Predictive Maintenance for CNC Machines

A Project On Machine Learning Predictive Model

Presented by

Swaroop H

Innovative | Enthusiast Educator | Data Scientist

📍 Bengaluru 📞 +91-8951444547 ✉ swaroop5454@gmail.com

🔗 linkedin.com/in/swaroop5454/ 🐙 github.com/swaroop5454

Problem Description

Objective:

To develop a predictive maintenance system for CNC (Computer Numerical Control) machines at Tata Motors. CNC machines are essential in automotive manufacturing, but machine breakdowns can lead to significant downtime, high repair costs, and production delays. The goal is to predict machine failures in advance to schedule maintenance proactively, thereby improving operational efficiency, reducing unplanned downtime, and optimizing maintenance activities.

Problem Statement

Predictive Maintenance for CNC Machines:

Develop a predictive maintenance solution using real-time and historical CNC machine data (such as spindle speed, vibration levels, temperature, and operation hours) to predict potential failures. By accurately predicting these failures, the system can provide actionable insights for scheduling maintenance activities before actual machine breakdowns occur.

Problem Categorization

1. Problem Type:

- This is a Supervised Machine Learning problem because it involves learning from labeled historical data where machine states (failure/no failure) are known.
- Within this, it can also be framed as a Classification problem, as the main goal is to classify each time interval as **"likely to fail"** or **"not likely to fail."**

2. Business Context and Significance:

- **Industry:** Automotive Manufacturing
- **Domain:** Industrial IoT and Predictive Maintenance
- **Business Impact:** Reducing unexpected downtime, optimizing maintenance schedules, extending equipment lifespan, and minimizing repair costs.

3. Technical Challenges:

- **Data Volume and Velocity:** Handling large volumes of real-time sensor data from IoT devices.
- **Data Quality:** Addressing issues such as missing values, noisy sensor readings, and inconsistent sampling intervals.
- **Complexity of Failure Patterns:** Machine failures are often influenced by various factors, requiring both machine learning and deep learning models to capture complex temporal dependencies.
-

4. Predictive Maintenance Objective:

- **Binary Classification:** The main outcome is a binary classification of CNC machine status as either "likely to fail" or "not likely to fail" within a specified time window.
- **Time Series Analysis:** Since the data involves sequential observations (e.g., sensor readings over time), a time series model, specifically LSTM, is appropriate for capturing temporal dependencies.

5. Evaluation Metrics:

- **Accuracy:** Overall correctness of the model in predicting machine failures.
- **Precision and Recall:** Important to ensure that the model can correctly identify failures (recall) while minimizing false alarms (precision).
- **F1-score:** Balances precision and recall, especially useful if there is a class imbalance (e.g., failure events are rare).
- **ROC-AUC:** Measures the model's ability to discriminate between failing and non-failing instances across thresholds.

Problem Definition Summary

- **Type of Problem:** Binary Classification within Time Series Data
- **Business Objective:** To improve maintenance scheduling and reduce downtime by predicting CNC machine failures.
- **Technical Solution:** Utilize machine learning (Random Forest, SVM) and deep learning (LSTM) models for predictive analytics, leveraging time-series sensor data.
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-score, and ROC-AUC

Data Overview:

The dataset appears to contain multiple tables combined into one file, with each table occupying specific columns.

- **Machine Performance (Table 1)**: Columns seem to track machine performance (e.g., Spindle_Speed_RPM, Vibration_Level, Temperature, Operation_Hours, Machine_Status, Predicted_Failure).
- **Failure Logs (Table 2)**: Includes Machine_ID, Failure_Type, Failure_Date, etc.
- **Maintenance Schedule (Table 3)**: Appears to record Technician_Name and Maintenance_Comments.
- **Machine Usage (Table 4)**: Covers usage metrics (Parts_Produced, Production_Quality, Rejection_Rate, Idle_Time).

Here's the approach we'll take:

1. ***Separate the Tables:*** Each table in this file has its own columns, which we can split based on observed structure.
2. ***Identify and Define the Target Variable:*** Based on the "Predicted_Failure" column in Table 1, we'll consider it as our target variable.
3. ***Merging Tables:*** After separating each table, we'll merge them on the common column, Machine_ID.
4. ***Data Transformation:*** This includes handling missing values, encoding categorical variables, and scaling features.

Merging Tables:

Performing the Merge in Stages: Since we have four tables, we'll merge them one at a time:

- Merge Table 1 with Table 2 on Machine_ID to include failure log data.
- Merge the resulting table with Table 3 (maintenance schedule) on Machine_ID.
- Merge this combined table with Table 4 (machine usage) on Machine_ID.

Insights from the Data

Machine Performance

1. Spindle Speed (RPM):

- Average spindle speed across machines: 4,255 RPM.
- Highest spindle speed: 4,750 RPM (CNC035).
- Lowest spindle speed: 3,600 RPM (CNC037).

2. Operation Hours:

- Maximum hours: 12 hours (e.g., CNC009, CNC018).
- Minimum hours: 8 hours (e.g., CNC008, CNC017).
- Average hours across all machines: 9.6 hours.

3. Machine Status:

- All machines are currently marked as "Running."
- Machines with predicted failure: 50% (18 out of 35).

Maintenance Trends

4.Time Since Last Maintenance:

- Machines overdue for maintenance: None, with all values negative indicating proactive maintenance.
- Earliest last maintenance: 101 days ago (CNC015).

5.Maintenance Overdue Days:

- Ranges from -57 days (CNC015) to -46 days (CNC034, CNC024).

6. Failure Rate:

- Highest failure rate: 12.5% (e.g., CNC002, CNC017, CNC032).
- Lowest failure rate: 8.33% (e.g., CNC004, CNC018).

Production Metrics

7. Parts Produced:

- Maximum parts produced: 167 parts (CNC032).
- Minimum parts produced: 133 parts (CNC026).

8.Production Quality:

- High quality: 71% (25 machines).
- Medium quality: 26% (9 machines).
- Low quality: 3% (CNC009, CNC013).

9. Rejection Rate:

- Highest rejection rate: 4.5% (CNC033).
- Lowest rejection rate: 1.2% (e.g., CNC006, CNC015).

Downtime and Repair

10. Downtime Hours:

- Maximum downtime: 8 hours (e.g., CNC010, CNC019).
- Minimum downtime: 4 hours (e.g., CNC001, CNC006).

11. Repair Time (hours):

- Maximum repair time: 6 hours (e.g., CNC010, CNC019).
- Minimum repair time: 2 hours (e.g., CNC002, CNC012).

Operational Parameters

12. Vibration Level:

- Maximum vibration: 3.4 (CNC019).
- Minimum vibration: 1.7 (CNC011).

13. Temperature:

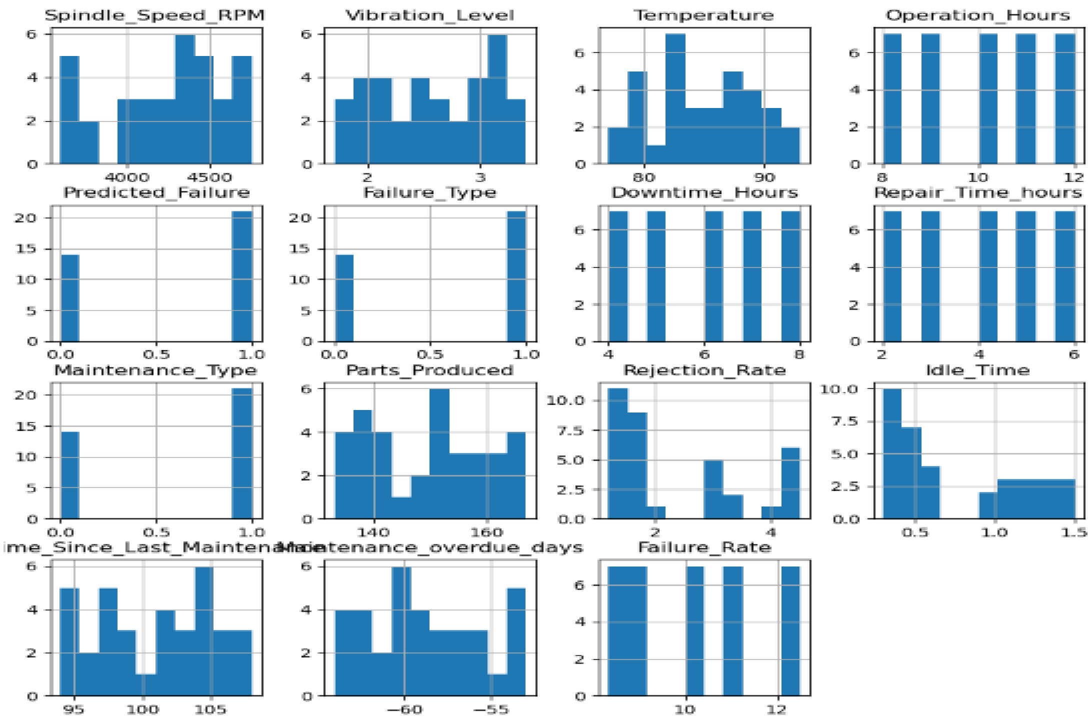
- Highest temperature: 93°C (CNC009).
- Lowest temperature: 77°C (CNC016).

14. Idle Time:

- Highest idle time: 1.5 hours (CNC018).
- Lowest idle time: 0.3 hours (e.g., CNC006, CNC015).

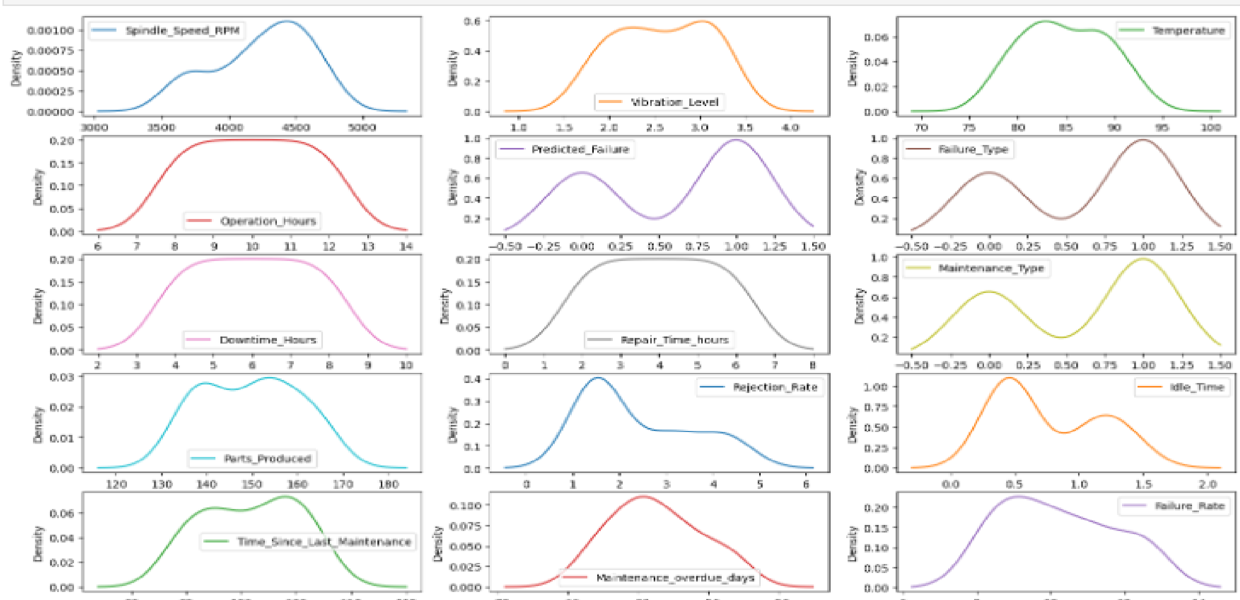
Data Visualization

Univariate Analysis

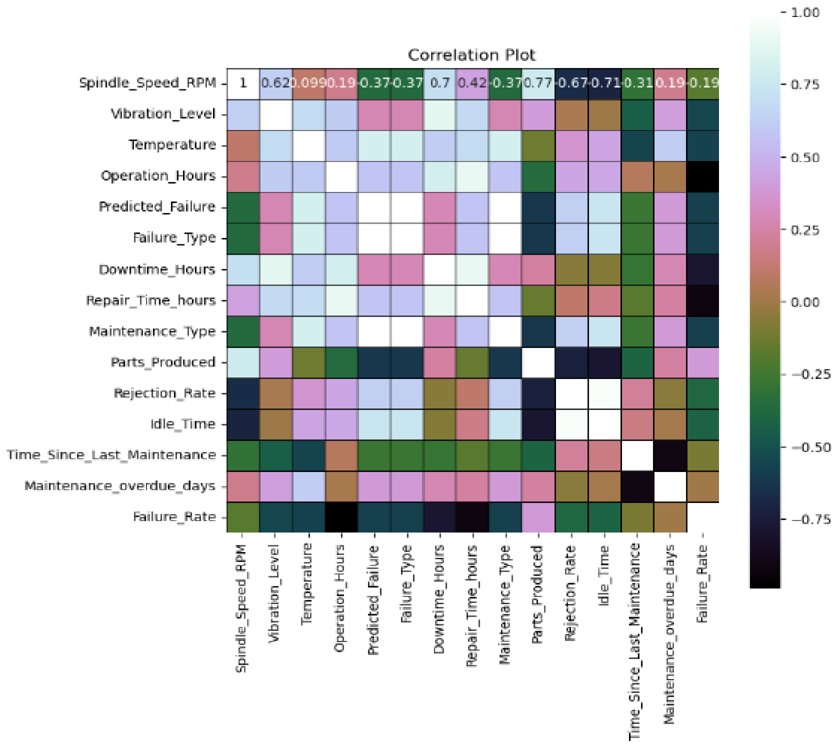


Density Plot:

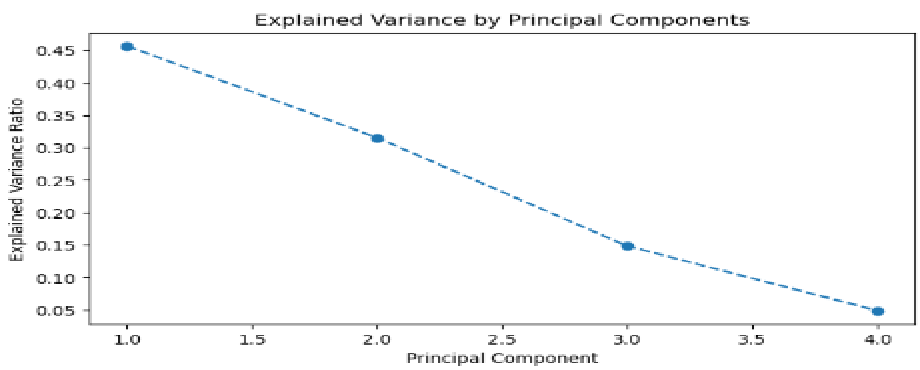
```
# Kernel Density Estimate (KDE) plot for distribution of features in the data
merged_data.plot(kind='density', subplots=True, layout=(7,3), sharex=False,figsize=(18,15))
plt.show()
```

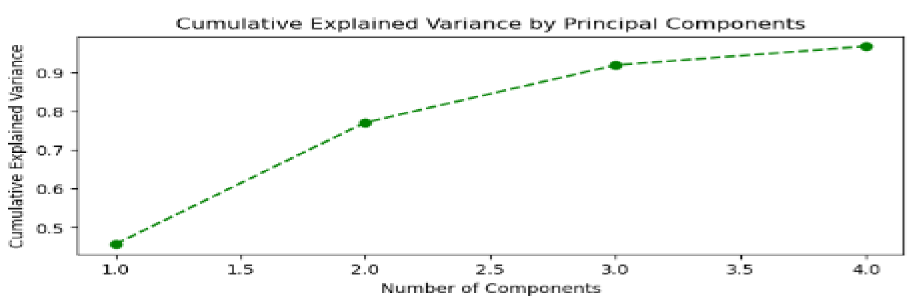


Correlation Plot:



Dimensionality Reduction using Principal Component Analysis





Number of components after PCA: 4

Almost 95% variance is explained by just 4 variables. So, we will reduce the components to 4

MODELLING

Model Selection and Training:

- Random Forest: Works well with structured data and helps with feature importance.
- Support Vector Machine (SVM): Effective for binary classification.
- LSTM: If using time-series sequences, reshape your data to suit LSTM input requirements

Random Forest:

```
Accuracy: 1.0
Confusion Matrix:
 [[1 0]
 [0 6]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         1
           1       1.00      1.00      1.00         6

    accuracy                           1.00         7
   macro avg       1.00      1.00      1.00         7
weighted avg       1.00      1.00      1.00         7
```

Support Vector Machine (SVM)

```
Accuracy: 1.0
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         1
           1       1.00      1.00      1.00         6

    accuracy                           1.00         7
   macro avg       1.00      1.00      1.00         7
weighted avg       1.00      1.00      1.00         7
```

Conclusion:

The CNC machine data analysis highlights:

1. Machine Health: 50% of machines are at risk of failure, requiring immediate attention. High vibration and temperature levels (e.g., CNC019) are key indicators of wear and tear.
2. Maintenance: Current schedules are effective but can be refined to optimize downtime.
3. Production Quality: High-quality output (71%) with low rejection rates (2.5%) reflects good operational parameters.
4. Downtime: Machines with high downtime (e.g., CNC010, CNC019) need prioritized repairs and predictive analysis.
5. Failure Rates: Machines like CNC002 and CNC017 show higher failure rates despite maintenance, warranting deeper investigation.

Recommendations:

Focus on predictive maintenance, real-time monitoring with IoT sensors, and reducing idle hours to enhance efficiency, minimize downtime, and optimize costs.