Predictive Maintenance for Industries Using Flask and Big data

Abstract

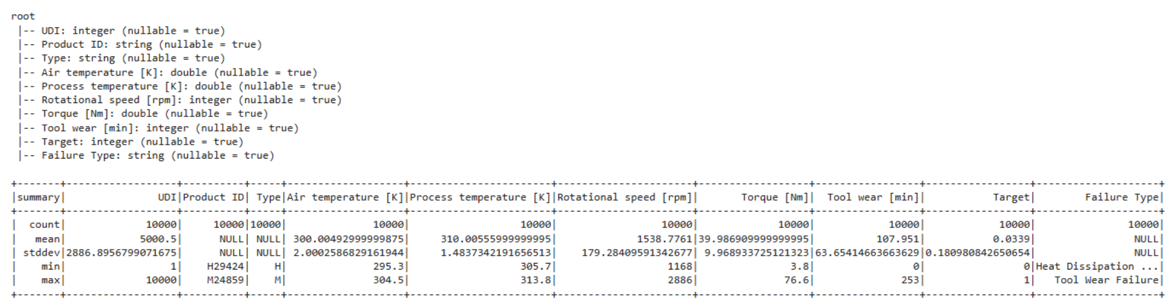
In industries looking to maximize equipment longevity and minimize downtime, predictive maintenance is essential. This study examines the use of machine learning methods, particularly the Gradient Boosting Classifier (GBT), to forecast industrial machinery breakdowns based on sensor data. Other classification models, such Random Forest and Logistic Regression, are compared to see how well the model predicts machine failure. With an accuracy of 97.81%, the data demonstrate that GBT performs better than alternative models. In order to enable real-time failure predictions based on incoming sensor data, the model has been implemented as a component of an API system that uses Flask.

Introduction

Using information from sensors mounted on equipment, predictive maintenance anticipates possible malfunctions before they happen. Predictive maintenance prolongs the life of vital equipment and lowers the costs of unscheduled downtime in sectors including manufacturing, oil and gas, and automotive. Predictive models can identify trends that point to a machine's risk of failure by using past data. In order to provide useful insights into the condition of machinery, this work investigates the application of machine learning models for failure prediction based on sensor data.

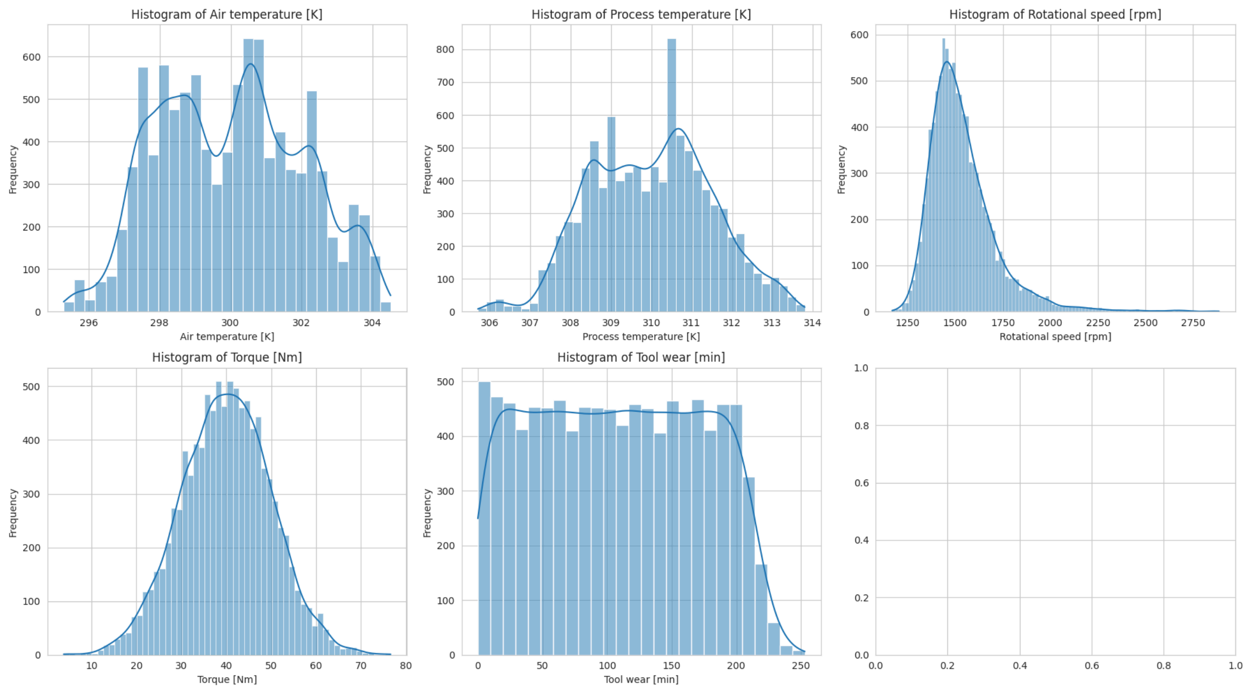
Exploratory Data Analysis:

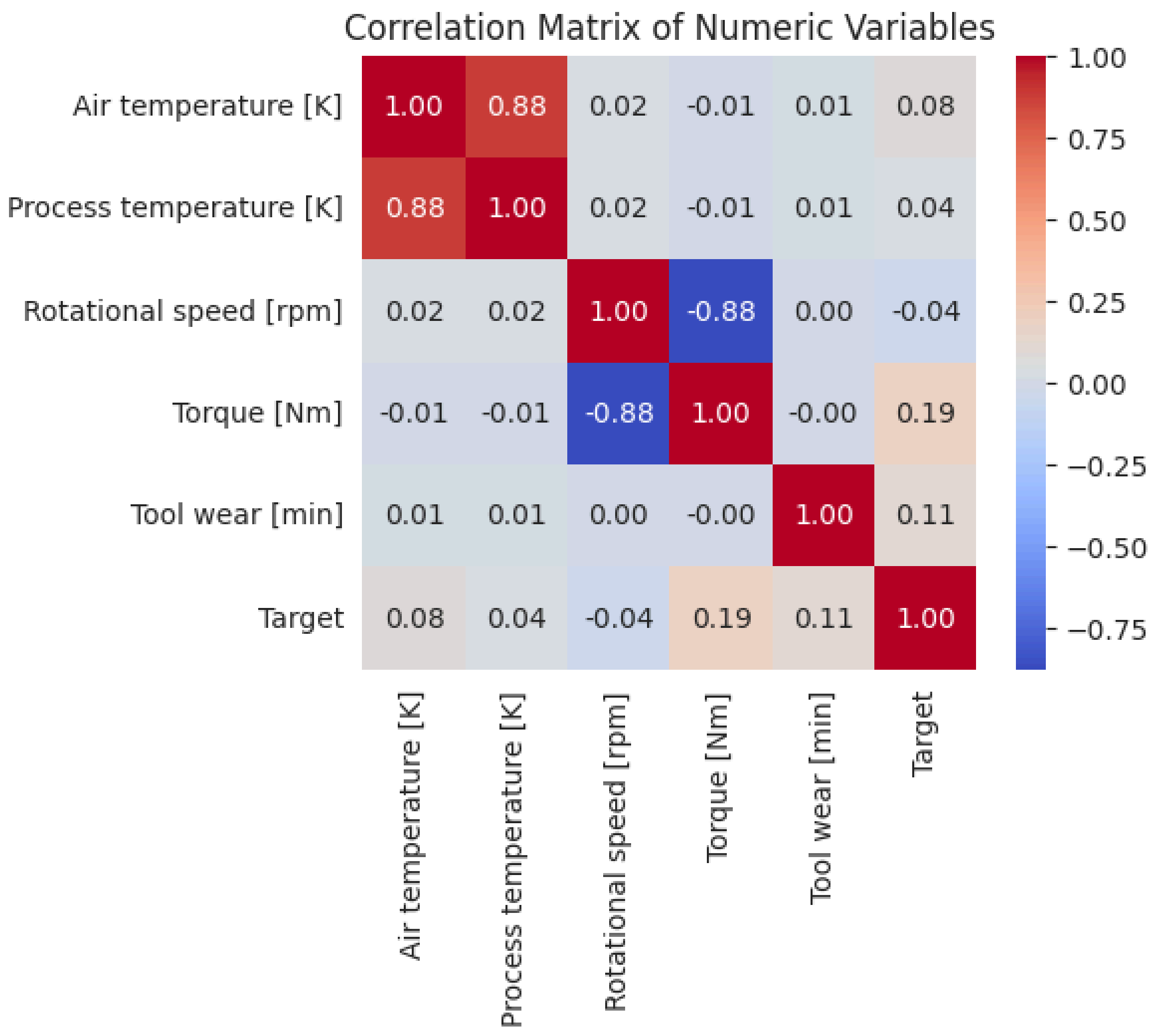
All the columns in our dataset are being displayed.



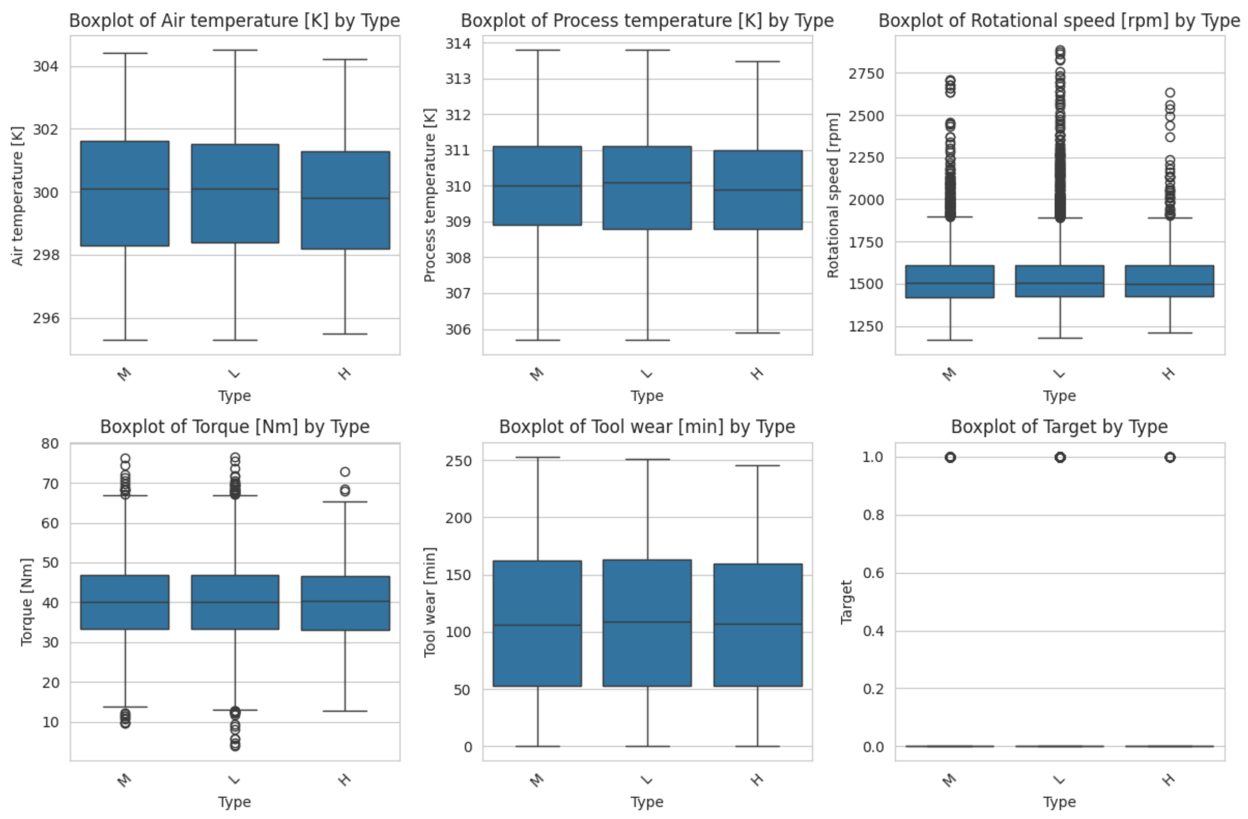
```
root
 |-- UDI: integer (nullable = true)
 |-- Product ID: string (nullable = true)
 |-- Type: string (nullable = true)
 |-- Air temperature [K]: double (nullable = true)
 |-- Process temperature [K]: double (nullable = true)
 |-- Rotational speed [rpm]: integer (nullable = true)
 |-- Torque [Nm]: double (nullable = true)
 |-- Tool wear [min]: integer (nullable = true)
 |-- Target: integer (nullable = true)
 |-- Failure Type: string (nullable = true)

+-------+------------------+----------+-----+-------------------+-----------------------+----------------------+------------------+-----------------+-----------------+--------------------+
|summary|               UDI|Product ID| Type|Air temperature [K]|Process temperature [K]|Rotational speed [rpm]|       Torque [Nm]|  Tool wear [min]|           Target|        Failure Type|
+-------+------------------+----------+-----+-------------------+-----------------------+----------------------+------------------+-----------------+-----------------+--------------------+
|  count|             10000|     10000|10000|              10000|                  10000|                 10000|             10000|            10000|            10000|               10000|
|   mean|            5000.5|      NULL| NULL| 300.00492999999875|     310.00555999999995|             1538.7761|39.986909999999995|          107.951|           0.0339|                NULL|
| stddev|2886.8956799071675|      NULL| NULL| 2.0002586829161944|      1.483734219165651 3|    179.28409591342677| 9.968933725121323|63.65414663663629|0.180980842650654|                NULL|
|    min|                 1|    H29424|    H|              295.3|                  305.7|                  1168|               3.8|                0|                0|Heat Dissipation ...|
|    max|             10000|    M24859|    M|              304.5|                  313.8|                  2886|              76.6|              253|                1|   Tool Wear Failure|
+-------+------------------+----------+-----+-------------------+-----------------------+----------------------+------------------+-----------------+-----------------+--------------------+
```

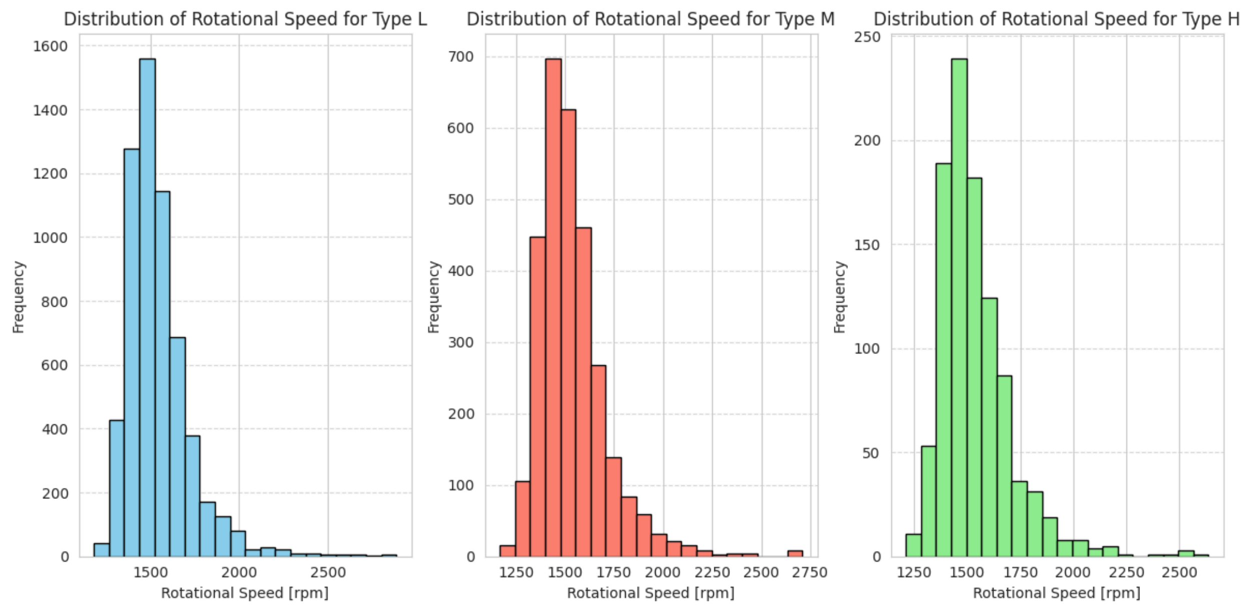

Histogram of Air temperature [K]
Air temperature [K]
Frequency
Histogram of Process temperature [K]
Process temperature [K]
Frequency
Histogram of Rotational speed [rpm]
Rotational speed [rpm]
Frequency
Histogram of Torque [Nm]
Torque [Nm]
Frequency
Histogram of Tool wear [min]
Tool wear [min]
Frequency



	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Target
Air temperature [K]	1.00	0.88	0.02	-0.01	0.01	0.08
Process temperature [K]	0.88	1.00	0.02	-0.01	0.01	0.04
Rotational speed [rpm]	0.02	0.02	1.00	-0.88	0.00	-0.04
Torque [Nm]	-0.01	-0.01	-0.88	1.00	-0.00	0.19
Tool wear [min]	0.01	0.01	0.00	-0.00	1.00	0.11
Target	0.08	0.04	-0.04	0.19	0.11	1.00



Boxplot of Air temperature [K] by Type
Air temperature [K]
Boxplot of Process temperature [K] by Type
Process temperature [K]
Boxplot of Rotational speed [rpm] by Type
Rotational speed [rpm]
Boxplot of Torque [Nm] by Type
Torque [Nm]
Boxplot of Tool wear [min] by Type
Tool wear [min]
Boxplot of Target by Type
Target
M
L
H
Type



Distribution of Rotational Speed for Type L
Distribution of Rotational Speed for Type M
Distribution of Rotational Speed for Type H
Frequency
Rotational Speed [rpm]

Existing Methods
The paper makes use of the

For anomaly detection, the study assesses models such as One-Class SVM, Isolation Forest, Elliptic Envelope, and Local Outlier Factor (LOF).

Classifiers such as Gaussian Naive Bayes, SVM, Random Forests, K-Nearest Neighbors (KNN), Logistic Regression, and Decision Trees are used to forecast failures.

The following are a few popular machine learning techniques for predictive maintenance:

- Random Forest: A group technique that constructs several decision trees and combines their forecasts.
- For binary classification tasks, such as predicting failure or no failure, logistic regression is a straightforward yet powerful technique.
- Gradient Boosting Machines (GBM) are boosting algorithms that optimize speed by building trees one after the other.

Methodology/Algorithm

The following procedures are used in this project to create and assess the predictive maintenance model:

1. Data Collection: Sensor data from machinery was gathered, including torque, rotational speed, air temperature, process temperature, and tool wear.

   Preprocessing Data:

2. addressing outliers and missing data.
3. Feature normalization with StandardScaler.
4. utilizing a mapping for machine kinds (L, M, and H) to encode category information.
5. Choosing a Model:
6. A number of models were assessed, such as Random Forest, Gradient Boosting Classifier (GBT), and Logistic Regression.
7. Train-test splits with an 80:20 ratio were used to train and validate the models.
8. Assessment of the Model:
9. The accuracy, precision, recall, and F1-score measures were used to evaluate each model's performance.
10. With the maximum accuracy of 97.81%, the Gradient Boosting Classifier was chosen as the top-performing model.

Results

Results as per the paper for failure prediction:

- Logistic Regression
  - 46.40% accuracy
  - 48.70% precision, 90.77% recall
  - 63.39% F1 score
- K-Nearest Neighbors (KNN)
  - 84.89% accuracy
  - 89.15% precision, 80.21% recall
  - 84.44% F1 score
- SVM with Linear Kernel
  - 51.12% accuracy
  - 51.12% precision, 100% recall
  - 67.66% F1 score
- SVM with RBF Kernel
  - 86.46% accuracy
  - 95.48% precision, 77.16% recall
  - 85.35% F1 score
- Gaussian Naive Bayes
  - 86.05% accuracy
  - 94.23% precision, 77.46% recall
  - 85.02% F1 score
- Decision Tree (with entropy criterion)
  - Best performer with 87.51% accuracy
  - 88.08% precision, 87.39% recall
  - 87.73% F1 score
- Random Forest (with entropy criterion, 10 estimators)
  - 84.35% accuracy
  - 88.28% precision, 80.00% recall
  - 83.94% F1 score

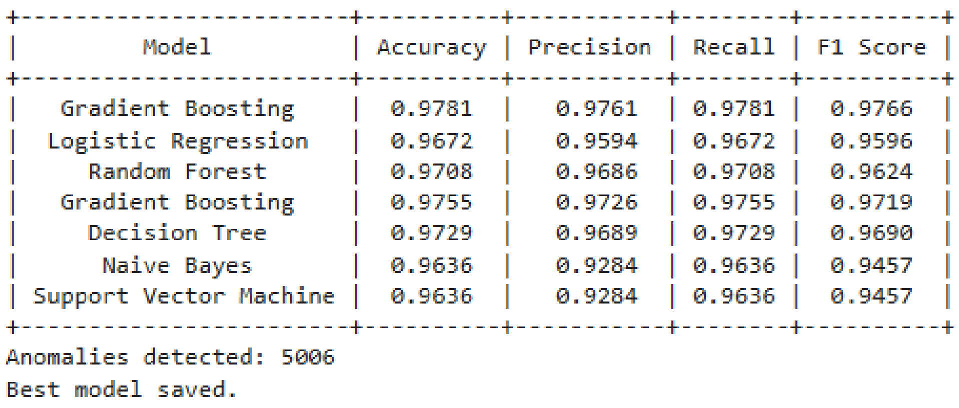
The following is a summary of the models' evaluation on our test set:

Accuracy of the Gradient Boosting Classifier (GBT): 97.81%

Accuracy of logistic regression: 96.72%

Random Forest: accuracy of 97.08%

Revision of Gradient Boosting: accuracy of 97.55%



```
+------------------------+----------+-----------+--------+----------+
|         Model          | Accuracy | Precision | Recall | F1 Score |
+------------------------+----------+-----------+--------+----------+
|   Gradient Boosting    |  0.9781  |   0.9761  | 0.9781 |  0.9766  |
|  Logistic Regression   |  0.9672  |   0.9594  | 0.9672 |  0.9596  |
|     Random Forest      |  0.9708  |   0.9686  | 0.9708 |  0.9624  |
|   Gradient Boosting    |  0.9755  |   0.9726  | 0.9755 |  0.9719  |
|     Decision Tree      |  0.9729  |   0.9689  | 0.9729 |  0.9690  |
|      Naive Bayes       |  0.9636  |   0.9284  | 0.9636 |  0.9457  |
| Support Vector Machine |  0.9636  |   0.9284  | 0.9636 |  0.9457  |
+------------------------+----------+-----------+--------+----------+
Anomalies detected: 5006
Best model saved.
```

When we compare both the results we can observe that our model has achieved best accuracies compared to the models achievement as per the paper.

We also used the user prompt such that user can enter all the details and our model predicts if their is failure or not.

Confusion Matrix and Classification Report for the **GBT model**:

- **Accuracy**: 97.81%
- **Precision: Both classes (failure and no failure) have high precision.**
- **Recall: Both classes showed high recall, demonstrating the model's resilience in forecasting failures.**

Comparison of Models:

- GBT was selected as the final model since it demonstrated the highest accuracy.
- In terms of accuracy, Random Forest and Logistic Regression fared similarly, albeit marginally worse than GBT.

Conclusion

Promising outcomes were shown when machine learning was applied to sectors for predictive maintenance. With an accuracy of 97.81%, the Gradient Boosting Classifier proved to be successful at forecasting equipment failures using sensor data. Predictive maintenance benefits greatly from the model's real-time predictions made possible by its distribution via a Flask-based API. Additional sensor data integration, model parameter adjustment, and the use of sophisticated algorithms could all lead to further advancements.

References

1. N. J. Arasu, K. Gowrisankar, M. Krishnamoorthy, and S. Jangoan, "Predictive Maintenance using Machine Learning in Industrial IoT," *International Journal of Innovative Science and Research Technology*, 2024. doi: 10.38124/ijisrt/ijisrt24mar984.
2. T. Harsh, T. Kumar, A. Mohanty, and A. Pandey, "Predictive Maintenance of Industrial Machines using ML and IoT," presented at the *AI and IoT Conference 2024*. doi: 10.1109/aiiot58432.2024.10574756.
3. D. Patel and P. Kalgutkar, "Predictive Maintenance for Industrial Equipment Using Machine Learning," *International Journal of Advanced Research in Science, Communication and Technology*, 2024. doi: 10.48175/ijarsct-19379.
4. R. K. Mishra, S. Suman, and P. S. Bali, "Predictive Maintenance in Industrial Systems Using Machine Learning," *International Journal of Innovative Science and Research Technology*, 2024. doi: 10.38124/ijisrt/ijisrt24mar1367.