PREDECTIVE MAINTENANCE FOR FACTORY EQUIPMENT

[1] Mrs. M.KAVITHA,[2] A. ROHITH,[3] B. AKSHAY REDDY,[4] J. VINOD,[5] K. SRITHAN,
[1]Assistant Professor, Department of DS, Sri Indu College Of Engineering & Technology
[2,3,4,5]U.G.Scholor, Department of DS, Sri Indu College Of Engineering & Technology, Hyderabad.

Abstract- Predictive maintenance is an important strategy in modern manufacturing that helps industries anticipate and prevent equipment failures, thereby reducing costly downtime and improving operational efficiency. This project proposes a machine learning based approach to predict factory equipment failure using sensor data collected from machines. By analyzing both historical and real time data through supervised learning models such as SVM, Decision Tree, Naïve Bayes, and Logistic Regression, the system can identify early signs of potential malfunction. Among the evaluated models, the Support Vector Machine (SVM) achieved the highest accuracy of 95%, showing its effectiveness in predictive maintenance applications. The proposed solution uses Python libraries along with the Kaggle Predictive Maintenance dataset for data analysis, visualization, and performance evaluation of the developed models.

I. Introduction

In industrial environments, maintaining uninterrupted production is essential to ensure the balance between supply and demand. Equipment failures can lead to unplanned downtime, resulting in significant financial losses and supply chain disruption. Traditional maintenance strategies, like reactive or scheduled maintenance, often prove inefficient as they either wait for breakdowns or apply uniform servicing schedules regardless of machine condition.

With the advent of the Industrial Internet of Things (IIoT), sensor technologies now allow real-time monitoring of machinery health. Sensors embedded in equipment continuously record variables such as temperature, rotation speed, pressure, and vibration. When analyzed with machine learning, this data can help predict the remaining useful life of components and identify signs of degradation early on.

Machine learning models are particularly well-suited for this task due to their ability to learn from historical data and generalize patterns to new inputs. Algorithms like SVM, Decision Tree, Naïve Bayes, and Logistic Regression can classify machine status based on input features, providing insights into potential failures before they occur.

The objective of this project is to develop a predictive maintenance model using the Predictive Maintenance Dataset from Kaggle. The project uses Python and Jupyter Notebook for implementation, with extensive data preprocessing, visualization, feature engineering, and performance evaluation carried out to select the most accurate model for deployment.

II. Literature Survey

Several studies have emphasized the importance of predictive maintenance in enhancing equipment reliability. Lee et al. (2014) presented a framework combining sensors and predictive analytics for early fault detection in smart factories. Their work highlighted the reduction of unplanned downtimes and maintenance costs through data-driven decision-making.

Khelif et al. (2017) explored the use of Support Vector Machines (SVM) and artificial neural networks for machine failure classification. Their results confirmed the robustness of SVM in handling high-dimensional and nonlinear sensor data, making it a preferred choice in industrial scenarios.

Another relevant contribution comes from Zhang et al. (2018), who applied Decision Trees and Random Forest algorithms to predict the failure probability of aircraft engines. They stressed the importance of model interpretability and real-time feedback in mission-critical systems.

A review by Ghosh and Chattopadhyay (2019) analyzed various machine learning methods for predictive maintenance, comparing their performance across industries. Their work supported the use of ensemble techniques and hybrid models to increase prediction reliability and reduce false alarms.

III. Proposed Method

The proposed method utilizes supervised machine learning algorithms to classify machine status based on sensor inputs. The methodology begins with data preprocessing, where missing values are checked, non-numeric values are converted, and features are normalized. The dataset used contains multiple sensor readings and labeled failure types, enabling multi-class classification.

Data visualization techniques, such as histograms, correlation matrices, and box plots, are applied to understand feature relationships and distributions. Feature selection is then performed to retain only the most relevant attributes, reducing model complexity and improving accuracy.

The dataset is split into training (80%) and testing (20%) sets. Several machine learning models including SVM, Decision Tree, Naïve Bayes, and Logistic Regression are trained and evaluated using metrics like accuracy, precision, recall, confusion matrix, and ROC curves.

SVM emerged as the top-performing model with an accuracy of 95%. It was followed by Decision Tree with 90%, and both Naïve Bayes and Logistic Regression with 88% each. The performance of each algorithm is visualized using bar graphs and tabular comparisons to highlight their strengths in handling predictive maintenance classification tasks.

In industrial environment production is the main factor to balance supply demand and to progress industry revenue but sometime due to machinery failure production will stop which effect supply and industry revenue. In olden days we don't have any possible way to predict machinery failure before time till it fully break down but now all industries are using sensors to monitor machinery health and by utilizing this sensor data we can predict machine health or its failure and it's available life and based on life technicians will arrange maintenance. Timely maintenance will make machines to work perfectly and production will continue non-stop.

To predict failure we can employ machine or deep learning algorithms which will get trained on past data and can predict future value by taking current input. This trained models can continuously read input from sensor data and then predict machine health or failure.

To make prediction accurate we have experimented with various machine and deep learning algorithms such as SVM, Decision Tree, Naïve Bayes, and Logistic Regression. Each algorithm performance is evaluated in terms of accuracy, precision, recall, Confusion Matrix, ROC Graph and FCSORE. All algorithms able to achieve accuracy of 90% and SVM manage to get an accuracy of 95%.

To train and test above mention algorithm performance we have utilize 'Predictive Maintenance Sensor Dataset' from KAGGLE repository and this dataset can be downloaded from below URL

https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification

In below screen we are showing dataset details

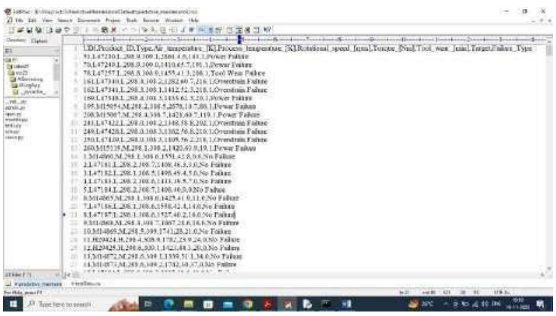


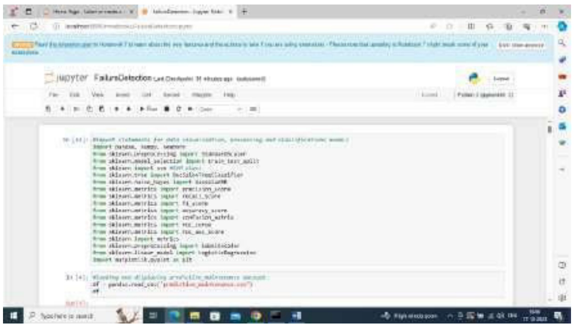
Fig. Dataset screenshot

In above dataset screen first row represents Dataset Column Names and remaining rows contains dataset values and **in last column we have "Failure Type'** as Machine Failure class labels. So by using above dataset we will train and test all algorithm performance.

Before training we have applied various data analysis such as Graph Visualization, Features Selection, data shuffling and normalization.

IV. RESULTS

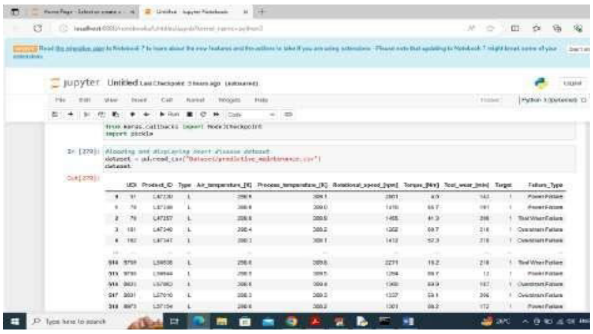
As this project contains data analysis and visualization we preffered Jupyter notebook. We have coded this project using JUPYTER NOTEBOOK and below are the code and output screens with blue color comments



In above screen importing require python classes and packages. For data visualization mainly matplotlib and seaborn are used. While for machine learning prediction sklearn library is used.

Different libraries used are ,

1. Numpy : basic numeric calculations
2. Pandas : loading and analyzing dataset
3. Matplotlib: plotting different graphs or data visualization
4. Seaborn: fro colorful visualizations
5. Sklearn : load machine learning algorithms , load train-test split function , for calculating different metrics.

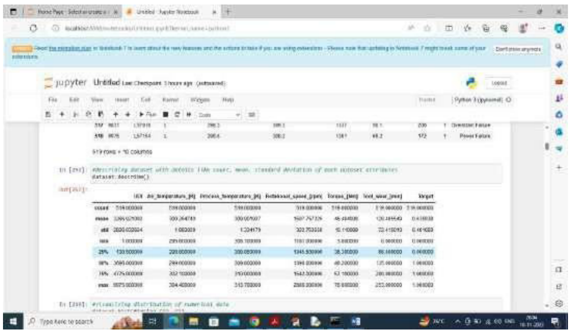


In above screen loading and displaying predictive maintenance dataset

1st row contains the attribute names like UDI, Product ID ,…

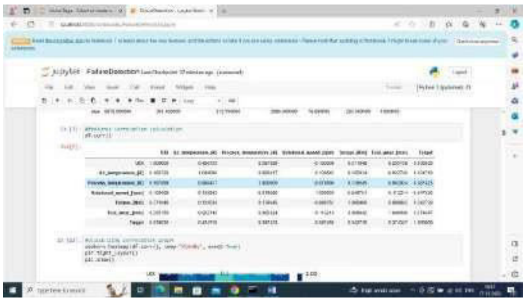
Remaining all rows has data values for those attributes

Last column has failure type like power failure , Tool wear failure , etc



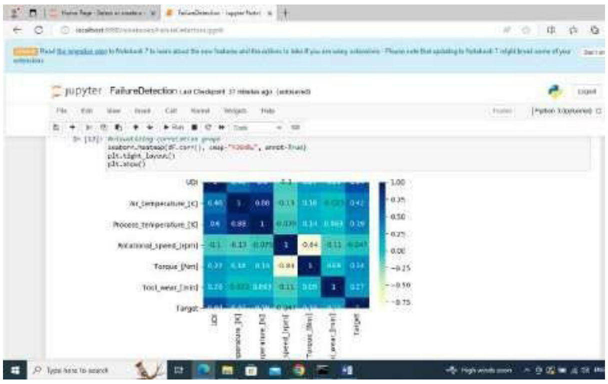
In above screen describing dataset values as 'Mean, standard deviation, min, max and other percentage of values

We used describe() function to get decription of dataset.



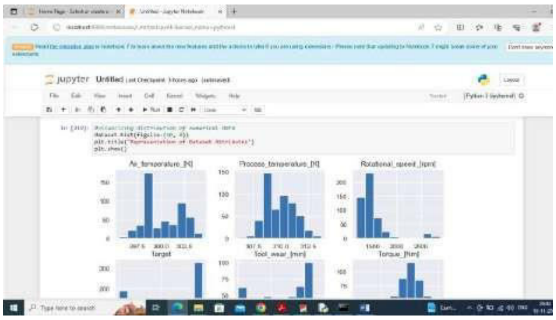
In above screen calculating correlation values for each features in dataset and the high value indicates highly correlated features

By using corr() function we calculated correlation values of each feature.



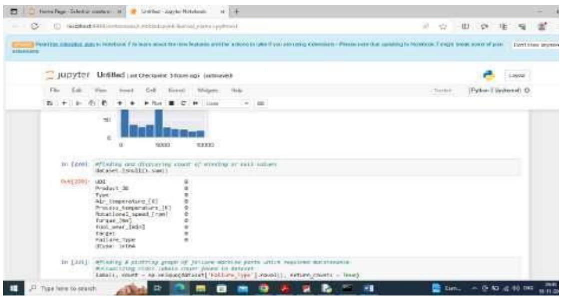
In above screen visualizing correlation graph

The obtained correlation values in previous graph is plotted using seaborn for colorful visualization and easy understanding.



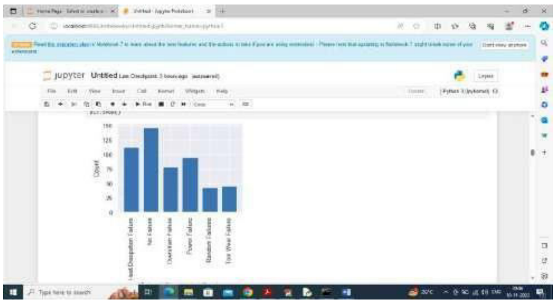
In above screen displaying graph of values distribution for each column values and by seeing above graph we can understand how values of columns distributed from one range to other range

By using .hist() we are plotting histogram of values from dataset. Histogram is the graphical representation of data. On x axis we have range of values and y-axis we have cound(repetation/frequency)



In above screen checking and displaying count of missing values and above dataset contains NO Missing values

We are finding any missing values in dataset so that we can fill it by using zeros or any meadian values but there is no missing value found in the dataset.



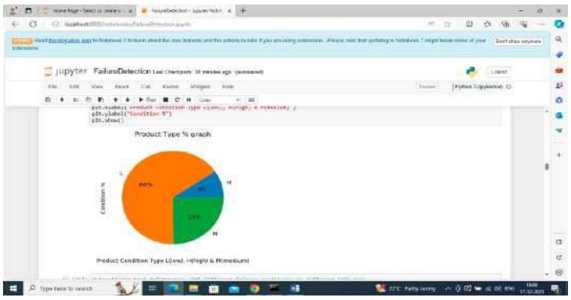
In above graph displaying different Failures found in the dataset where x-axis represents 'Failure Name' and y-axis represents Number of instances or samples found under that failure

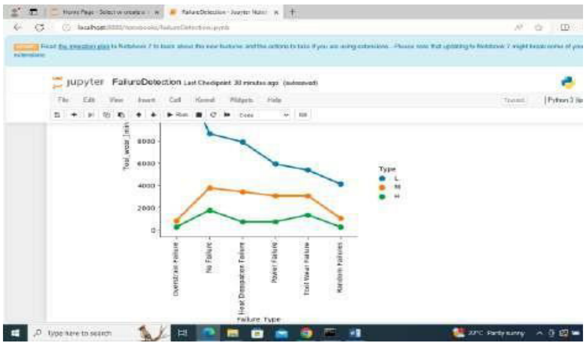
By using unique() function for last column in the dataset, we found the count of each failure type

x-axis : failure type

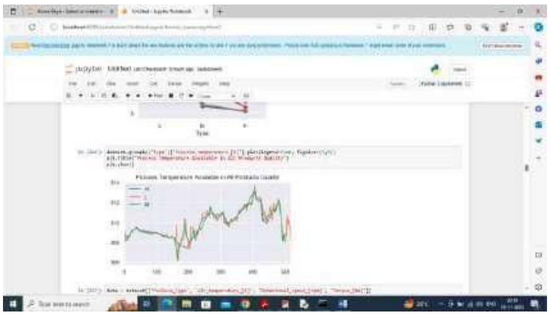
y-axis : count or frequency of that failure

In above graph displaying product maintenance status where L represents Low Quality, M represents Medium and H represents High and in above graph we can see % of product quality in machine

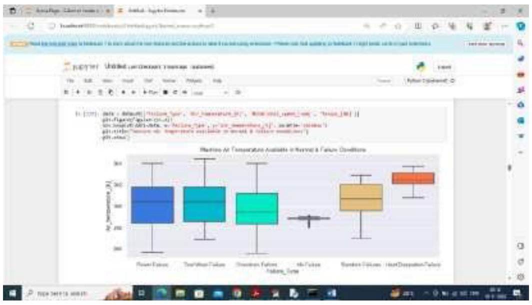




In above graph displaying type of maintenance required under different available life time where x-axis represents type of failure and y-axis represents machine life and each life represents type of maintenance under different available life time.

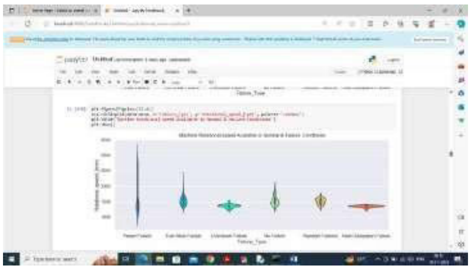


In above graph displaying Machine Process Temperature for various condition of product where x-axis represents Number of Records and y-axis represents Process Temperature and different color lines represents High, Low and Medium product condition



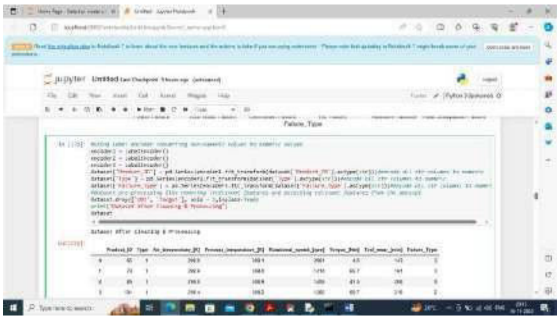
In above graph displaying Air Temperature for different Failure

In above graph plotted box type of plot to know the machine air temperature available in normal and failure conditions



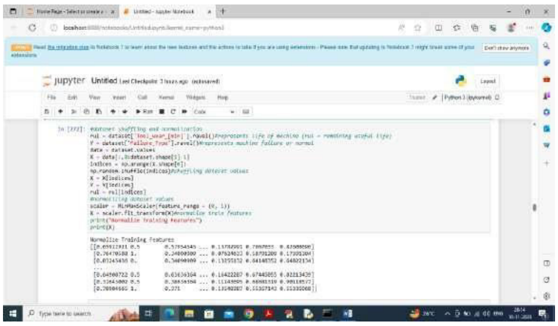
In above graph displaying Machine Rotation Speed for different Failure Condition

Using seaborn plotted violine type of plot.

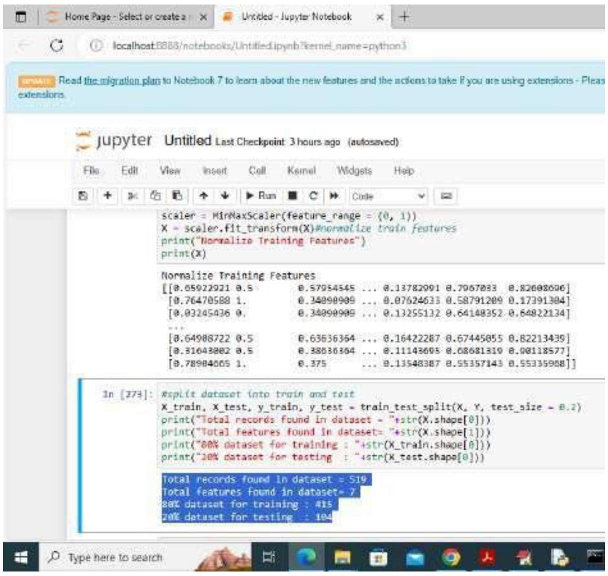


Using above code we are applying data processing to convert non-numeric values to numeric values and after conversion we can see all values are in numeric format as all ML algorithms take input as numeric format so we have converted

In above screen applying various features



processing like Features Selection, shuffling and normalization and after normalization we can see normalized values



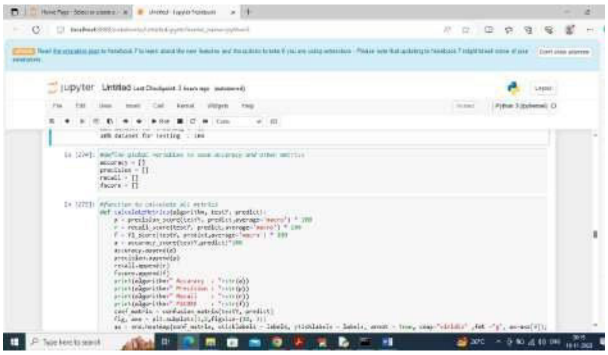
In above screen splitting dataset into train and test where application use 80% dataset size for training and 20% for testing

Xtrain : 80% of the attributes values

Ytrain : labels of 80% attributes (failure type)

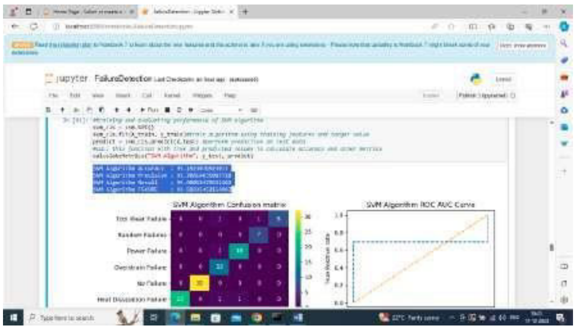
Xtest : 20% test data for testing/ prediction

Ytest : labels of 20% test data (failure type)



In above screen defining function to calculate accuracy and other metrics

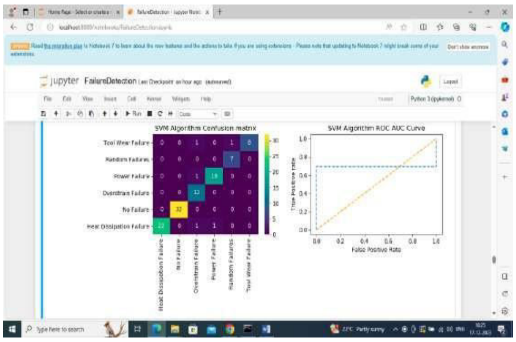
Here we are calculating different metrics by comparing predicted results and actual labels(testY)



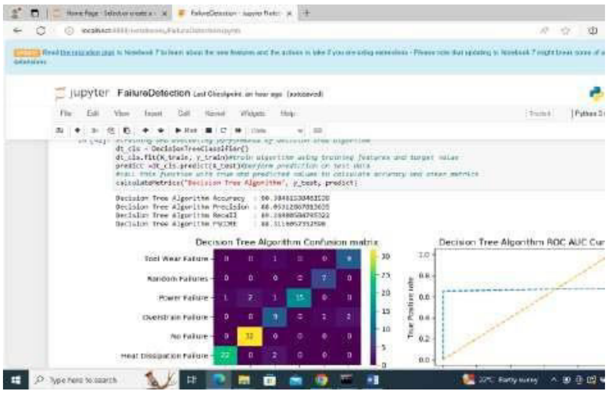
In above screen training SVM algorithm and then performing prediction on test data and after prediction SVM got accuracy as 95% and can see other metrics also and below are the SVM performance graph

SVM : support vector machine

Its is famous supervised machine learning classifier. It works with hyperplane concepts



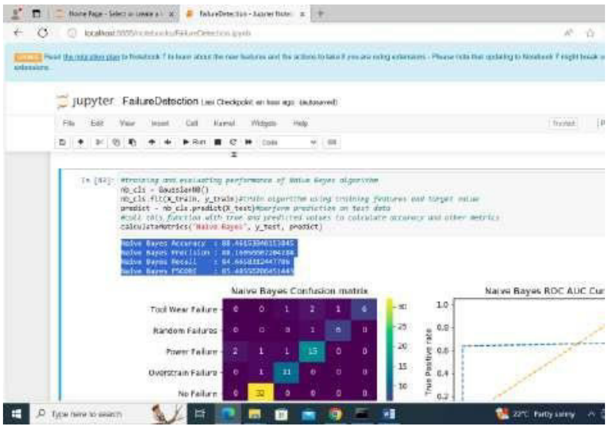
In above confusion matrix graph x-axis represents “Predicted Labels” and y-axis represents “True Labels” and all different color boxes in diagnol represents correct prediction count and remaining all blue boxes contains incorrect prediction count which are very few. In Roc curve graph x-axis represents False Positive Rate and y-axis represents True Positive Rate and if blue line goes below orange line then all predictions are incorrect and if goes above orange line then all predictions are correct and in above ROC graph we can see only few predictions are incorrect



In above screen training Decision Tree algorithm and then it got 90% accuracy and can see other metrics also

Decision Tree classifier

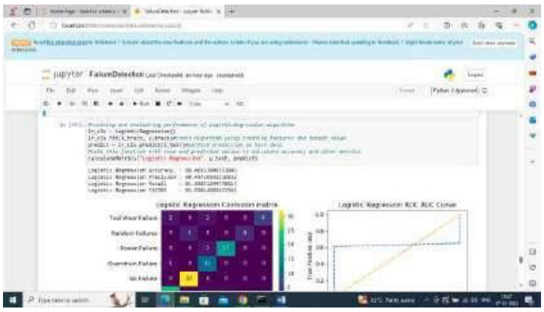
This is supervised machine learning classifier.



In above screen Naïve Bayes got 88% accuracy and can see other metrics also

Naïve Bayes

This is also famous ML classifier. It is also loaded from sklearn



In above screen logistic regression got 88% accuracy

Logistic Regression

This is also famous ML classifier , used for prediction and loaded using sklearn libraries.

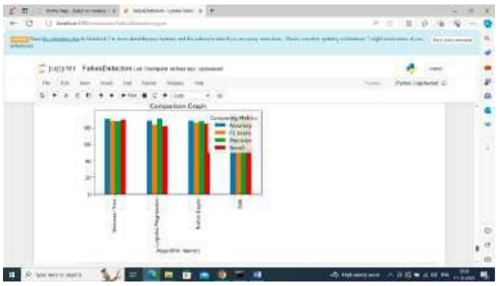


Fig. performance analysis of different ML algorithms

In above graph displaying all algorithm performance where x-axis represents algorithm names and y-axis represents accuracy and other metrics in different color bars and in all algorithms SVM got high performance

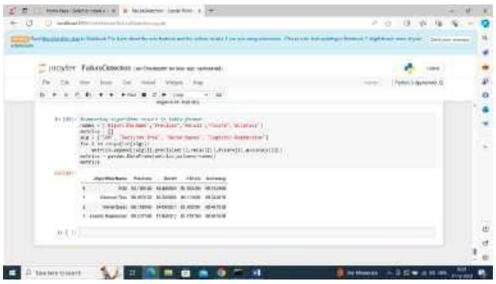


Fig. performace analysis is shown in value format

In above screen displaying all algorithm performance in tabular format and in all algorithm names **SVM got high accuracy**

V. Conclusion

This project demonstrates the effectiveness of machine learning models in predicting factory equipment failure using real-time sensor data. Among the evaluated algorithms, SVM achieved the highest accuracy, making it a strong candidate for deployment in industrial predictive maintenance systems. By leveraging data analytics, industries can proactively manage maintenance schedules, reduce downtimes, and enhance productivity. Future work can focus on incorporating deep learning techniques and real-time deployment in edge computing environments.

References

1. Lee, J., Bagheri, B., & Kao, H. A. (2014). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3, 18–23.
2. Khelif, R., Medjaher, K., & Zerhouni, N. (2017). Data-driven prognostics for predicting remaining useful life based on usage similarity. *Reliability Engineering & System Safety*, 167, 365–379.
3. Zhang, J., Verma, A., & Swamy, M. (2018). Machine learning for aircraft engine predictive maintenance. *IEEE Aerospace Conference*.
4. Ghosh, A., & Chattopadhyay, S. (2019). Predictive maintenance: A comprehensive review. *IEEE Transactions on Instrumentation and Measurement*, 68(12), 4203–4215.
5. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
6. Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4, 77–90.
7. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
8. Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
9. Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2017). Review, analysis and synthesis of predictive maintenance models with machine learning. *Computers in Industry*, 109, 122–140.
10. https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification