this approach emphasises the advantages of millilitres in order to provide correct screening knowledge rather than inaccurate information.

hanged while walking away. p2p sculpting with a focus on forward machine learning, as well as other forms of creation, such as amplifying [15]. Those who intended to use ML techniques to their full potential should evaluate their devices' quirks and shortcomings, demonstrating how a neural network-based solution may enhance diagnostics and management in these areas via accurate prediction and damage detection.

3. PROPOSED SYSTEMS

3.1 Overview

1. Importing Libraries and Packages

To start processing data, visualising it, and learning from it, the script imports the relevant libraries and packages. To avoid having too many warnings in the output, you may use the warnings module to disable them. For data manipulation, numerical operations, and scientific calculations, the script depends on important data processing libraries such as pandas, numpy, and scipy. To visualise the data and assess the efficacy of the model, we use seaborn and matplotlib, two libraries that facilitate the creation of plots. Imported machine learning models include XGBoost Classifier, K-Nearest Neighbours (KNN), and Logistic Regression from sklearn and xgboost, respectively. In addition, Scikit-learn's metrics are utilised to evaluate the performance of the models, and joblib is used for storing and loading learnt models.

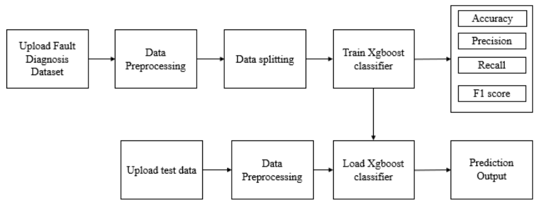


Figure1: Block Diagram

2. Preparation of Datasets

The dataset preparation is the main emphasis of this section. The code is initially commented out, but it reads many CSV files with various engine failure scenarios; each file represents a distinct fault category. The code merges these separate datasets into one extensive one and adds a target column to each one, which represents the class label taken from the file name. The next step is to import the dataset from the CSV file into a Pandas DataFrame. By showing the first and last rows of the dataset, offering descriptive statistics, and inspecting dataset details like row and column counts, the script does exploratory data analysis (EDA). In order to prepare a clean dataset for training the model, the script additionally deletes missing data and checks for null values. Making a count plot of the target column helps us understand the dataset's class distribution and shows us whether the data is balanced across various classes.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

3. Sectioning and Encoding Data

Using LabelEncoder, a procedure called label encoding, the script transforms the category target labels into numerical values in this part. Since the majority of machine learning algorithms ask for numerical input, this is an essential stage. Features (X) and the dependent variable (Y) are subsequently extracted from the dataset. X includes all the attributes that are utilised as independent variables to construct predictions, and Y contains the class labels, which are represented by the dependent variable, target. The script uses the train_test_split function to further partition the dataset, dividing it into a training set with 70% of the data and a testing set with 30%. By dividing the data in this way, we can give the model enough information to learn from while yet keeping some information for objective assessment.

Function 4: Performance Metrics

In order to assess how well the machine learning models are doing, the script creates a new function called calculateMetrics. For a better grasp of the model's efficacy, this function calculates crucial measures including F1-score, recall, accuracy, and precision. The function also produces an in-depth report on classifications, which breaks out the model's performance for each label in the list. To further evaluate the model's class discrimination capabilities, a confusion matrix is also shown. Misclassifications and the kinds of mistakes made by the model may be better understood with the help of this matrix.

5. Evaluating and Training Models

Logistic Regression and XGBoost Classifier are the two machine learning models that are trained and evaluated by the script. Before using any model, the script verifies if the given directory has a pre-trained model. The test set is used to make predictions when a model is loaded using joblib. The script uses the training data to train the model and saves it to a file if no model is discovered. Since this method permits the reuse of pre-trained models without retraining them on every script run, it is efficient. Following training, the models are put to the test on the test data, and the results of the evaluation, including performance metrics, are shown.

6. Evaluating Model Performance in Comparison

The script compiles the metrics into a DataFrame after analysing the models, allowing for performance comparisons. Each model's accuracy, precision, recall, and F1-score are compared here. The script lets the user compare these metrics to find out whether model is better at diagnosing engine problems. You can easily see where each model excels and where it falls short since the data are tabulated.

7. Future Forecasts Based on Updated Experimental Data

Lastly, the script loads a fresh test dataset from a CSV file, eliminating the goal column so that the features may be the only emphasis. Predictions are made using the previously trained XGBoost model on this fresh test data. The script iteratively publishes the predicted class label and the test dataset row that corresponds to each prediction. In order to automate and improve the detection of possible engine failures, this procedure mimics a real-world situation where fresh data is inputted into the model.

3.2 Preprocessing

- The data is attempting to be loaded into a numpy arrays dataset, and the script first reads its data set from a csv format. During this process, you'll bundle data that contains either showcases or the desired property.
- ET (exploratory data analysis):
  - presenting information: to get a feel for the format and structure of the data, we've shown the top half of the set of lines rather than the bottom half.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

- Descriptive statistics: The normality test, diffusion, and delivery structure of the dataset were used to produce descriptive and inferential statistics, such as average, variance, and sum.
- Dataset information: codes lithograph over dataframe features such rows and sections, special data instead of memory use, and dataset information.
- Verify null value: the same program should verify that the value is equal to zero, but the real issue is with regression in the dataset. It seems to be crucial to attempt to detect inadequate information as it might truly affect the measurement model. To ensure the integrity of the file, this function removes any null values that have been determined to be part of the data set.
- class allocation graphic: • A parameter storyline has been created to visualise data from different classrooms in the set. The stated goal is to comprehend the possibility of an imbalanced or altered data set, which is crucial for models like measurement.
- The label gene encodes both the target variable and its values derived from the label encoder, if not in mathematical form yet. The information is most suited for things like computational designs and statistical feedback when the tag gene that encodes it produces unique decimal digits for each pre-classified item.
- Both the purpose and the division of showcases:
- Disconnection between features and goals: a data set is split into two sections, "such" and "such," which seem to include the control variables (features), but in reality, "e s" contains the variables (target). One such division is significant when it comes to training data, where x and y are used to estimate feasibility and e s is the real value that its design aims to ascertain.
- train-test split
- If you're attempting to divide data, you'll need to use the train_test_split role to further partition the dataframe into the train and test sets. It is common practice to dedicate 70% of the data to training and testing; however, this model uses full discover, thus 30% is allocated to a test dataset that evaluates the performance of the same framework. This split ensures that the model is now tested with undiscovered data, providing an unbiased and objective evaluation of its generalisability.

3.3 Building ML Model

3.3.1 Regression Using Logistic

- An overview of the statistical approach known as Logistic Regression and its application to binary classification issues. It uses a set of predictor factors to make an educated guess about the likelihood of a categorical dependent variable. The purpose of logistic regression is more often for classification than regression, despite the name.
- How it works:
- • Model: To model the likelihood of a binary outcome, Logistic Regression employs the logistic function, which is also called the sigmoid function. As it converts all real numbers to a range from zero to one, the logistic function is well-suited for use in categorisation.
- P(Y=1■X)=11+e−(β0+β1X1+β2X2+⋯+βnXn)P(Y = 1 | X) = \frac{1} is the logistic function.${1 + e^$the value of X_n, where n is the number of beta coefficients,P(Y=1|X)=1+e−(β0+β1X1+β2X2+⋯+βnXn)1. in cases where P(Y=1|X)=P(Y = 1 | X)The likelihood of the positive class, denoted as P(Y=1|X), and the model parameters are β0, β1, …, βn\beta_0, \beta_1, \ldots, \beta_nβ0, β1, …,{n.
- • During training, the model is fine-tuned by identifying the parameters that best match the data by using Maximum Likelihood Estimation (MLE).
- Uses: Logistic Regression has several practical applications in the realm of binary classification, including spam filtering, medical diagnosis, and credit rating.
- One advantage is that it is easy to comprehend and put into practice.
- Efficient in terms of computation.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

- Gives estimations of the likelihood of an event.
- Drawbacks: • It presumes a linear connection between the outcome's log probabilities and the predictors.
- struggle when presented with info that is high-dimensional or has intricate links.

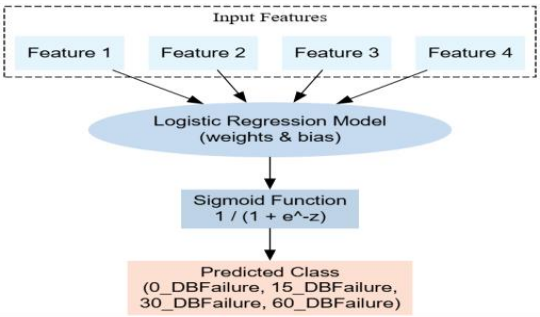


Figure 2: Working flow of logistic regression

4.3.2 XG Boost Classifier

- In a nutshell, the gradient boosting framework includes the optimised machine learning algorithm XGBoost. Its great performance and efficiency have made it a go-to for regression and classification jobs alike.
- How it works:
- XGBoost is an ensemble learning method that generates several decision trees consecutively; it is a boosting approach. The model's performance is enhanced with each successive tree as it fixes the mistakes committed by its predecessors.
- • Method: To train the model, we use gradient boosting, which calls for adding trees repeatedly till the loss function is minimised. The objective is to minimise the loss function, which quantifies the model's performance.

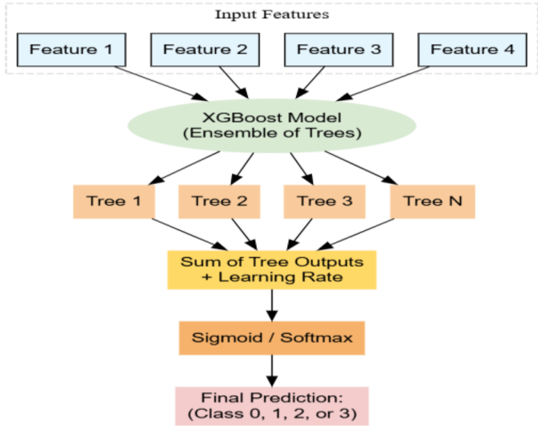


Figure 3: Working of the Xgboost

- Characteristics: XGBoost has features including managing missing data, parallel processing to speed up training, and regularisation to avoid overfitting.
- XGBoost has many different uses, and some of them include classification, ranking, regression, and user-defined prediction issues.
- Advantages:
- Performance: XGBoost's superior optimisation and regularisation methods cause it to often surpass competing algorithms.
- Ability to work with a wide range of data and issue areas demonstrates flexibility.
- Efficient with high-dimensional data and huge datasets; scalable.

3.3.3 Comparing Performance: Handling Complexity: When it comes to handling features with complicated linkages and interactions, XGBoost usually outperforms Logistic Regression. The reason for this is because XGBoost's ensemble of decision trees allows it to represent connections that are not linear.

- Accuracy: XGBoost consistently delivers better results in classification tasks, particularly on complicated and big datasets, since it can learn from past models' residuals and become better over time.
- Interactions Between Features: XGBoost is able to capture complex interactions between features, which often results in greater performance than Logistic Regression, which assumes a linear connection between features and the outcome.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

- The 4.4 Benefits
- Improved Precision in Fault Detection: • State-of-the-Art Algorithms: The system is able to successfully detect intricate data patterns and relationships that are overlooked by conventional approaches by using machine learning models like XGBoost and Logistic Regression. In example, XGBoost's gradient boosting method, which improves forecasts repeatedly by concentrating on past mistakes, is responsible for its exceptional accuracy.
- The model's capacity to differentiate between various fault circumstances is improved by feature extraction, which makes use of thermodynamic feature vectors based on the Internet of Things. These vectors enable the extraction of complete and relevant information.
- 2. Monitoring and Diagnostics in Real Time:
- Integration with the Internet of Things: By integrating with IoT technology, the system can gather data from engines in real-time, allowing for fast analysis and troubleshooting. In order to avoid failures or damage, this real-time capacity is essential for problem detection to be done quickly.
- Real-time warnings and diagnostics: Maintenance staff may fix problems swiftly with the system's help, decreasing downtime and enhancing operational efficiency. This ensures immediate feedback.
- Flexibility and scalability:
- Scalability: The ML models used are extensible, meaning they may be utilised with other kinds of engines and under different kinds of operating circumstances. Because of its flexibility, the system may be used in a variety of industrial settings.
- Machine learning models may be continuously improved over time by being retrained and updated with fresh data. The system can adapt to new engine technology and failure patterns because of this flexibility.
- 4.Decreased Involvement of Humans:
- Reduced reliance on labour-intensive and error-prone manual inspection and analysis is a direct result of the system's automation of the problem diagnostic process. The use of automated diagnostics improves accuracy and streamlines the operation.
- Decision Support: The system supports maintenance teams in making informed choices by providing actionable insights and suggestions based on the data, eliminating the need for laborious manual computations.
- 5.Segmented Predictive Maintenance: • Predictive Upkeep: The system's capacity to anticipate possible problems in advance enables proactive scheduling of maintenance. Better maintenance planning, fewer unexpected failures, and more efficient use of maintenance resources are all outcomes of this predictive method.
- • Financial Benefits: The system aids in cutting maintenance expenses and increasing overall cost-efficiency by minimising the frequency of human checks and eliminating unexpected breakdowns.
- Sixth, Thorough Data Analysis: • Data Visualisation: The system has state-of-the-art data visualisation capabilities for analysing and understanding diagnostic outcomes. Better decisions may be made with the assistance of this feature's analysis of fault patterns, trends, and system performance.
- Comprehensive Reporting: The system produces comprehensive reports on the engine's status, fault states, and diagnostic results, offering useful information for continuous tracking and evaluation.

4. RESULTS

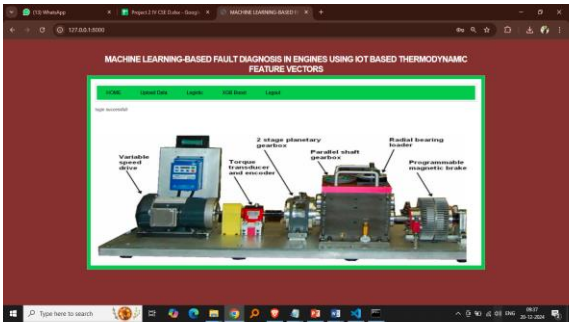


Figure4: shows that the admin page

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

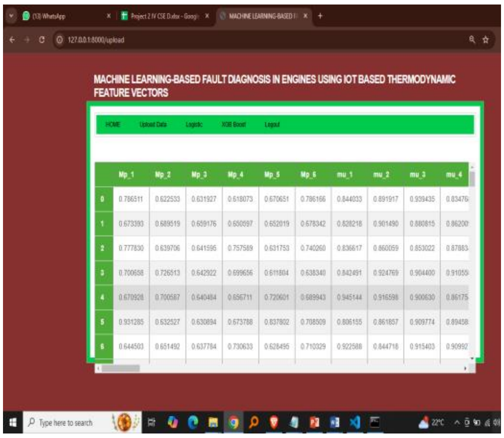


Figure 5: Represents all the rows and columns of the dataset.

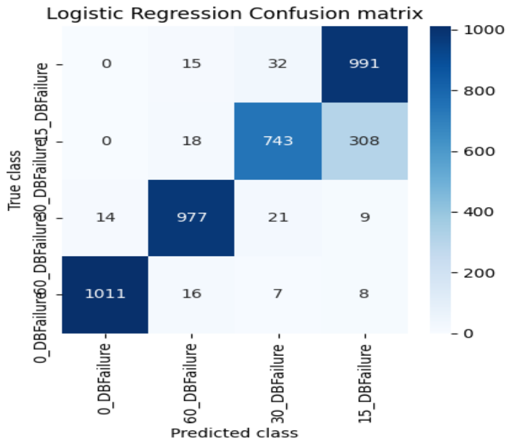


Figure 6: Confusion matrix of Logistic Regression Algorithm

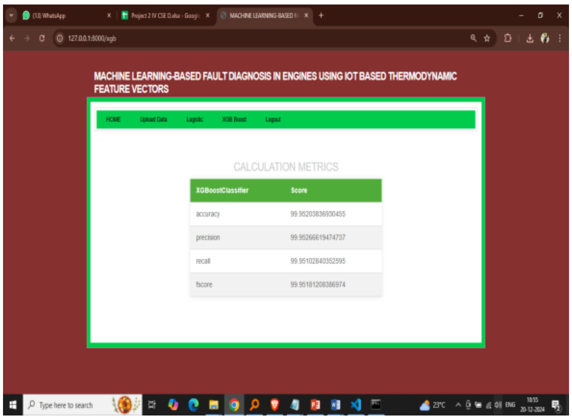


Figure 7: shows that the XGBoostClassifier

5. CONCLUSION

This research proved that Logistic Regression and XGBoost Classifier models, two machine learning applications, can accurately diagnose engine problems. The models were trained in a stable environment because to the thorough data preparation procedure, which included data cleaning, encoding, and train-test separation. With an impressive accuracy of about 89.26%, the Logistic Regression model demonstrated an excellent equilibrium among precision, recall, and F1-score. With a remarkable accuracy of 99.95%, the XGBoost Classifier demonstrated its better ability to handle complicated, high-dimensional thermodynamic feature vectors, surpassing its rival. The assessment metrics and visualizations, including confusion matrices, confirm the models' dependability and highlight the real-world value of using sophisticated algorithms in predictive maintenance systems. A smooth interface for real-time defect diagnostics was provided with the addition of user authentication and file upload features, which further enhanced the project. Early problem identification with data science approaches integrated into a web-based framework has been beneficial, which might reduce operational downtime and maintenance costs and open the door to further advanced research in industrial predictive analytics.

REFERENCES

[1] J. Kolb, K. Hameyer, "Classification of Tolerances in Permanent Magnet Synchronous Machines With Machine Learning", *IEEE Transactions on Energy Conversion*, vol.39, no.2, pp.831-838, 2024.

[2] Hristo Milushev, Georgi Enchev, Nikolay Djagarov, "Simulating a Lack of Phase Fault on SynRM and Integrate Results with Machine Learning", *2023 International Conference Automatics and Informatics (ICAI)*, pp.18-22, 2023.

[3] Billa Harikrishna, S Kanagalakshmi, "Machine Learning based Fault Detection for Non-Linear Systems", *2023 9th International Conference on Smart Computing and Communications (ICSCC)*, pp.647-652, 2023.

[4] David Molinero, Daniel Santamargarita, Emilio Bueno, Marta Marrón, Miroslav Vasić, "Artificial Neural Network Based Thermal

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Model for a Three-Phase Medium Frequency Transformer", *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*, pp.1-6, 2023.

[5] Moritz Benninger, Marcus Liebschner, Christian Kreischer, "Increasing the Robustness of Fault Detection for Induction Motors based on Neural Networks and the Winding Function Method", *2023 24th International Conference on the Computation of Electromagnetic Fields (COMPUMAG)*, pp.1-4, 2023.

[6] Antonio Femia, Giacomo Sala, Luca Vancini, Gabriele Rizzoli, Michele Mengoni, Luca Zarri, Angelo Tani, "A Machine-Learning-Based Interturn Short-Circuit Diagnosis for Multi-Three-Phase Brushless Motors", *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, vol.4, no.3, pp.855-865, 2023.

[7] B Rajesh Kumar, M Gowrisankar, S Ramana, P Dhatchina Aakash, S Aravind, "Load Fault Diagnosis in Induction Motor using Artificial Intelligence Algorithm", *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp.103-108, 2022.

[8] Ciro Attaianese, Matilde D'Arpino, Mauro Di Monaco, Luigi Pio Di Noia, "Modeling and Detection of Phase Current Sensor Gain Faults in PMSM Drives", *IEEE Access*, vol.10, pp.80106-80118, 2022.

[9] Qian Yang, Muhammed Ali Gultekin, Vahe Seferian, Krishna Pattipati, Ali M. Bazzi, Francesco A. N. Palmieri, Ravi Rajamani, Shailesh N. Joshi, Muhamed Farooq, Hiroshi Ukegawa, "Incipient Residual-Based Anomaly Detection in Power Electronic Devices", *IEEE Transactions on Power Electronics*, vol.37, no.6, pp.7315-7332, 2022.

[10] Rongchen Zhou, Weiwei Liu, Shuya Lei, Wensi Zhang, Lin Huang, "A Machine Learning Based Sameness Recognition Method for Power System Management Information", *2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, pp.3318-3322, 2021.

[11] Denis Bobylev, Tuhin Choudhury, Jesse O. Miettinen, Risto Viitala, Emil Kurvinen, Jussi Sopanen, "Simulation-Based Transfer Learning for Support Stiffness Identification", *IEEE Access*, vol.9, pp.120652-120664, 2021.

[12] Jiyu Zhang, Mutasim Salman, Wesley Zanardelli, Siddharth Ballal, Bojian Cao, "An Integrated Fault Isolation and Prognosis Method for Electric Drive Systems of Battery Electric Vehicles", *IEEE Transactions on Transportation Electrification*, vol.7, no.1, pp.317-328, 2021.

[13] Liangjun Feng, Chunhui Zhao, "Fault Description Based Attribute Transfer for Zero-Sample Industrial Fault Diagnosis", *IEEE Transactions on Industrial Informatics*, vol.17, no.3, pp.1852-1862, 2021.

[14] Sagar Verma, Nicolas Henwood, Marc Castella, Al Kassem Jebai, Jean-Christophe Pesquet, "Neural Networks based Speed-Torque Estimators for Induction Motors and Performance Metrics", *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pp.495-500, 2020.

[15] Thomas Guillod, Panteleimon Papamanolis, Johann W. Kolar, "Artificial Neural Network (ANN) Based Fast and Accurate Inductor Modeling and Design", *IEEE Open Journal of Power Electronics*, vol.1, pp.284-299, 2020.