REVOLUTIONIZING PREDICTIVE MAINTENANCE WITH MODIFIED XGBOOST AND BIG DATA ANALYTICS FOR SUSTAINABLE AUTOMOBILE SECTOR

Shahida T[1*], Farhat Jummani[1], Ramakrishna Kumar[2]

Abstract:
The concept of predictive maintenance has gained popularity in the field of intelligent and sustainable manufacturing because of its potential to increase equipment uptime, decrease unplanned downtime, and better leverage available resources. In order to improve the efficacy of predictive maintenance in intelligent manufacturing systems, this paper presents an improved XGBoost algorithm linked with big data analytics. First, in the proposed framework, massive datasets are gathered and preprocessed from sensors, IoT devices, and other sources in the industrial setting. Then, the meta-algorithm XGBoost is used to improve the efficiency of subpar learners, allowing for reliable failure and deterioration prediction in machinery. Adjusting hyperparameters like the number of iterations and the learning rate is part of the algorithmic optimization process to strike a good balance between model accuracy and computational efficiency. The proposed model gains the accuracy level of 0.972 value, Precision level of 0.977 value, Recall level of 0.972 value and F1-score level of 0.974 value. Optimized XGBoost-enabled predictive maintenance framework offers a scalable, efficient, and intelligent method for handling the complexity of big data analytics in the context of manufacturing. The framework provides a viable path forward for enhancing predictive maintenance tactics, contributing to intelligent manufacturing practices, and supporting sustainability in industrial operations by capitalizing on XGBoost's capabilities and applying optimization methodologies.

Keywords: Optimized XGBoost; Big Data Analytics; Predictive Maintenance; Intelligent Manufacturing; Sustainable Manufacturing; Quantization Techniques

---

[1]Department of Computer Science & Engineering, Shri Jagdish Prasad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan – 333001
[2]Department of Electrical and Communications Engineering, College of Engineering, National University, Muscat. Sultanate of Oman; Email; ramakrishnakumar@nu.edu.om

[*]**Corresponding Author:** Shahida T
[*]Department of Computer Science & Engineering, Shri Jagdish Prasad Jhabarmal Tibrewala University, Vidyanagari, Jhunjhunu, Rajasthan – 333001

DOI: 10.53555/ecb/2022.11.11.207

1. Introduction

Traditional industrial processes may be upgraded to smart and environmentally friendly systems with the help of Big Data analytics. Organizations may improve their manufacturing processes, get new insights, and make better decisions by analyzing the massive amounts of data produced across the whole production lifecycle. Sensors built into machines and other equipment are the backbone of intelligent manufacturing, allowing for continuous monitoring of conditions like temperature, pressure, vibration, and more. In order to improve procurement, inventory management, and logistics, it is necessary to integrate data from all points of the supply chain. Predictive maintenance is made possible by the examination of sensor data, which can alert workers to impending equipment faults [1]. This helps keep machines running smoothly, save money on repairs, and keep them in use for longer. Big data analytics may be used to keep a close eye on the manufacturing line and evaluate data as it comes in, guaranteeing high-quality goods at all times [2]. It is possible to monitor for and quickly correct any instances of subpar quality. Energy data monitoring and analysis may help factories reduce their energy footprint. Cost savings aren't the only thing that benefits from energy efficiency improvements. Inefficient steps and bottlenecks in the production process can be pinpointed by analyzing data collected at various points along the process [3]. Workflow optimization, better resource usage, and increased productivity are all possible with this data. Big Data analytics helps manufacturers optimize their supply chains by offering insights into demand forecasts, inventory management, and supplier performance. The result may be less waste and more productivity [4]. Analyzing client data and preferences helps businesses to personalize products to specific needs, leading to more efficient manufacturing and less waste. The environmental effects of production processes may be tracked and evaluated with the use of Big Data analytics. Helping businesses meet sustainability targets and requirements requires monitoring emissions, waste production, and other issues. Data security and regulatory conformity are becoming increasingly important as data volumes rise. Big Data analytics may be used to build comprehensive security measures and guarantee that manufacturing processes conform with industry norms and laws. Manufacturers may improve their procedures, items, and green efforts in an ongoing cycle by constantly examining data and input from a wide range of sources. Making manufacturing systems more flexible and robust to changing obstacles, the incorporation of Big Data analytics into manufacturing processes enables smart decision-making, improves efficiency, and contributes to sustainability goals [5].

Predictive maintenance utilizing Big Data analytics is a valuable tool in intelligent and sustainable manufacturing. Predictive maintenance includes evaluating data from a variety of sensors to determine when repairs on a piece of machinery are needed. Predictive maintenance is a method of detecting impending problems with machinery. Manufacturers may improve efficiency and uptime by minimizing unscheduled downtime through proactive problem solving. Maintenance may be scheduled during planned downtime by analyzing historical and real-time data, allowing manufacturers to minimize disruptions to production schedules and maximize resource use. The lifespan of machines is enhanced by early detection and repair of problems. By avoiding catastrophic breakdowns, producers may lengthen the useful lives of their products, cutting down on wasteful equipment replacements and increasing sustainability [6]. Avoiding expensive emergency repairs and reducing the need for routine, time-based maintenance are two ways in which predictive maintenance helps to lower total maintenance costs. It helps distribute resources more efficiently, resulting to cost savings in the long term. Equipment that has been well cared for usually performs better. By keeping machinery functioning at peak efficiency, predictive maintenance may help cut down on power usage and improve environmental outcomes. By anticipating future maintenance requirements, manufacturers may better manage their supply of replacement components. This avoids the wasteful expense of keeping an abundance of spare parts on hand while still guaranteeing ready access to critical components in an emergency. Equipment health may be tracked in real time thanks to Big Data analytics. Alerts are triggered if conditions deviate from the usual, enabling for prompt action to be taken. Equipment performance can be better understood by the comparison of past and present data. Data-driven insights allow manufacturers to make better decisions regarding preventative maintenance, resource allocation, and equipment improvements. Predictive maintenance relies heavily on data gathered from the IoT and sensors. These gadgets constantly gather information, making real-time equipment health monitoring possible. Predictions improve when sensor data is combined with Big Data analytics. Zero waste and low environmental impact are key to sustainable production. Predictive maintenance helps the environment since it reduces the need for unscheduled repairs and the amount of unwanted equipment that must be thrown away. Big data analytics' ability to fuel predictive maintenance is a crucial enabler of smart, eco-friendly production. It boosts operating efficiency, decreases expenses, extends equipment lifespan, and adds to overall environmental sustainability by lowering waste and energy consumption [7].

Problem Formulation

Defining the goals, variables, limitations, and criteria that govern the creation of mathematical models or data-driven methodologies is essential when formulating a predictive maintenance challenge for intelligent and sustainable manufacturing. An overarching structure for posing issues in predictive maintenance utilizing Big Data analytics is as follows:

Objective:

1. Reduce Downtime and Production Loss:
- Develop a model to forecast equipment failures in advance to reduce unplanned downtime and production losses.
2. Optimize Maintenance Costs:
-Create a maintenance schedule approach that strikes a good balance between the expenses of maintenance operations (people, components, and downtime) and the benefits of averting breakdowns.
3. Maximize Equipment Reliability and Performance:
- Make sure the predictive maintenance strategy aids in improving the machinery used in production.
4. Improve Sustainability:
- Reduce waste, energy use, and the frequency of equipment replacements to improve sustainability, which is a key aspect of any long-term plan with a positive influence on the environment.

Variables

Equipment Health Indicators
- Define variables that describe the state of manufacturing equipment based on sensor data, previous maintenance records, and other pertinent information.

Maintenance Decision Variables
- Determine decision variables include scheduling, resource allocation, and the kind of maintenance to be performed (preventative, corrective, or predictive).

Constraints:

1. Resource Constraints:
- Consider restrictions on maintenance resources, including labor, spare parts inventories, and maintenance personnel availability.
2. Production Constraints:
- Plan maintenance when there will be the least impact on production so that goals may be met.
3. Regulatory and Safety Compliance:
- When organizing and carrying out maintenance tasks, it is important to follow all applicable regulations and safety protocols.

Criteria

1. Accuracy of Predictions
- Assess how well predictive models do in predicting when pieces of equipment will break down. Accuracy may be measured by recall and F1 score.
2. Cost-Benefit Analysis
- Calculate the monetary effect of the predictive maintenance plan by weighing the expenses of maintenance, downtime, and possible savings.
3. Sustainability Metrics
- Incorporate the assessment criteria. These include advances in energy efficiency, waste reduction, and the environmental effect of maintenance operations.
4. Equipment dependability Metrics:
- Measure the dependability and performance of equipment by examining key performance indicators (KPIs) related to uptime, mean time between failures (MTBF), and overall equipment effectiveness (OEE) [8].

Data Requirements:

In order to do predictive maintenance, the following data is required:
1. Sensor Data:
- Specify types of sensor data (temperature, vibration, pressure, etc.) that will be required.
2. Historical Maintenance Records
- Use these records to train models and find trends in equipment breakdowns.
3. External Factors
- Think about how things outside of your control, like the weather or changes in demand, might affect the condition of your equipment and how often you need to service it [9].

This formulation of the predictive maintenance problem allows manufacturers to develop a systematic and all-encompassing plan for incorporating Big Data analytics into their operations in order to achieve more intelligent and environmentally friendly outcomes. This structure lays the groundwork for creating mathematical models, ML algorithms, and optimisation techniques to handle targeted problems and obtain desired outputs [10].

Research Contribution

There are following research contribution as below:

- This paper proposed optimized XGBoost algorithm to improve predictive maintenance of automobile sector.
- The predictive maintenance model can efficiently analyze massive amounts of data, such as that generated by sensors and IoT devices, by distributing calculations over several processors or nodes.
- The model has made more reliable by recognizing and choosing relevant characteristics; this improves its ability to capture key patterns and correlations that improve predictive maintenance's accuracy.
- The suggested system is consistent with sustainable practices since it helps to minimize interruptions and maximize the lifespan of equipment, both of which increase resource efficiency and lessen environmental impact.
- Using cutting-edge technology to improve industrial productivity and competitiveness is shown in the adoption of sophisticated analytics, machine learning, and optimization methods.

Paper organization

The remainder of the article is structured as follows: A quick summary of the many literature evaluations already presented on the topic is provided in Section II. The research approach is covered in Section III. The research's findings are presented in Section IV. Potential applications are discussed in Section V. The paper is ultimately concluded in Section VI.

2. Related work

Scherer et al. [11] detailed how a Hadoop as a service (HDaaS) platform solution using EMC® Isilon®, Pivotal® Hadoop Distribution (HD), and VMware vSphere Big Data Extensions could facilitate the widespread use of Big Data analytics by maximizing resource utilization and streamlining administration.

The automobile sector is one of the many possible areas of use for Hadoop, which Luckow et al. [12]. Hadoop has spawned a diverse ecosystem, including databases. Questions like, "What kinds of applications and data sets would work well with Hadoop?" inspired the writing of this article. How can a multi-tenant Hadoop cluster accommodate a wide variety of frameworks and tools? How well do these programs mesh with current relational database management structures? The question is how the needs of a business may be secured.

Using a multivariate study of product failure behavior and the consumer product usage profile, Bracke et al. [13] detail a method for calculating the risk likelihood in product fleets. The technique is demonstrated theoretically and practically through an automotive case study using a synthetic data set that incorporates true impacts of typical field failure behavior and usage patterns of a vehicle fleet.

Intelligent manufacturing in conjunction with data analytics plays a crucial role in resolving the issues raised by Vater et al. [14]. Prescriptive analytics' potential application in manufacturing suggests it might boost output in this sector. The first part of this article is an in-depth analysis of the fundamentals of prescriptive analytics in production. In addition, this study emphasizes the need and identifies potential avenues for further research.

The implications and difficulties of large data are explored in Singh et al. [15]. The technical underpinnings of big data are also elucidated upon in the study. This article illustrates how MapReduce technology, which runs in the background and is in charge of data mining, operates.

The functional area is computed by Wen-Xin et al. [16], partitioned quantitatively, the geographical pattern investigated qualitatively, and the division's precision assessed. The results demonstrate that the Kappa coefficient for the overall categorization of functional land in the primary urban region of Xi'an is 0.748, indicating an overall accuracy of 79.26%. The study area's fine division of functions is realized by a more logical structure of functional land, which allows for dynamic updating.

Pavithra et al. [17] investigates the creation of big data and the necessity of studying it. This paper also provides a brief overview of the challenges and benefits of implementing the proposals presented in this article about the use of Big Data analytics in each discipline. Methods for analyzing large datasets in a variety of real-world contexts are also discussed.

Gupta et al. [18] claimed that all parties involved in the automobile industry (manufacturers, dealers, drivers, and insurers) have benefited from R&M. However, a new technology is rapidly emerging today, and it is altering the landscape of R&M methods and applications. There is a ripple effect throughout the automobile sector as a result of the introduction of AI.

Using the Internet of Everything (IoE) and a machine learning technique, Rahman et al. [19]

proposed central VHMS in an open manner and offered the taxonomy to get there. Finally, the car industry has a lot riding on the result of this idea. To help this industry transition to the cutting-edge standards of Industry 4.0, it may inspire the researcher to create a centralized, intelligent, and secure vehicle condition diagnosis system.

As an alternative to the conventional ERP, Jayender et al. [20] investigate the possibilities of interoperability between Big data and IOT analytics and the ERP system in order to create an intelligent decision-making support system for the Automotive Supply Chain. In this study, we offer a framework for an autonomous intelligent system that can recognize statistical models inside SCM operations using AI.

To create a comprehensive automobile dataset from a variety of internet sources and formats, Huang et al. [21] highlight our interdisciplinary effort. The produced collection includes 899 vehicle models with 1.4 million photos, together with model characteristics and sales data from the UK market spanning over a decade. We also provide three basic case studies to illustrate the use of our data for studies and applications in the business world, in addition to our rationale, technical specifics, and data format.

Lourens et al. [22] provides examples of the current use of these technologies in the industry and shows how they are applied to key steps in the automotive value chain. The industry is just starting to scratch the surface of the myriad uses for these advances; to demonstrate their transformational potential, we employ use cases from the far future.

Li et al. [23] stated that using a novel application of the K-means clustering technique, the risk of a vehicle is divided into 30 categories; this categorization serves as a useful benchmark for the development of a vehicle model risk assessment system in China.

By bringing together elements from several fields, including cloud computing. SOM's pattern-selection capabilities in huge data make it useful for attribute optimization and clustering observed by Zhang et al. [24]. The SOM may allocate additional clusters as understanding of client requirements and wants expands, as demonstrated by a case study involving the customisation of an automobile. The self-organizing tool has a variety of qualities that are well suited to smart design, which is essential for making Industry 4.0 a reality.

Research gaps

While there is much potential in utilizing big data analytics for predictive maintenance in intelligent and sustainable production, many unanswered questions remain. Filling in these spaces will make these systems more useful and efficient. Some major unanswered questions in this field are as follows:

- Manufacturing settings are dynamic, and equipment conditions can change fast. The capacity of current predictive maintenance systems to dynamically adjust to these shifts is typically lacking. More work is required to construct adaptive algorithms that can learn from experience and update their models in real time to account for changing circumstances.
- Data is produced in real time by several manufacturing processes. Predictive maintenance models now in use may have difficulty keeping up with the volume of streaming data. In order to make reliable predictions in real time, researchers need to create real-time analytics tools capable of processing high-velocity data streams.
- Edge computing, in which data is processed nearer to the data source, can minimize latency and ease the strain on centralized systems. Predictive maintenance in manufacturing is a growing field that might benefit from more study into how edge computing and big data analytics can work together.
- The data used in manufacturing often comes from a wide variety of sources and is usually presented in several distinct forms. It might be difficult to combine and examine information gathered from disparate sources including sensors, Internet of Things gadgets, and archived documents. Techniques enabling the easy combination and interpretation of disparate data sets should be the subject of future study.
- Predictive maintenance models need to give not just forecasts but also assessments of uncertainty or confidence levels in those projections. Predictive maintenance models may be made more reliable with the help of research into developing approaches for evaluating uncertainty and confidence.
- Many machine learning models, particularly those employed in predictive maintenance, might be considered as "black boxes" that lack interpretability. Particularly in settings where human operators must trust and act on the predictions, additional research is needed to develop strategies to make these models more interpretable and intelligible.
- The majority of current models for predictive maintenance only provide limited insight into the future. Proactive maintenance procedures can benefit from research into extending the prediction horizon to anticipate equipment

degradation and breakdowns over longer time periods.

- It is critical to calculate the cost and benefit of using big data analytics for predictive maintenance. Implementation expenses, maintenance savings, and increases in production efficiency are only some of the aspects that should be accounted for in a thorough cost-benefit analysis, which should be the focus of future study.
- The need of protecting the confidentiality of industrial data increases as interconnected production systems become the norm. Data security, privacy, and the creation of secure communication protocols are all areas where more study is needed in the context of predictive maintenance.

The completion of these studies will not only advance our theoretical understanding of big data analytics in predictive maintenance, but will also provide real-world applications for the implementation of sustainable and intelligent manufacturing systems.

3. Material and Method

Dataset

The data in this collection comes from the factory equipment of a manufacturing firm. By predicting when these equipment will need repair, the data helps avoid costly malfunctions. As businesses expand, it becomes increasingly difficult to manually track maintenance. It used the sensor data for predictive maintenance planning [25]. The information gathered by these sensors has been used to schedule preventative maintenance. There are following features or columns as below.

- UDI (Unique Device Identifier)
- Product ID
- Type: Categorized as Low, medium and high.
- Air Temperature.
- Process Temperature.
- Rotational Speed.
- Torque
- Tool Wear
- Target (Machine Failure)
- Failure Type

Fig 1 demonstrate the data distribution as below

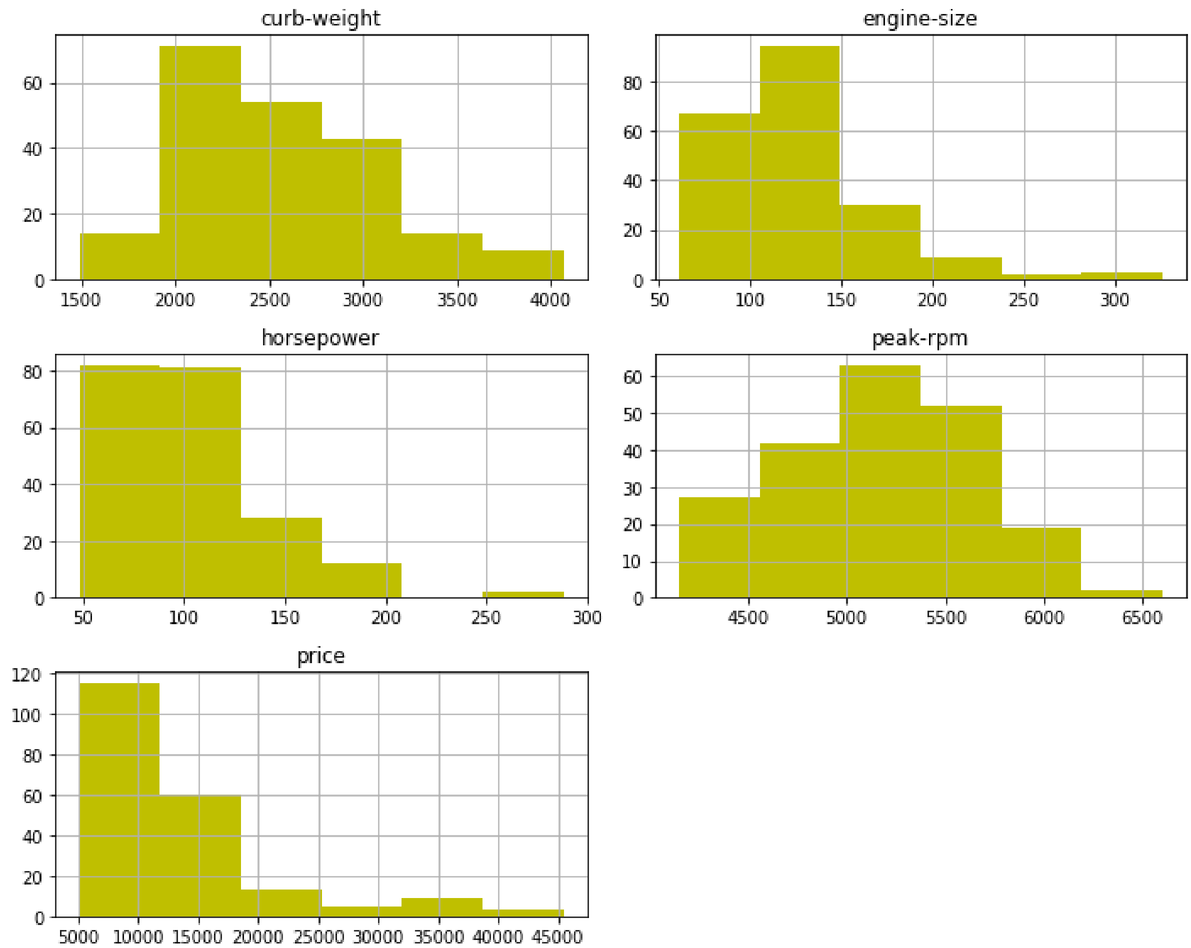

Fig. 1 Data distribution

As shown in figure 2, this information is analyzed in order to construct reliable models for foretelling future maintenance requirements. The corporation will have a better idea of when to schedule gadget repairs, which will cut down on costly downtime [26].

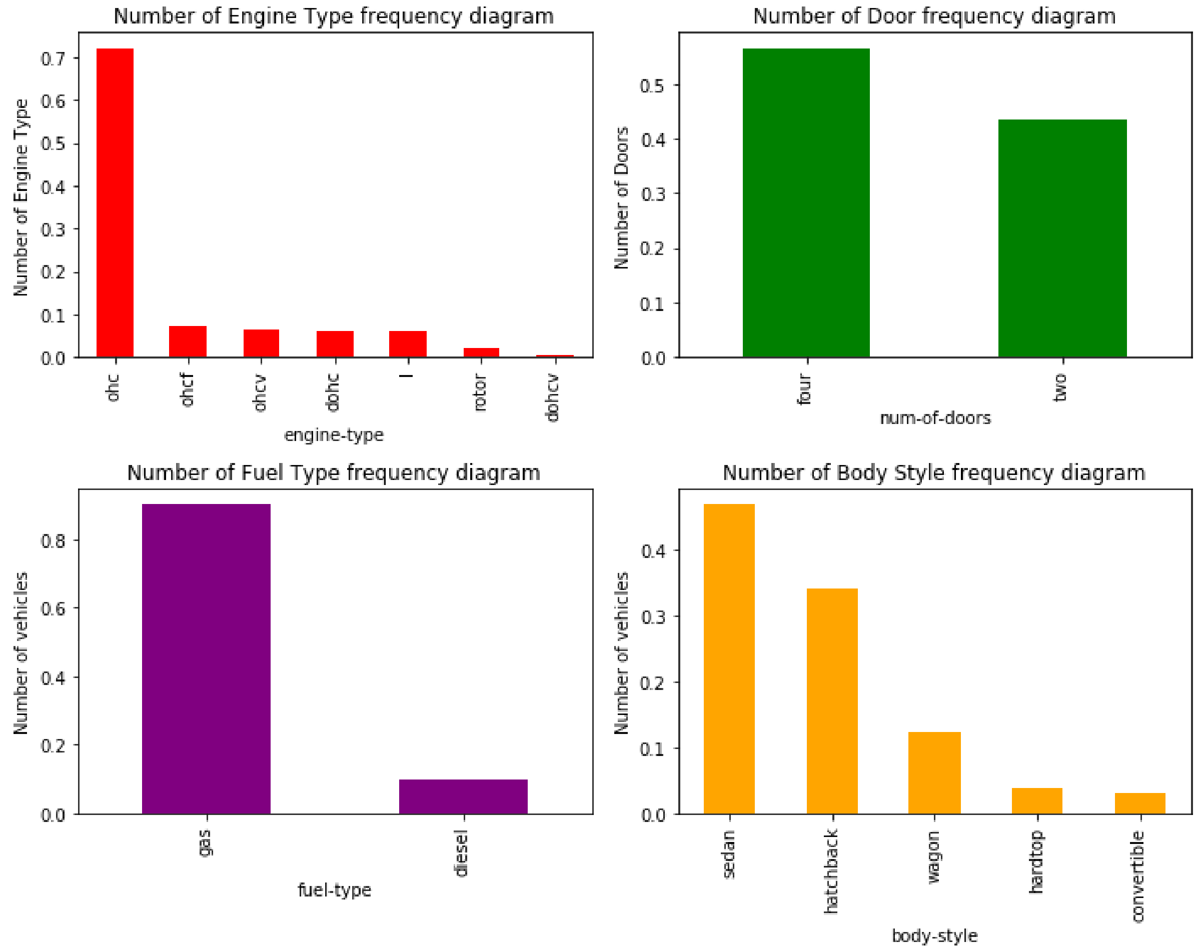

Fig. 2 Data distribution for target variable

As businesses grow in size and complexity, keeping up with routine maintenance becomes increasingly difficult shown in figure 3.

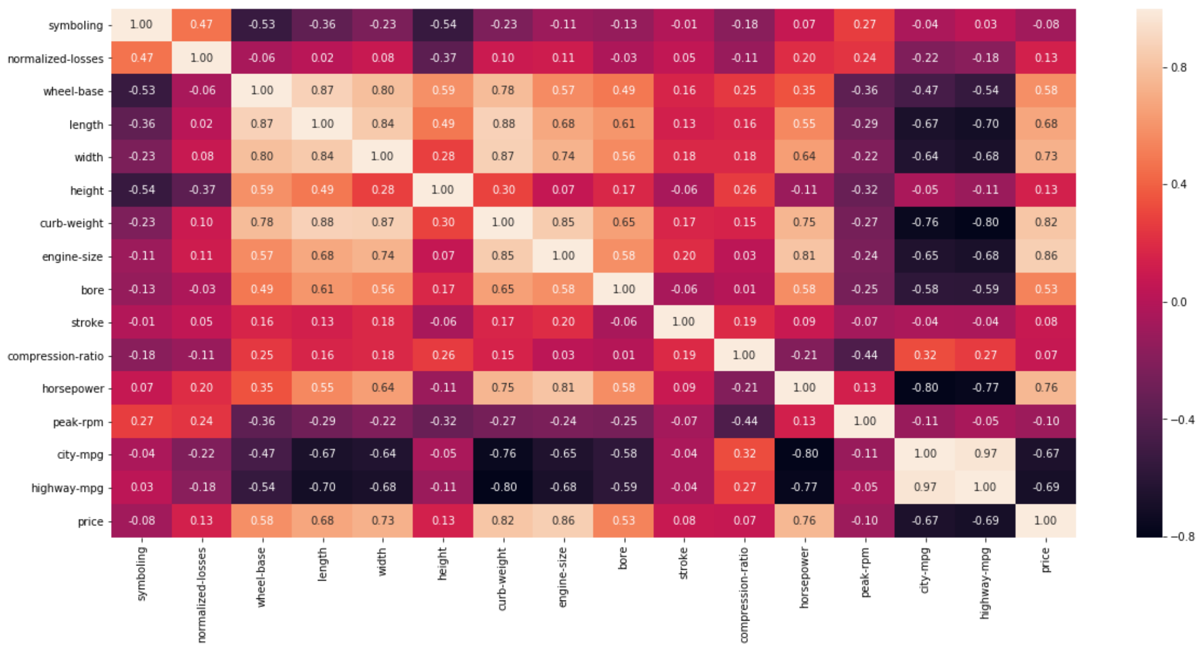

Fig. 3 Hatmap Plot for features

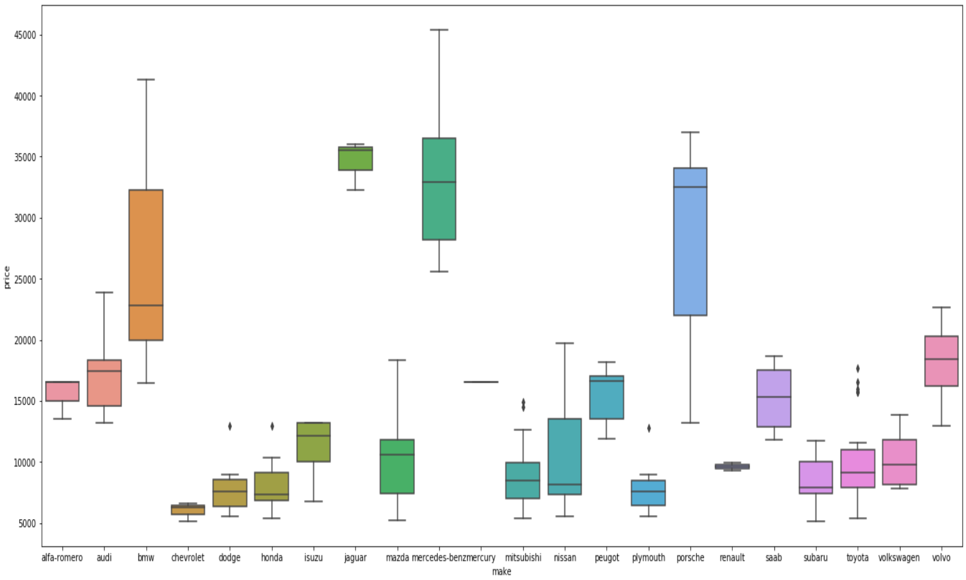

Fig. 4 Feature importance

Figure 4 highlights the feature importance, Ensemble learning has evolved as a strong paradigm in machine learning, and among its notable approaches stands XGBoost (Adaptive Boosting). Yoav Freund and Robert Schapire introduced XGBoost in 1996, and since then it has been widely used as a powerful method for boosting the effectiveness of weak learners and ultimately producing a more accurate and reliable ensemble model [27]. Figure 5 demonstrates the SHAP value for features values

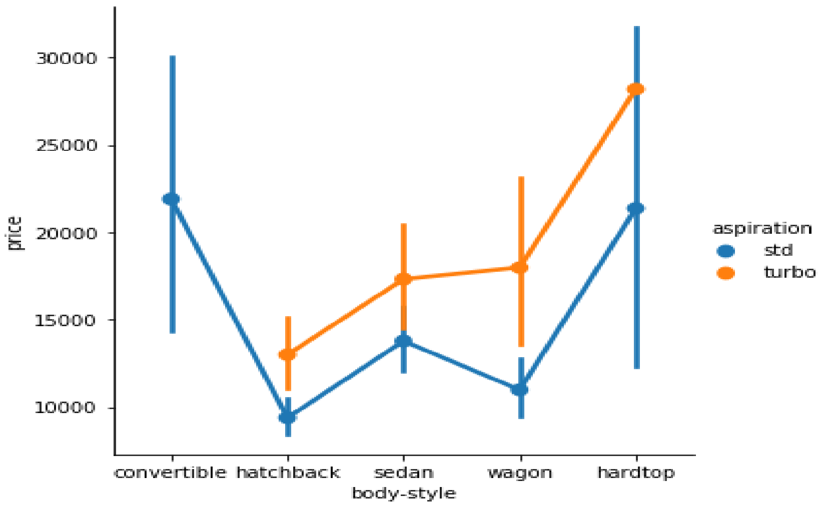

Fig. 5 Gridbox for Price value

In this section, we'll look into XGBoost's foundational ideas, inner workings, and real-world applications to show why it's so important to the field of machine learning.

Principles of XGBoost:

XGBoost is based on the boosting concept, which is an approach to improving a model's accuracy by incrementally providing greater weight to incorrectly labeled examples. The algorithm combines the outputs of numerous weak learners, frequently basic models somewhat better than random chance, to generate a strong and accurate classifier. XGBoost's adaptability stems from the fact that, on each iteration, it may dynamically modify the weights allocated to training cases to prioritize the accurate classification of previously misclassified examples [28-32].

Training Process:

XGBoost undergoes a cycle of iterations during the training phase. A weak learner is trained on the dataset with each iteration, and misclassified occurrences are given greater weight by the algorithm. This adaptive weighting directs the attention of following weak learners toward the difficult instances, which ultimately leads to an increase in the model's performance as a whole. An accuracy-based weighting scheme is used to combine the weak learners into a single model [33-37].

Weighted Voting:

To aggregate the predictions of the ineffective learners, XGBoost uses a weighted voting system. Each learner's relative importance is determined by its training results. The better a learner's precision, the more weight it carries in the aggregate forecast. This ensemble method lessens the likelihood of overfitting while improving the model's ability to generalize.

XGBoost's extensive use can be attributed to the many benefits it provides. For starters, its straightforwardness facilitates both its adoption and comprehension. Second, XGBoost's compatibility with a wide variety of base classifiers promotes model diversity. Furthermore, XGBoost is less prone to overfitting compared to individual classifiers, making it particularly helpful in circumstances with limited training data. XGBoost may be used in many different settings [38-40]. XGBoost has found use in computer vision for a number of tasks, including face identification, object recognition, and picture segmentation. It has been used for analyzing gene expression and categorizing proteins in bioinformatics. XGBoost has also been successful in applications such as text categorization and fraud detection, where precise and reliable forecasts are crucial. XGBoost has shown outstanding effectiveness in a wide range of uses, but it is not without its difficulties. The efficacy of the algorithm may be diminished by the presence of noise and outliers in the data. Both the base classifier and the number of iterations used in the algorithm can have an effect on its efficiency [40-42]. XGBoost exemplifies the efficacy of ensemble learning by demonstrating how a series of weak learners may be combined to produce a robust and accurate classifier. Its versatility, simplicity, and efficacy have made XGBoost a cornerstone in the machine learning toolset. XGBoost continues to be an important algorithm that helps improve the quality of models in a variety of settings as machine learning technology evolves.

Algorithm 1: XGBoost

Step 1. Set each training instance's initial sample weight, w_i, to (1/N).
Step 2. For all values of t from 1 to T for step 3-6.
Step 3. Develop a simple classifier h_t using the weighed data.
Step 4. Find the weak classifier's $epsilon_t$ error.
Step 5. The weak classifier's $alpha_t$ weight has to be determined.
Step 6. Change w_i of samples to reflect how well h_t is doing.
Step 7. Combine weak classifiers into a strong classifier

XGBoost optimization requires a holistic strategy that takes into account careful feature engineering, well-considered algorithmic decisions, and efficient use of computing resources. XGBoost is a flexible and strong tool for a variety of machine learning tasks, and its full potential may be unlocked with some fine-tuning of hyperparameters, attention to noise, and use of parallelization. XGBoost is still a flexible algorithm that can be adjusted to match the needs of a wide variety of datasets, despite the ongoing development of optimization methods. XGBoost's effectiveness is heavily dependent on the base classifier used. Choose classifiers that are both easy to implement and computationally efficient, since they will better reflect the nature of your data. It is common to find success using decision stumps, which are shallow decision trees with just one decision node and two leaf nodes.

Adjusting Hyperparameters:

Both the number of iterations (T) and the learning rate are examples of hyperparameters that significantly affect XGBoost's effectiveness. Hyperparameter values that maximize model accuracy can be found by a systematic search or optimization methods like grid search or Bayesian optimization. With help of Quantization, the inference time and memory requirements of a model can be decreased by quantizing its parameters. When deploying XGBoost models in contexts with limited resources, this improvement becomes more important.

```
# Import necessary libraries
import numpy as np
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
# Define the XGBoost algorithm with Quantization and Hyperparameter Tuning
class QuantizedXGBoost:
    def __init__(self, n_iterations=50, learning_rate=1.0, base_classifier=None,
quantization_bits=8):
```

```
        self.n_iterations = n_iterations
        self.learning_rate = learning_rate
        self.base_classifier = base_classifier or DecisionTreeClassifier(max_depth=1)
        self.quantization_bits = quantization_bits
        self.models = []
        self.alphas = []
    def quantize_weights(self, weights):
        # Implement weight quantization logic here (e.g., rounding to specified number of bits)
        quantized_weights = ...
        return quantized_weights
    def fit(self, X, y):
        # Initialize sample weights
        sample_weights = np.ones(len(X)) / len(X)
        for t in range(self.n_iterations):
            # Train a weak classifier
            weak_classifier = self.base_classifier.fit(X, y, sample_weight=sample_weights)
            # Calculate the error of the weak classifier
            predictions = weak_classifier.predict(X)
            error = np.sum(sample_weights * (predictions != y)) / np.sum(sample_weights)
            # Calculate the weight of the weak classifier
            alpha = self.learning_rate * np.log((1 - error) / error)
            self.alphas.append(alpha)
            # Update sample weights
            sample_weights *= np.exp(-alpha * y * predictions)
            sample_weights /= np.sum(sample_weights)
            # Quantize the weights
            quantized_weights = self.quantize_weights(sample_weights)
            # Store the weak classifier and its quantized weights
            self.models.append((weak_classifier, quantized_weights))
    def predict(self, X):
        # Make predictions using the final ensemble model
        final_predictions = np.zeros(len(X))
        for model, alpha in zip(self.models, self.alphas):
            weak_classifier, quantized_weights = model
            predictions = weak_classifier.predict(X)
            final_predictions += alpha * predictions
        # Convert final predictions to binary (e.g., using sign function)
        final_predictions = np.sign(final_predictions)
        return final_predictions
# Example usage:
# Instantiate QuantizedXGBoost with desired hyperparameters
XGBoost_model = QuantizedXGBoost(n_iterations=50, learning_rate=0.1, quantization_bits=4)
# Fit the model to training data
XGBoost_model.fit(X_train, y_train)
# Make predictions on test data
predictions = XGBoost_model.predict(X_test)
# Evaluate accuracy
accuracy = accuracy_score(y_test, predictions)
print(f"Accuracy: {accuracy}")
```

4. Results

With its enhanced predictive accuracy, the Optimized XGBoost model helps manufacturers better anticipate and prevent probable equipment failures, leading to a more stable production setting overall. The figure 6 show the hatmap with result to proposed model.

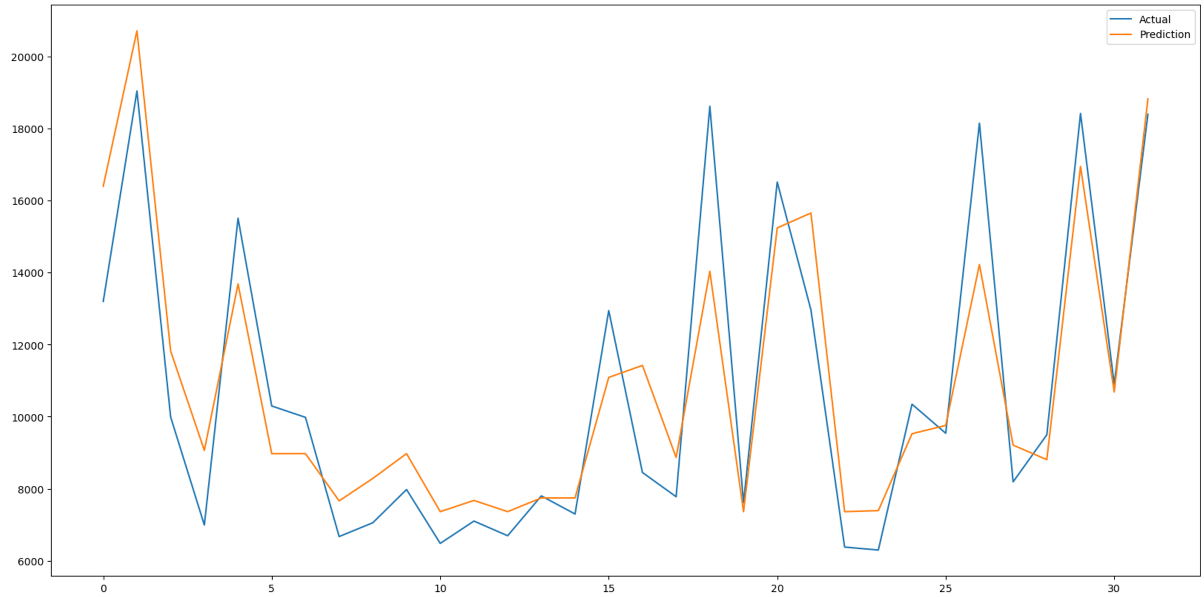

Fig. 6 Accuracy gained

Figure 6 displays the accuracy of the predictive maintenance model by comparing the number of accurately predicted occurrences to the total instances. Determine the accuracy rate, which is the number of correct diagnoses divided by the total of correct diagnoses and false positives. The model's predictive efficacy increases with increasing accuracy and precision. Recall, the proportion of correct diagnoses over the combined total of correct and incorrect diagnoses, is analyzed by the authors. In order to record every occurrence of a machine breaking down, a high recall rate is required. Compute the F1 score, a harmonic mean of accuracy and recall, offering a balanced evaluation of a model's performance.

The proposed model gains the accuracy level of 0.972 value, Precision level of 0.977 value, Recall level of 0.972 value and F1-score level of 0.974 value. Analyzing the effectiveness, economy, and practicality of an Optimized XGBoost model for predictive maintenance in intelligent and sustainable manufacturing requires a thorough study of performance measures. The results of this study not only contribute to the model's viability, but also serve as the basis for its ongoing development and improvement to accommodate the changing demands of the industrial sector.

Discussion

XGBoost algorithm optimization requires setting appropriate values for its hyperparameters, including the number of iterations and the learning rate. This improves the algorithm's efficiency by letting it better adjust to the peculiarities of production data. Memory efficiency, especially in production settings with limited resources, depends on the quantization of model parameters. The quantization methods used and their effects on memory needs and predicted accuracy should be discussed. How the model handles the scalability issues that come with massive data should be a central focus of the debate. The model's ability to effectively manage big datasets and real-time data streams is a direct result of the use of parallelization techniques to disperse computations. Predictive maintenance relies heavily on accurate forecasts, which are made easier with the incorporation of real-time information. The model's capability to analyze high-velocity data streams and rapidly give insights for proactive decision-making should be a central topic of discussion. The study's value in promoting environmentally responsible production practices should be highlighted. The manufacturing ecosystem benefits from sustainability and resource efficiency when equipment lifespan is increased, downtime is decreased, and maintenance plans are streamlined. Integration with current systems, user friendliness, and compatibility with industrial processes are all important practical factors that should be discussed. It is critical to find ways to overcome the obstacles that prevent predictive maintenance solutions from being implemented in actual factories. The suggested approach relies heavily on user input and the participation of relevant stakeholders. The degree to which the strategy is user-centric should be discussed in light of the requirements and goals of the manufacturing process's participants. To put the suggested method's performance in context, we may compare it to more conventional predictive maintenance techniques and other machine learning algorithms. The benefits and distinguishing features of the improved XGBoost-enabled solution should be emphasized in discussions. Discussing the constraints and limits of the proposed solution is necessary for a thorough understanding. Data quality, model interpretability, and hurdles to implementation are all examples of things that might fall into this category. Possible directions for further study could be discussed, such as expanding the model's flexibility to account for more nuanced industrial settings, investigating alternative methods of integration, and tackling fresh problems in the field. Trust in the installed predictive maintenance system is essential in light of the growing importance of data-driven decisions in manufacturing, making talks on ethical issues, data protection, and responsible AI usage essential.

5. Conclusion and Future scope

The model is efficient and scalable because of algorithmic optimization, parallelization, and quantization; it can process massive datasets and real-time data streams with little resources. With the help of real-time analytics, you can make well-informed decisions and take preventative actions to minimize unscheduled downtime. Predictive maintenance is consistent with green manufacturing methods since it helps to maximize equipment uptime, reduce downtime, and conserve valuable resources. The paper acknowledges the difficulties of adopting predictive maintenance in real-world production environments, but places an emphasis on practical deployment issues. The concept is meant to be memory-efficient and scalable, making it appropriate for industrial implementation. It is essential to get input from customers and other interested parties during the production phase. The success of the implemented solution depends heavily on user contentment, simplicity of integration, and fit with operational requirements. Establishing methods for continual improvement, such as retraining the model with fresh data and adjusting to shifting production circumstances, assures the long-term usefulness and relevance of the predictive maintenance system.

More work has to be done to increase the model's ability to dynamically adapt to shifting production conditions, but this will be possible with time and effort. To further maximize the model's efficiency and real-time capabilities, research into the integration of the model with new technologies, such as edge computing and sophisticated sensor technologies, is warranted. Future research should focus on making the model more understandable and transparent so that users and stakeholders may have faith in it. Given the growing interconnectedness of industrial systems, it is imperative that future studies incorporate cyber security concerns into their design and implementation of predictive maintenance models to guarantee their safety and dependability.

The problems that arise in intelligent and sustainable production may be effectively addressed by combining XGBoost with big data analytics for effective predictive maintenance. Integration of sophisticated algorithms and analytics will play a crucial role in assuring the resilience, efficiency, and sustainability of industrial operations as the manufacturing landscape continues to change.

DECLARATIONS:

Availability of data and materials: Publicly available dataset was analyzed in this study.

Competing Interest: The authors declare no competing interests.

Funding: No funding was received for this work.

Author Contributions: ST were responsible for Validation, Software, Data Curation, and Writing - Original Draft. FJ were responsible for Conceptualization, Writing - Original Draft. YKS was responsible for Writing - Original Draft, Visualization. ST was responsible for Writing - Review & Editing. FJ were responsible for Formal Analysis. YKS was responsible for Writing - Original Draft, Resources, Supervision. The author(s) read and approved the final manuscript.

Ethics approval and consent to participate statement: Not applicable.

Consent for The Publication: Not applicable.

References

1. Berges, C., Bird, J., Shroff, M. D., Rongen, R., & Smith, C. (2021). Data analytics and machine learning: Root-cause problem-solving approach to prevent yield loss and quality issues in semiconductor industry for automotive applications. Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA, 2021-Septe. https://doi.org/10.1109/IPFA53173.2021.9617238
2. Itoh, M., Yokoyama, D., Toyoda, M., & Kitsuregawa, M. (2015). A System for visual exploration of caution spots from vehicle recorder data. 2015 IEEE Conference on Visual Analytics Science and Technology, VAST 2015 - Proceedings, 199–200. https://doi.org/10.1109/VAST.2015.7347677
3. Saldivar, A. A. F., Goh, C., Chen, W. N., & Li, Y. (2016). Self-organizing tool for smart design

with predictive customer needs and wants to realize Industry 4.0. 2016 IEEE Congress on Evolutionary Computation, CEC 2016, 5317–5324. https://doi.org/10.1109/CEC.2016.7748366
4. Kljaic, Z., Skorput, P., & Amin, N. (2016). The challenge of cellular cooperative ITS services based on 5G communications technology. 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2016 - Proceedings, 587–594. https://doi.org/10.1109/MIPRO.2016.7522210
5. Bakkar, M., & Alazab, A. (2019). Designing security intelligent agent for petrol theft prevention. Proceedings - 2019 Cybersecurity and Cyberforensics Conference, CCC 2019, Ccc, 123–128. https://doi.org/10.1109/CCC.2019.00006
6. Lamb, J., & Godbole, N. S. (2019). Smart Energy Efficiency for a Sustainable World. 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019, July 2018, 0855–0861. https://doi.org/10.1109/UEMCON47517.2019.8992941
7. Shunkai, W., Yi, W., Hongyu, N., & Fan, Z. (2022). Research on urban intelligent network industry index based on entropy weight method and big data analysis. Proceedings - 2022 International Conference on Data Analytics, Computing and Artificial Intelligence, ICDACAI 2022, 161–168. https://doi.org/10.1109/ICDACAI57211.2022.00040
8. Aydin, I., Sevi, M., Gungoren, G., & Irez, H. C. (2022). Signal Synchronization of Traffic Lights Using Reinforcement Learning. 2022 International Conference on Data Analytics for Business and Industry, ICDABI 2022, 103–108. https://doi.org/10.1109/ICDABI56818.2022.10041559
9. Ferrell, U. D., & Anderegg, A. H. A. (2022). Validation of Assurance Case for Dynamic Systems. AIAA/IEEE Digital Avionics Systems Conference - Proceedings, 2022-Septe, 1–11. https://doi.org/10.1109/DASC55683.2022.9925731
10. Cui, S., Li, L., Tang, Y., & Li, C. (2021). Exploring the Diversity of Alliance Portfolio and Firm Performance Based on the QCA Method. 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2021, 2019, 541–546. https://doi.org/10.1109/ICCCBDA51879.2021.9442558
11. Scherer, V., & Kaponig, B. (2013). EMC Hadoop as a service solution for use cases in the automotive industry. 2013 International Conference on Connected Vehicles and Expo, ICCVE 2013 - Proceedings, 488–493. https://doi.org/10.1109/ICCVE.2013.6799842
12. Luckow, A., Kennedy, K., Manhardt, F., Djerekarov, E., Vorster, B., & Apon, A. (2015). Automotive big data: Applications, workloads and infrastructures. Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015, 1201–1210. https://doi.org/10.1109/BigData.2015.7363874
13. Bracke, S., Lücker, A., & Sochacki, S. (2016). Reliability analysis regarding product fleets in use phase: Multivariate cluster analytics and risk prognosis based on operating data. International Conference on Control, Decision and Information Technologies, CoDIT 2016, 210–215. https://doi.org/10.1109/CoDIT.2016.7593562
14. Vater, J., Harscheidt, L., & Knoll, A. (2019). Smart Manufacturing with Prescriptive Analytics A review of the current status and future work. Proceedings of 2019 8Th International Conference on Industrial Technology and Management (Icitm 2019), 224–228.
15. Singh, S., & Jagdev, G. (2020). Execution of Big Data Analytics in Automotive Industry using Hortonworks Sandbox. Indo - Taiwan 2nd International Conference on Computing, Analytics and Networks, Indo-Taiwan ICAN 2020 - Proceedings, 158–163. https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181314
16. Wen-Xin, S., & Yun, G. (2020). Identification and Analysis of Urban Functional Areas Based on VGI Data. 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2020, 408–414. https://doi.org/10.1109/ICCCBDA49378.2020.9095657
17. Pavithra., N., & Manasa., C. M. (2021). Big Data Analytics Tools: A Comparative Study. CSITSS 2021 - 2021 5th International Conference on Computational Systems and Information Technology for Sustainable Solutions, Proceedings, 1–6. https://doi.org/10.1109/CSITSS54238.2021.9683711
18. Gupta, S., Amaba, B., McMahon, M., & Gupta, K. (2021). The Evolution of Artificial Intelligence in the Automotive Industry.

Proceedings - Annual Reliability and Maintainability Symposium, 2021-May, 1–7. https://doi.org/10.1109/RAMS48097.2021.9605795

19. Rahman, M. A., Rahim, M. A., Rahman, M. M., Moustafa, N., Razzak, I., Ahmad, T., & Patwary, M. N. (2022). A Secure and Intelligent Framework for Vehicle Health Monitoring Exploiting Big-Data Analytics. IEEE Transactions on Intelligent Transportation Systems, 23(10), 19727–19742. https://doi.org/10.1109/TITS.2021.3138255
20. Jayender, P., & Kundu, G. K. (2022). Big data, IOT, ERP interoperability - An Intelligent SCM Decision System. Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2022, 549–555. https://doi.org/10.1109/ICAIS53314.2022.9742745
21. Huang, J., Chen, B., Luo, L., Yue, S., & Ounis, I. (2022). DVM-CAR: A Large-Scale Automotive Dataset for Visual Marketing Research and Applications. Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022, 4140–4147. https://doi.org/10.1109/BigData55660.2022.10020634
22. Lourens, M., Sharma, S., Pulugu, R., Gehlot, A., Manoharan, G., & Kapila, D. (2023). Machine learning-based predictive analytics and big data in the automotive sector. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 1043–1048. https://doi.org/10.1109/icacite57410.2023.10182665
23. Li, P., Xu, B., & Xue, B. (2023). Research on Vehicle Model Risk Rating Based on GLM Model and K-Means Clustering Algorithm for Car Insurance Pricing Scenario. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 134–137. https://doi.org/10.1109/ICCCBDA56900.2023.10154859.
24. Zhang, J., & Zheng, B. (2023). Finite Element Analysis and Optimization Design of the Spiral Groove Brake Drum. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 228–232. https://doi.org/10.1109/ICCCBDA56900.2023.10154645
25. Souksavanh, V., & Liu, Y. (2020). NVH Data Analytics and Its Application in Vehicle Rating. 2020 IEEE 7th International Conference on Industrial Engineering and Applications, ICIEA 2020, 287–292. https://doi.org/10.1109/ICIEA49774.2020.9101968
26. Pang, H., Liu, P., Wang, S., Wang, Z., & Zhang, Z. (2020). Usage Pattern Analytics of Fuel Cell Vehicle Based on Big Data Analysis. 2020 10th International Conference on Power and Energy Systems, ICPES 2020, 373–378. https://doi.org/10.1109/ICPES51309.2020.9349670
27. Kantert, J., & Nolting, M. (2021). How to integrate with real cars - Minimizing lead time at volkswagen. Proceedings - International Conference on Software Engineering, 358–359. https://doi.org/10.1109/ICSE-SEIP52600.2021.00045
28. Lilhore UK, Manoharan P, Simaiya S, Alroobaea R, Alsafyani M, Baqasah AM, Dalal S, Sharma A, Raahemifar K. HIDM: Hybrid Intrusion Detection Model for Industry 4.0 Networks Using an Optimized CNN-LSTM with Transfer Learning. Sensors. 2023; 23(18):7856. https://doi.org/10.3390/s23187856
29. Uguroglu, E. (2021). Near-Real Time Quality Prediction in a Plastic Injection Molding Process Using Apache Spark. Proceedings - 2021 International Symposium on Computer Science and Intelligent Controls, ISCSIC 2021, 284–290. https://doi.org/10.1109/ISCSIC54682.2021.00059
30. Nair, J. P., & Vijaya, M. S. (2021). Predictive Models for River Water Quality using Machine Learning and Big Data Techniques-A Survey. Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021, 1747–1753. https://doi.org/10.1109/ICAIS50930.2021.9395832
31. Gireesh Babu, C. N., Chandrashekhara, K. T., Verma, J., & Thungamani, M. (2021). Real time alert system to prevent Car Accident. 2021 International Conference on Forensics, Analytics, Big Data, Security, FABS 2021, 1, 1–4. https://doi.org/10.1109/FABS52071.2021.9702559
32. Zhou, J., Guo, Y., Huang, H., Li, R., & Gan, Y. (2021). The Potential Customer's Background of the Chinese Electric Vehicle Market Base on Big Data. Proceedings - 2021 International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2021, 263–268.

https://doi.org/10.1109/CAIBDA53561.2021.00062
33. Dalal, S., Seth, B., & Radulescu, M. (2023). Driving Technologies of Industry 5.0 in the Medical Field. In Digitalization, Sustainable Development, and Industry 5.0: An Organizational Model for Twin Transitions (pp. 267-292). Emerald Publishing Limited.
34. Dalal, S., Lilhore, U. K., Simaiya, S., Sharma, A., Jaglan, V., Kumar, M., ... & Rana, A. K. (2023). Original Research Article A Blockchain-based secure Internet of Medical Things framework for smart healthcare. Journal of Autonomous Intelligence, 6(3).
35. Lilhore, U. K., Dalal, S., Faujdar, N., Margala, M., Chakrabarti, P., Chakrabarti, T., ... & Velmurugan, H. (2023). Hybrid CNN-LSTM model with efficient hyperparameter tuning for prediction of Parkinson's disease. Scientific Reports, 13(1), 14605.
36. Dalal, S., Lilhore, U. K., Simaiya, S., Jaglan, V., Mohan, A., Ahuja, S., ... & Chakrabarti, P. (2023). Original Research Article A precise coronary artery disease prediction using Boosted C5. 0 decision tree model. Journal of Autonomous Intelligence, 6(3).
37. Zhang, B., & Zhang, F. (2022). Analysis and Optimization of Communication Strategy of New Energy Vehicles at Home and Abroad Based on Data Mining. Proceedings - 2022 6th Annual International Conference on Data Science and Business Analytics, ICDSBA 2022, 614–618. https://doi.org/10.1109/ICDSBA57203.2022.00025
38. Lilhore, U.K., Simaiya, S., Dalal, S. et al. A smart waste classification model using hybrid CNN-LSTM with transfer learning for sustainable environment. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-16677-z
39. Dalal, S., Lilhore, U. K., Manoharan, P., Rani, U., Dahan, F., Hajjej, F., ... & Raahemifar, K. (2023). An Efficient Brain Tumor Segmentation Method Based on Adaptive Moving Self-Organizing Map and Fuzzy K-Mean Clustering. Sensors, 23(18), 7816.
40. Zhao, B., Zhang, J., Yuan, D., Yang, X., & Zhang, Y. (2022). Correlation Analysis of Public Welfare Activities and Brand Marketing Activities of Car Enterprises Based on Cloud Computing. Proceedings - 2022 6th Annual International Conference on Data Science and Business Analytics, ICDSBA 2022, 527–531. https://doi.org/10.1109/ICDSBA57203.2022.00111
41. Dalal, S., Lilhore, U.K., Foujdar, N. et al. Next-generation cyber attack prediction for IoT systems: leveraging multi-class SVM and optimized CHAID decision tree. J Cloud Comp 12, 137 (2023). https://doi.org/10.1186/s13677-023-00517-4
42. Zheng, B., & Yao, C. (2023). Automobile Profession Responds to the Development of Automobile Intelligence. 2023 8th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2023, 419–423. https://doi.org/10.1109/ICCCBDA56900.2023.10154651