Predictive Maintenance for Industrial Equipment

by Guide

-Dr.V.Auxilla Osvin Nancy.,M.Tec

Done by

-MADHAVV N S

-CSE-B

-Rajalakshmi Engineering Colleg

Introduction

1. Industry 4.0 demands uninterrupted machine operations.
2. Unplanned downtime costs $50+ billion annually (ISA).
3. Traditional maintenance (reactive/preventive) is cost-inefficient.
4. Need for data-driven, proactive strategies.
5. Predictive Maintenance (PdM) = monitor + predict failures using data.Uses sensor data, machine learning, and IIoT.

Benefits:

- Reduces downtime and costs
- Enhances safety and efficiency
- Improves spare parts and labor planning

Literature Survey

- Predictive maintenance (PdM) helps forecast equipment failures using sensor data and machine learning.
- Widodo & Yang (2007) used SVM and RVM on vibration signals for early machinery fault prediction.
- Zhang et al. (2019) applied deep learning to time-series sensor data; emphasized the importance of temporal feature extraction.
- Ahmed et al. (2020) found Random Forest and Gradient Boosting outperform traditional classifiers on manufacturing datasets.
- Liu et al. (2021) combined Random Forest with Fast Fourier Transform (FFT) to detect real-time mechanical anomalies.
- Logistic Regression remains popular due to simplicity, interpretability, and successful use in safety-critical industries (Khan et al., 2018).
- SVMs are suitable for high-dimensional sensor data and can define non-linear decision boundaries.
- Gao & Chen (2017) applied SVM to diagnose rotating machinery using vibration and current data features.
- Data augmentation techniques like Gaussian noise, SMOTE, and bootstrapping are used to improve model generalization (Shorten & Khoshgoftaar, 2019).
- Gaussian noise was used in this project to simulate operational variability and improve robustness.
- Common benchmark datasets include NASA's C-MAPSS and Kaggle/UCI Predictive Maintenance datasets.
- These datasets typically contain sensor data such as temperature, torque, pressure, and machine age.

Problem Statement

- Build a predictive maintenance model to forecast equipment failures using sensor data, aiming to reduce downtime and maintenance costs.
- Use the "Predictive Maintenance Dataset" with machine operational settings, sensor measurements, and historical failure events.
- Clean the data, handle missing values, engineer features, and normalize the sensor data for analysis.
- Evaluate machine learning models such as Random Forest Classifier, Logistic Regression, and SVM for optimizing through hyperparameter tuning and cross-validation.
- Extract insights from the model to develop predictive maintenance schedules and strategies to improve equipment reliability and reduce costs.

Dataset Source and Structure

Dataset Source

No of Features	10 columns
No of Records	10,000 rows

Dataset Feature Description

- **UDI :** Unique identifier for each record, ensuring traceability
- **Product ID :** A categorical identifier representing the specific product or machine.
- **Type :** Category of product, which can impact operational characteristics. (M,L and H)
- **Air Temperature [K]** : Ambient air temperature, measured in Kelvin, influencing operational conditions.
- **Process Temperature [K]** : Internal process temperature in Kelvin, key for understanding internal heating dynamics.
- **Rotational Speed [rpm]** : Speed of the equipment's rotation in revolutions per minute, reflecting workload levels.
- **Torque [Nm] :** Torque applied, measured in Newton meters.
- **Tool Wear [min]** : Time (in minutes) of tool usage, indicating wear over time.
- **Target:** Binary indicator if there's a failure (1) or not (0).
- **Failure Type :** Specifies failure category (No Failure, Power Failure, Tool wear Failure, Overstrain Failure, Random Failure and Heat Dissipation Failure)

Dataset and Preprocessing

Dataset includes time-series and tabular data from industrial machines

Key features:

- Vibration levels
- Temperature
- Pressure
- RPM
- Voltage and current
- Operating hours

Preprocessing steps:

- Handle missing values
- Normalize features
- Detect and remove outliers
- Ensure data consistency for ML models

Model Uesd:

- Logistic Regression (LR)
- Random Forest (RF)
- Support Vector Machine (SVM)

Data Acquisition and Cleaning

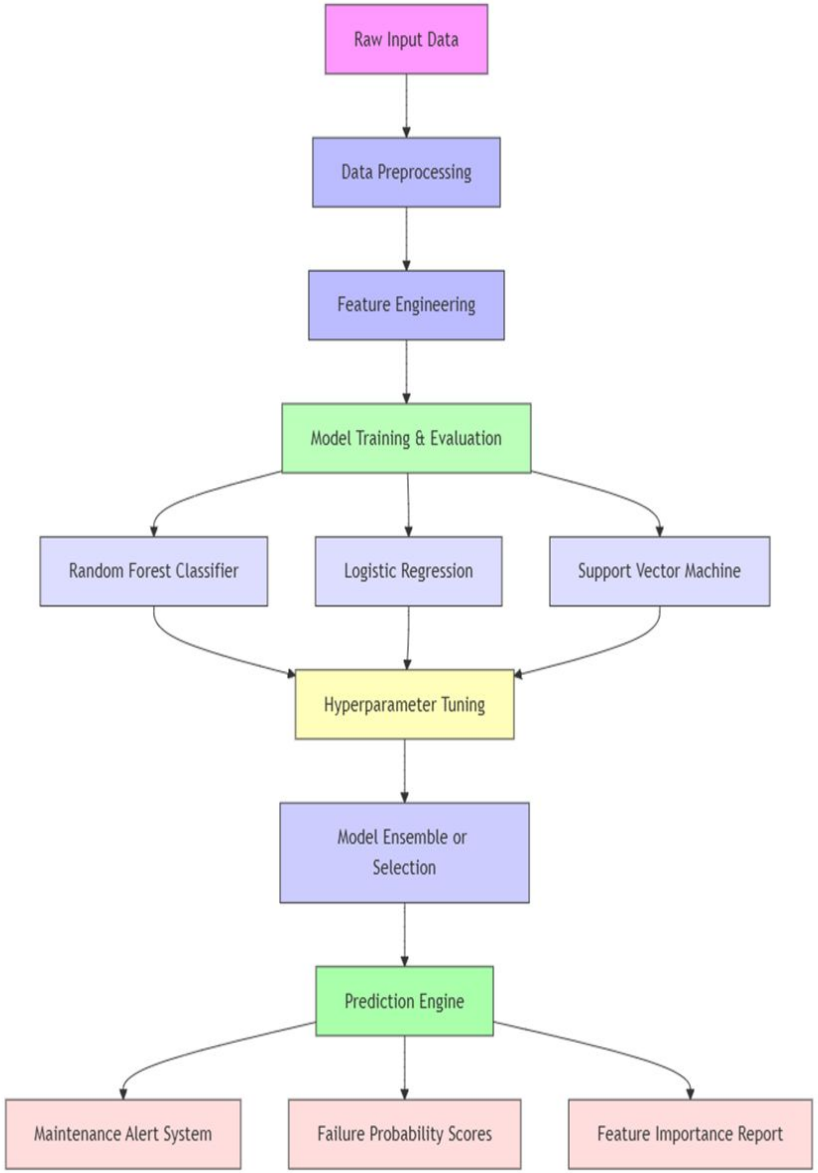
	UDI	Product ID	Type	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Target	Failure Type
0	1	M14860	M	298.1	308.6	1551	42.8	0	0	No Failure
1	2	L47181	L	298.2	308.7	1408	46.3	3	0	No Failure
2	3	L47182	L	298.1	308.5	1498	49.4	5	0	No Failure
3	4	L47183	L	298.2	308.6	1433	39.5	7	0	No Failure
4	5	L47184	L	298.2	308.7	1408	40.0	9	0	No Failure

No missing values detected.

No duplicate rows detected.

- Which captures the difference between "Process temperature [K]" and "Air temperature [K]".
- This transformation highlights the temperature gradient, which could influence the system's behavior or failure rates.
- Which captures the interaction feature, torque_speed_interaction, by multiplying "Torque [Nm]" and "Rotational speed [rpm]".
- This feature represents the combined effect of torque and rotational speed, which could directly correlate with the machine's Stress

System Architecture



Methodology

- Goal: Predict equipment failure using supervised machine learning
- Pipeline structured in 5 phases:
  - Data collection and preprocessing
  - Feature engineering
  - Model selection and training
  - Evaluation
  - Data augmentation
- Objective: High accuracy, real-world usability, and scalability

Implementation

The entire implementation was carried out using **Python programming language** due to its rich ecosystem of data science libraries.

Development and testing were performed on **Google Colab**, providing GPU support, scalability, and cloud-based reproducibility.

The following **Python libraries** were used:

- pandas and numpy for data manipulation
- matplotlib and seaborn for visualization
- scikit-learn for machine learning models and evaluation
- imbalanced-learn for handling class imbalance (SMOTE, etc.)

Step-by-step Implementation Flow:

1. **Data Loading**: Imported CSV dataset into pandas DataFrame
2. **Data Cleaning**:
   - Removed duplicates and null values
   - Handled outliers using interquartile range (IQR) method
3. **Feature Engineering**:
   - Created features like temp_diff and torque_speed_interaction
   - Performed correlation analysis
4. **Data Normalization**:
   - Applied StandardScaler to scale features for SVM and Logistic Regression

Implementation

1. **Train-Test Split**:
   - Split data into 80% training and 20% testing using train_test_split()
2. **Model Training**:
   - Trained Logistic Regression, Random Forest, and SVM
   - Used GridSearchCV for hyperparameter tuning
3. **Model Evaluation**:
   - Calculated MAE, MSE, and R^2 Score
   - Compared model performance visually using bar charts and confusion matrices
4. **Data Augmentation**:
   - Added Gaussian noise to training features
   - Retrained models to improve robustness and generalization

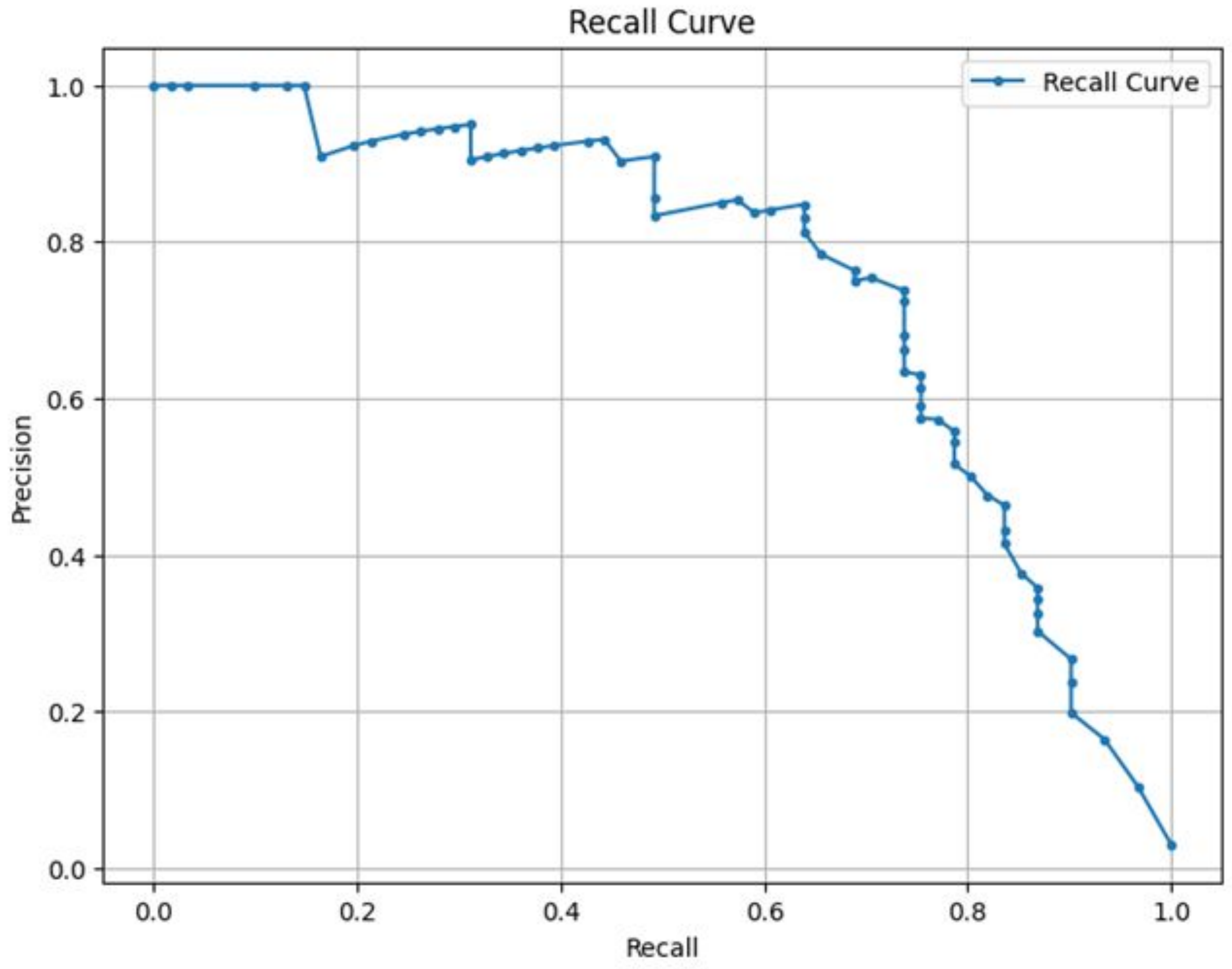
Final models were saved and ready for integration with real-time sensor dashboards or edge computing devices.

The system is designed to be modular and **easily deployable** in industrial monitoring setups.

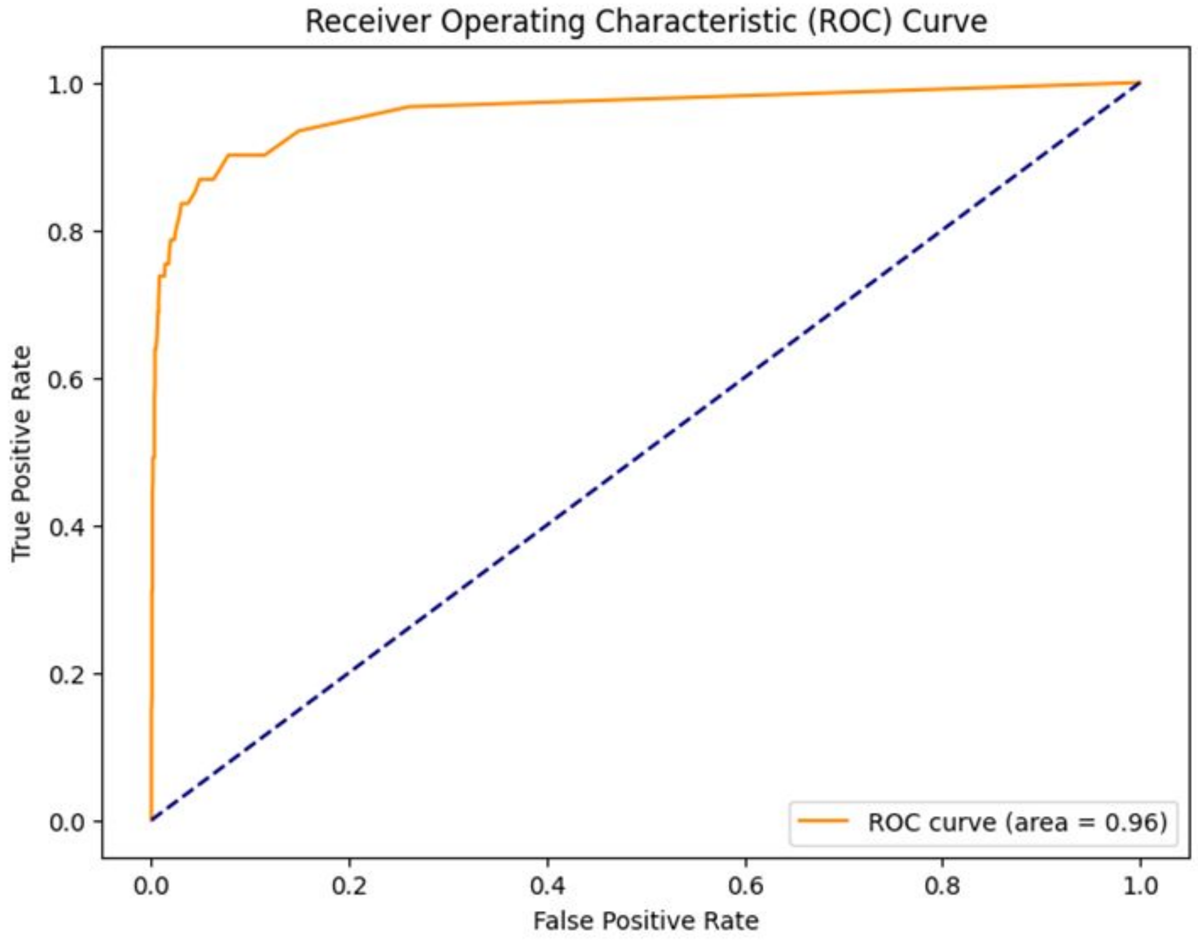
Results

Model	Accuracy (↑ Better)	Precision	Recall	F1 Score	ROC AUC	RANK
Logistic Regression	0.88	0.86	0.83	0.84	0.90	3
Random Forest	0.94	0.93	0.92	0.92	0.96	1
SVM	0.91	0.89	0.88	0.88	0.93	2

Results



Recall Curve
Recall Curve
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0.0
0.2
0.4
0.6
0.8
1.0
Recall



Receiver Operating Characteristic (ROC) Curve
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
ROC curve (area = 0.96)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate

Conclusion and Future Work

- Developed a supervised machine learning system for predictive maintenance using real-time sensor data.
- Implemented three models: Random Forest, Logistic Regression, and SVM, each offering unique strengths in classification performance.
- Identified torque, temperature, and rotational speed as key indicators of equipment failure.
- Feature engineering and Gaussian noise augmentation significantly improved model robustness and generalization.
- Random Forest delivered the best overall performance, while Logistic Regression provided interpretability.
- Extend the model to support real-time prediction using IoT-based sensor feeds.
- Deploy trained models on edge devices for faster inference in on-site environments.
- Incorporate deep learning architectures (e.g., LSTM, GRU) to handle sequential and time-series data more effectively.
- Expand dataset coverage by integrating multiple machine types and failure modes.

Reference

1. Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP, 16*, 3–8.
2. Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal, 13*(3), 2213–2227.
3. Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483–1510.
4. Wen, L., Li, X., Gao, L., & Zhang, Y. (2017). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics, 65*(7), 5990–5998.
5. Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On predictive maintenance in industry 4.0: Overview, models, and challenges. *Applied Sciences, 12*(16), 8081.
6. Grall, A., Dieulle, L., Bérenguer, C., & Roussignol, M. (2002). Continuous-time predictive-maintenance scheduling for a deteriorating system. *IEEE Transactions on Reliability, 51*(2), 141–150.
7. Vlasov, A. I., Grigoriev, P. V., Krivoshein, A. I., Shakhnov, V. A., Filin, S. S., & Migalin, V. S. (2018). Smart management of technologies: predictive maintenance of industrial equipment using wireless sensor networks. *Entrepreneurship and Sustainability Issues, 6*(2), 489–502.
8. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
9. Vijay Kumar, E., Chaturvedi, S. K., & Deshpandé, A. W. (2009). Maintenance of industrial equipment: Degree of certainty with fuzzy modelling using predictive maintenance. *International Journal of Quality & Reliability Management, 26*(2), 196–211.
10. Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. D. P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering, 137*, 106024.
11. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics, 11*(3), 812–820.
12. Wang, K., & Yu, D. (2014). Dimensionality reduction of high-dimensional data using PCA and SVM. *Journal of Computers, 9*(4), 971–979.
13. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2020). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing, 104761*.
14. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.
15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*.

THANK YOU