Introduction

Predicting machine failures is crucial for maintenance and reliability engineering. By spotting potential issues before they arise, companies can cut down on downtime, save money, and boost operational efficiency. In this Notebook, we'll dive into the essential concepts, techniques, and best practices for effective machine failure prediction.

Why Is Machine Failure Prediction Important?

1. **Cost Savings:** Unplanned equipment failures can lead to costly repairs, production delays, and lost revenue. Predictive maintenance helps prevent these issues by allowing timely interventions.
2. **Safety:** Machine failures can pose safety risks to operators and other personnel. Predicting failures in advance enables proactive measures to mitigate these risks.
3. **Optimized Maintenance:** Rather than relying on fixed schedules (which may be inefficient), predictive maintenance focuses on specific equipment conditions. This targeted approach optimizes maintenance efforts.

Load Required Libraries

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import numpy as np
from google.colab import drive
from sklearn.preprocessing import StandardScaler
# data splitting
from sklearn.model_selection import train_test_split
# data modeling
from sklearn.metrics import confusion_matrix,accuracy_score,roc_curve,classification_report
from sklearn.linear_model import LogisticRegression
```

Load Dataset

```
df = pd.read_csv('/content/machine failure.csv')
```

```
df.head()
```

	UDI	Product ID	Type	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	TWF	HDF	PWF	OSF	RNF
0	1	M14860	M	298.1	308.6	1551	42.8	0	0	0	0	0	0	0
1	2	L47181	L	298.2	308.7	1408	46.3	3	0	0	0	0	0	0
2	3	L47182	L	298.1	308.5	1498	49.4	5	0	0	0	0	0	0
3	4	L47183	L	298.2	308.6	1433	39.5	7	0	0	0	0	0	0
4	5	L47184	L	298.2	308.7	1408	40.0	9	0	0	0	0	0	0

Next steps: Generate code with df View recommended plots

```
df.tail()
```

	UDI	Product ID	Type	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	TWF	HDF	PWF	OSF	RNF
9995	9996	M24855	M	298.8	308.4	1604	29.5	14	0	0	0	0	0	0
9996	9997	H39410	H	298.9	308.4	1632	31.8	17	0	0	0	0	0	0
9997	9998	M24857	M	299.0	308.6	1645	33.4	22	0	0	0	0	0	0
9998	9999	H39412	H	299.0	308.7	1408	48.5	25	0	0	0	0	0	0
9999	10000	M24859	M	299.0	308.7	1500	40.2	30	0	0	0	0	0	0

```
df = df.drop('UDI', axis=1)
```

```
df.isnull().sum()
```

```
Product ID                 0
Type                       0
Air temperature [K]        0
Process temperature [K]    0
Rotational speed [rpm]     0
Torque [Nm]                0
Tool wear [min]            0
Machine failure            0
TWF                        0
HDF                        0
PWF                        0
OSF                        0
```

```
RNF                    0
dtype: int64
```

```
df.dropna()
```

	Product ID	Type	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	TWF	HDF	PWF	OSF	RNF
0	M14860	M	298.1	308.6	1551	42.8	0	0	0	0	0	0	0
1	L47181	L	298.2	308.7	1408	46.3	3	0	0	0	0	0	0
2	L47182	L	298.1	308.5	1498	49.4	5	0	0	0	0	0	0
3	L47183	L	298.2	308.6	1433	39.5	7	0	0	0	0	0	0
4	L47184	L	298.2	308.7	1408	40.0	9	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
9995	M24855	M	298.8	308.4	1604	29.5	14	0	0	0	0	0	0
9996	H39410	H	298.9	308.4	1632	31.8	17	0	0	0	0	0	0
9997	M24857	M	299.0	308.6	1645	33.4	22	0	0	0	0	0	0
9998	H39412	H	299.0	308.7	1408	48.5	25	0	0	0	0	0	0
9999	M24859	M	299.0	308.7	1500	40.2	30	0	0	0	0	0	0

10000 rows × 13 columns

```
df.shape
```

```
(10000, 13)
```

```
df.describe()
```

	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	TWF	HDF	PWF	OSF	RNF
count	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.00000
mean	300.004930	310.005560	1538.776100	39.986910	107.951000	0.033900	0.004600	0.011500	0.009500	0.009800	0.00190
std	2.000259	1.483734	179.284096	9.968934	63.654147	0.180981	0.067671	0.106625	0.097009	0.098514	0.04355
min	295.300000	305.700000	1168.000000	3.800000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
25%	298.300000	308.800000	1423.000000	33.200000	53.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
50%	300.100000	310.100000	1503.000000	40.100000	108.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
75%	301.500000	311.100000	1612.000000	46.800000	162.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
max	304.500000	313.800000	2886.000000	76.600000	253.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.00000

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 13 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Product ID               10000 non-null  object 
 1   Type                     10000 non-null  object 
 2   Air temperature [K]      10000 non-null  float64
 3   Process temperature [K]  10000 non-null  float64
 4   Rotational speed [rpm]   10000 non-null  int64  
 5   Torque [Nm]              10000 non-null  float64
 6   Tool wear [min]          10000 non-null  int64  
 7   Machine failure          10000 non-null  int64  
 8   TWF                      10000 non-null  int64  
 9   HDF                      10000 non-null  int64  
 10  PWF                      10000 non-null  int64  
 11  OSF                      10000 non-null  int64  
 12  RNF                      10000 non-null  int64  
dtypes: float64(3), int64(8), object(2)
memory usage: 1015.8+ KB
```

```
df.dtypes
```

```
Product ID                  object
Type                        object
Air temperature [K]        float64
Process temperature [K]    float64
Rotational speed [rpm]       int64
Torque [Nm]                float64
Tool wear [min]              int64
Machine failure              int64
```

```
TWF                          int64
HDF                          int64
PWF                          int64
OSF                          int64
RNF                          int64
dtype: object
```

Checking for the unique values in Target column

```
df['Machine failure'].nunique()
```

```
2
```

```
df.nunique()
```

```
Product ID                 10000
Type                           3
Air temperature [K]           93
Process temperature [K]       82
Rotational speed [rpm]       941
Torque [Nm]                  577
Tool wear [min]              246
Machine failure                2
TWF                            2
HDF                            2
PWF                            2
OSF                            2
RNF                            2
dtype: int64
```

```
features = df.columns.tolist()
features
```

```
['Product ID',
 'Type',
 'Air temperature [K]',
 'Process temperature [K]',
 'Rotational speed [rpm]',
 'Torque [Nm]',
 'Tool wear [min]',
 'Machine failure',
 'TWF',
 'HDF',
 'PWF',
 'OSF',
 'RNF']
```

```
# Categorical features
cat_features = ['Product ID', 'Type']
```

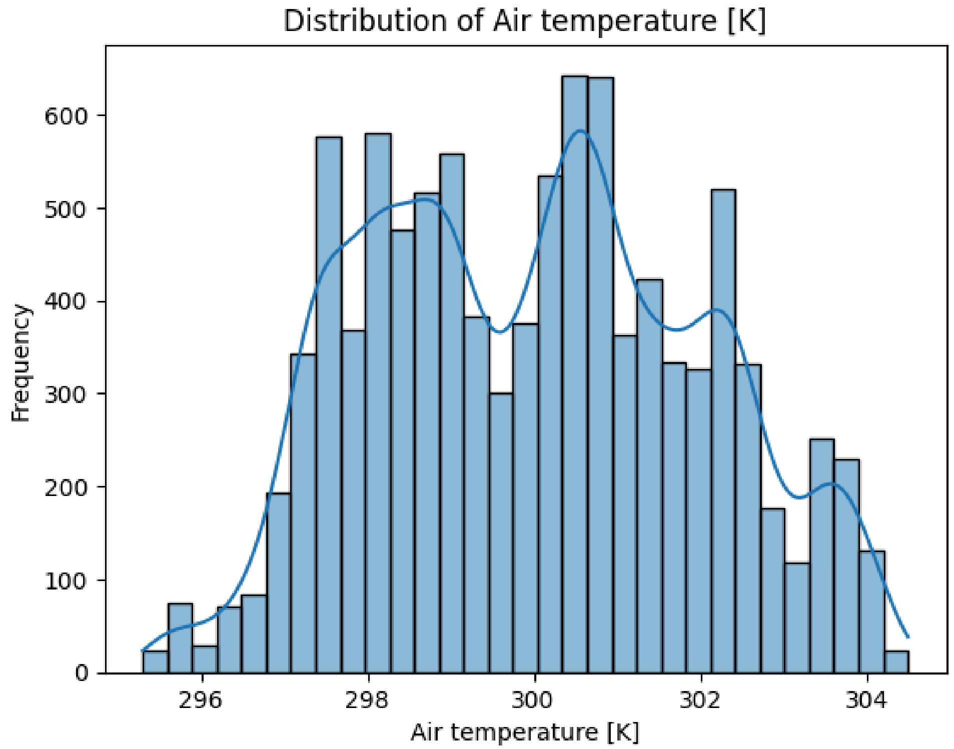
```
df=df.drop(["Product ID"], axis=1)
```

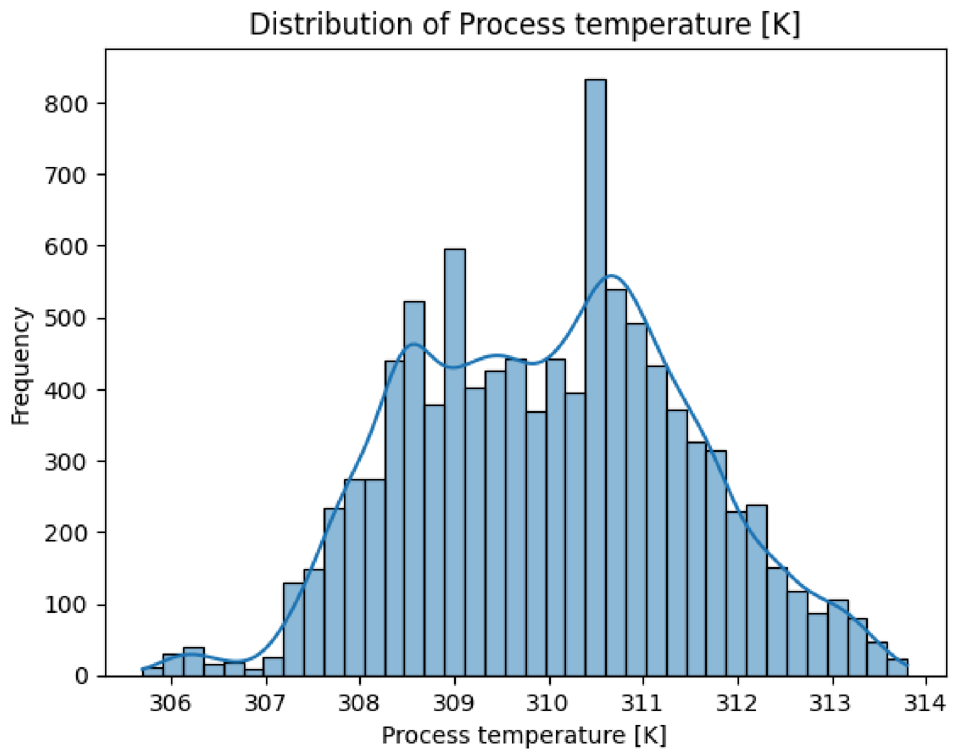
```
# Numerical features
num_features = [f for f in features if f not in (cat_features)]
num_features
```

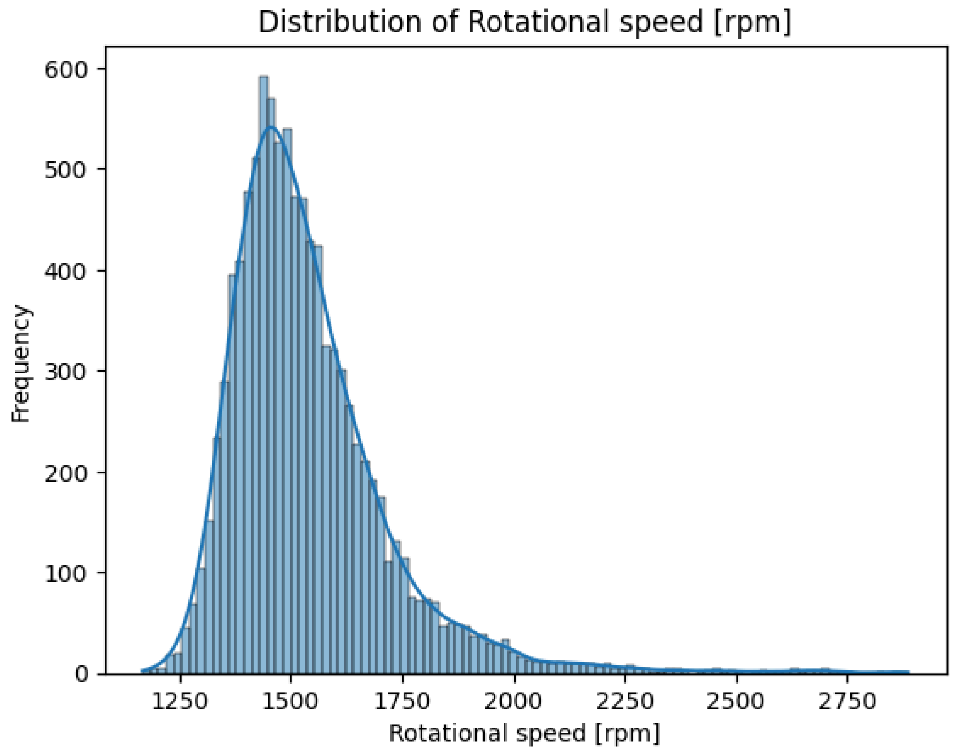
```
['Air temperature [K]',
 'Process temperature [K]',
 'Rotational speed [rpm]',
 'Torque [Nm]',
 'Tool wear [min]',
 'Machine failure',
 'TWF',
 'HDF',
 'PWF',
 'OSF',
 'RNF']
```

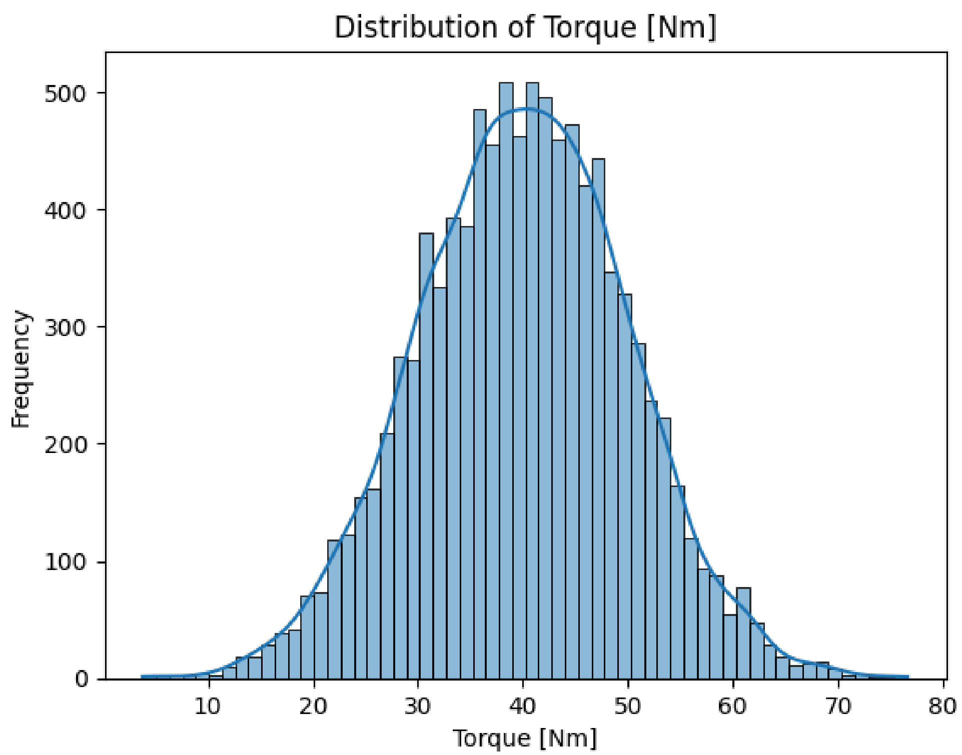
Data Visualization

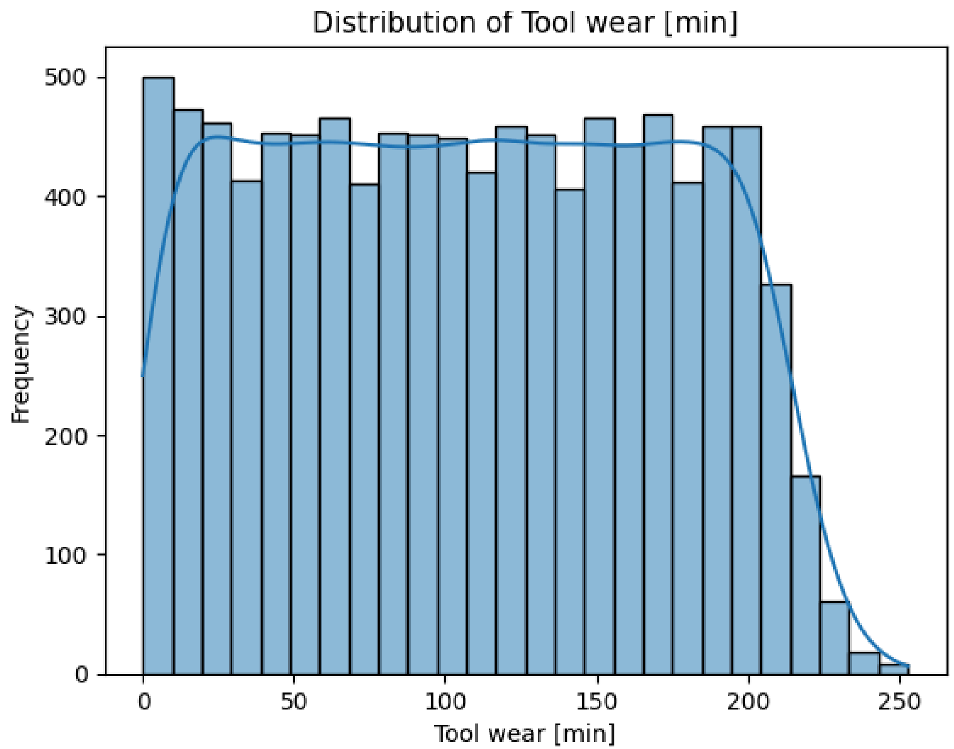
```
for feature in num_features:

    sns.histplot(df[feature], kde=True)
    plt.title(f'Distribution of {feature}')
    plt.xlabel(feature)
    plt.ylabel('Frequency')
    plt.show()
```

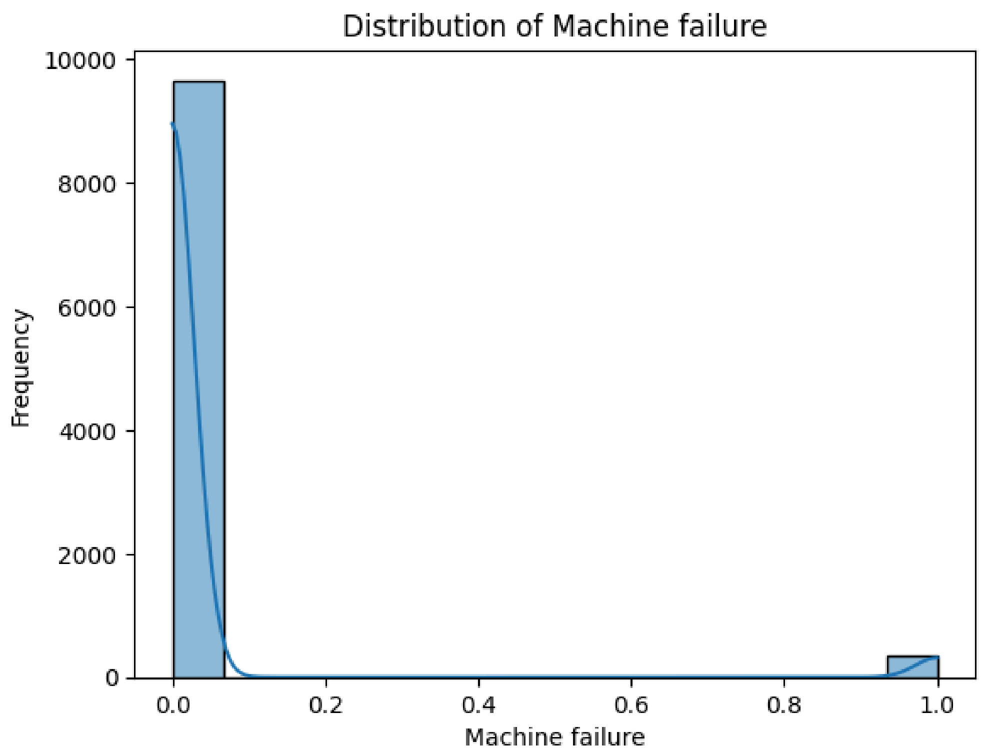

Distribution of Air temperature [K]
Frequency
0
100
200
300
400
500
600
296
298
300
302
304
Air temperature [K]


Distribution of Process temperature [K]
Frequency
0
100
200
300
400
500
600
700
800
306
307
308
309
310
311
312
313
314
Process temperature [K]


Distribution of Rotational speed [rpm]
Frequency
Rotational speed [rpm]
0
100
200
300
400
500
600
1250
1500
1750
2000
2250
2500
2750

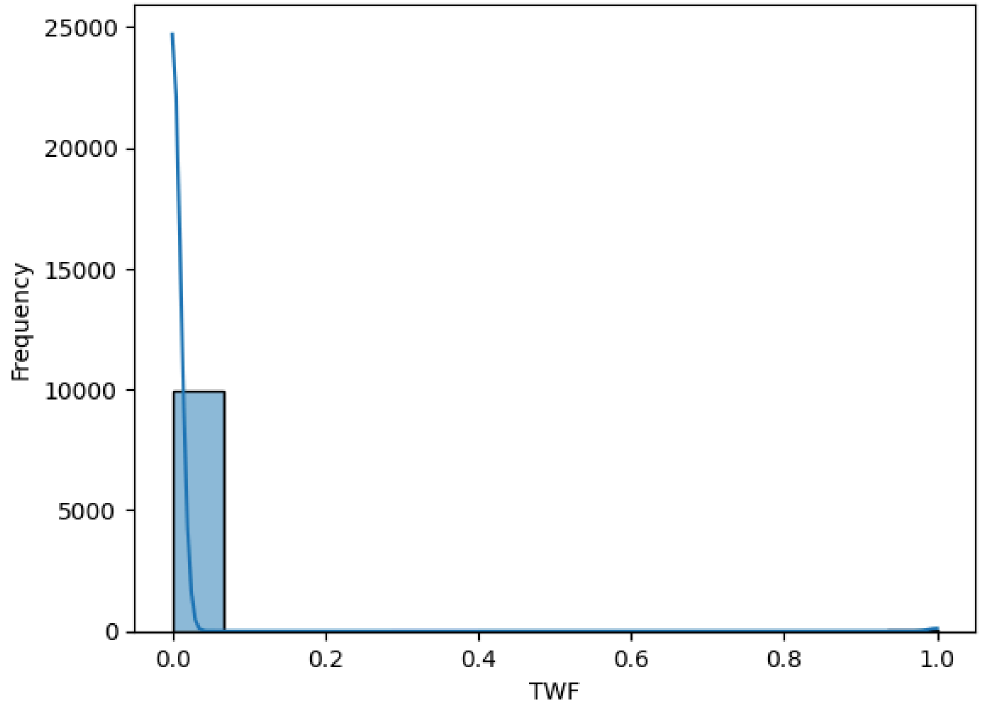

Distribution of Torque [Nm]
Frequency
Torque [Nm]
0
100
200
300
400
500
10
20
30
40
50
60
70
80



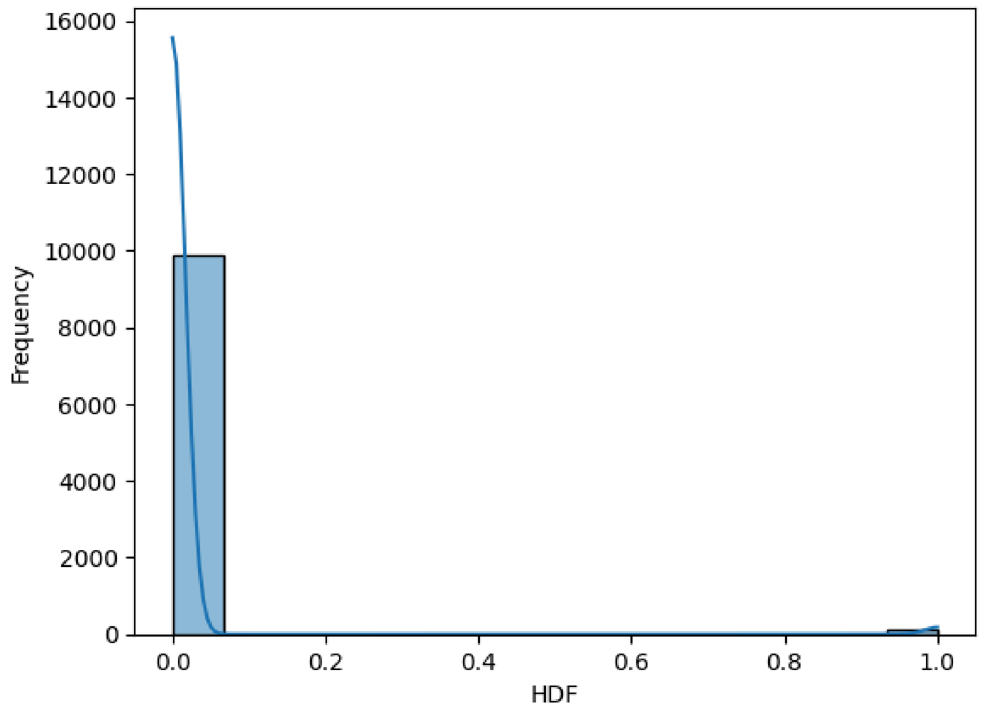


Distribution of TWF

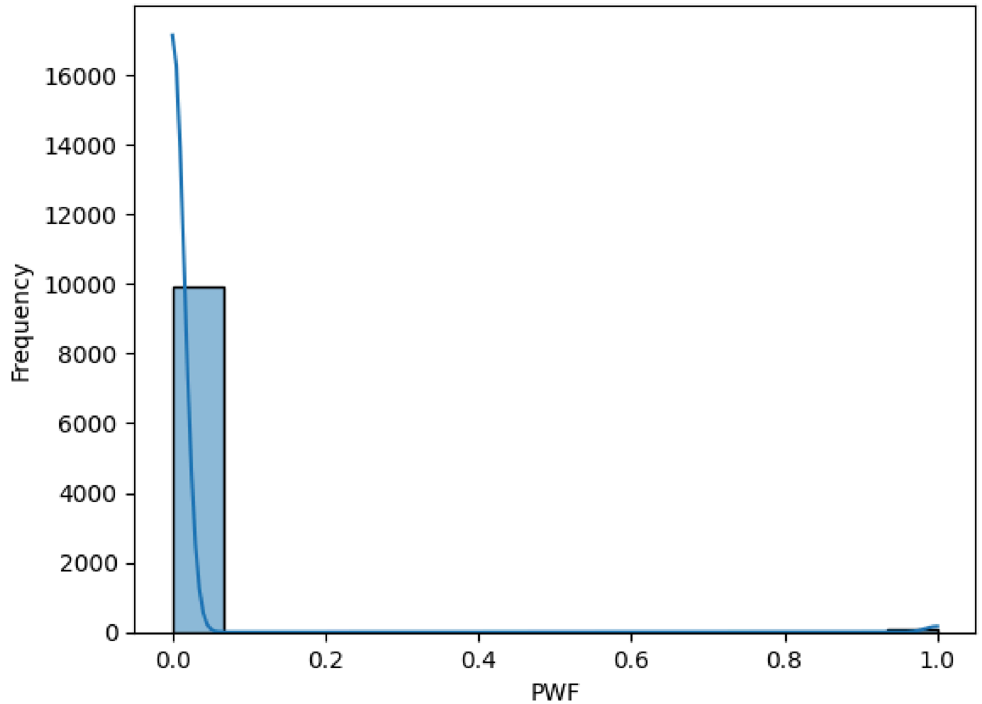
Distribution of TWF



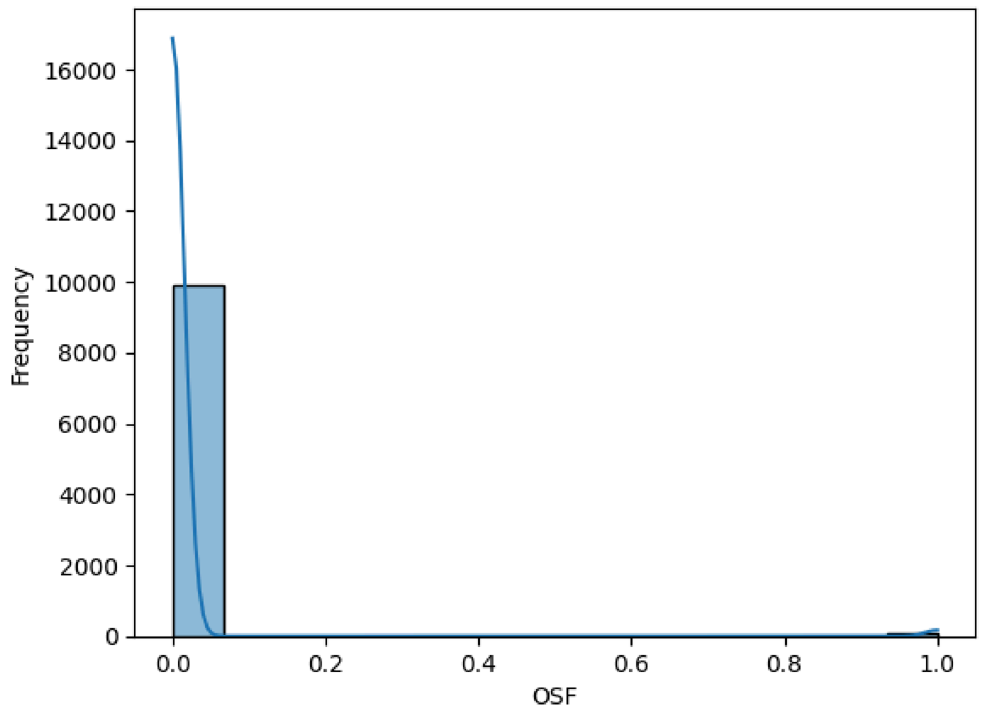
Distribution of HDF



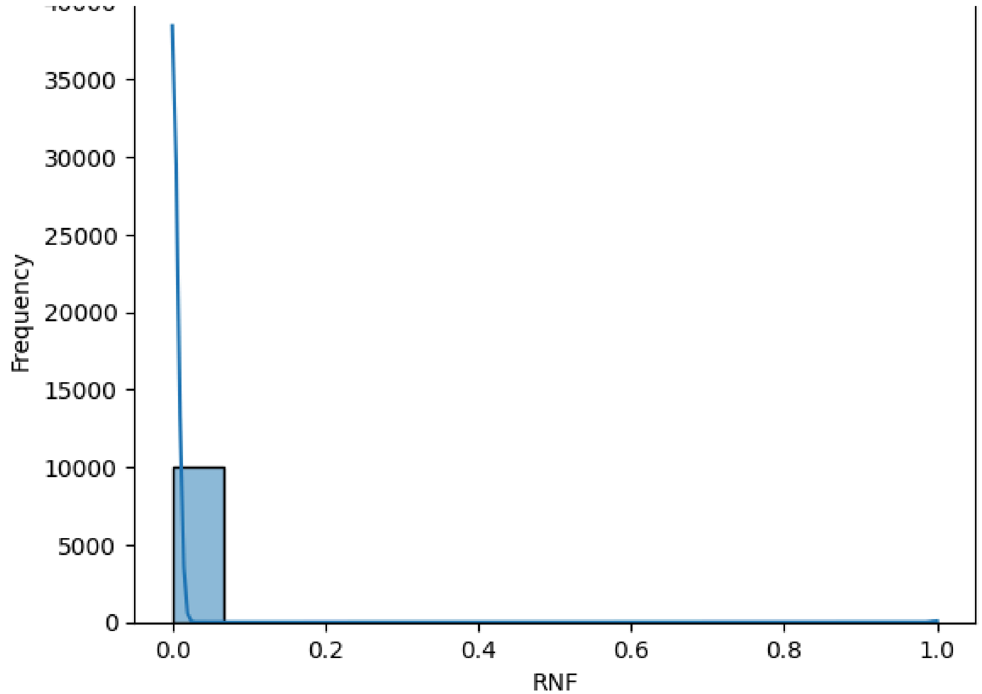
Distribution of PWF



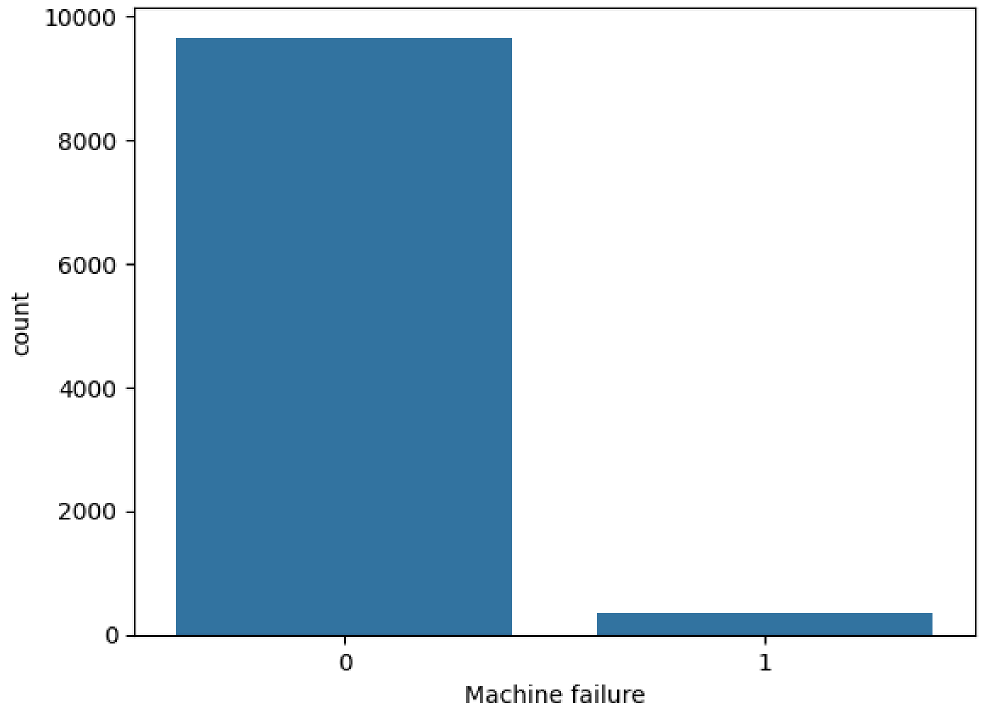
Distribution of OSF



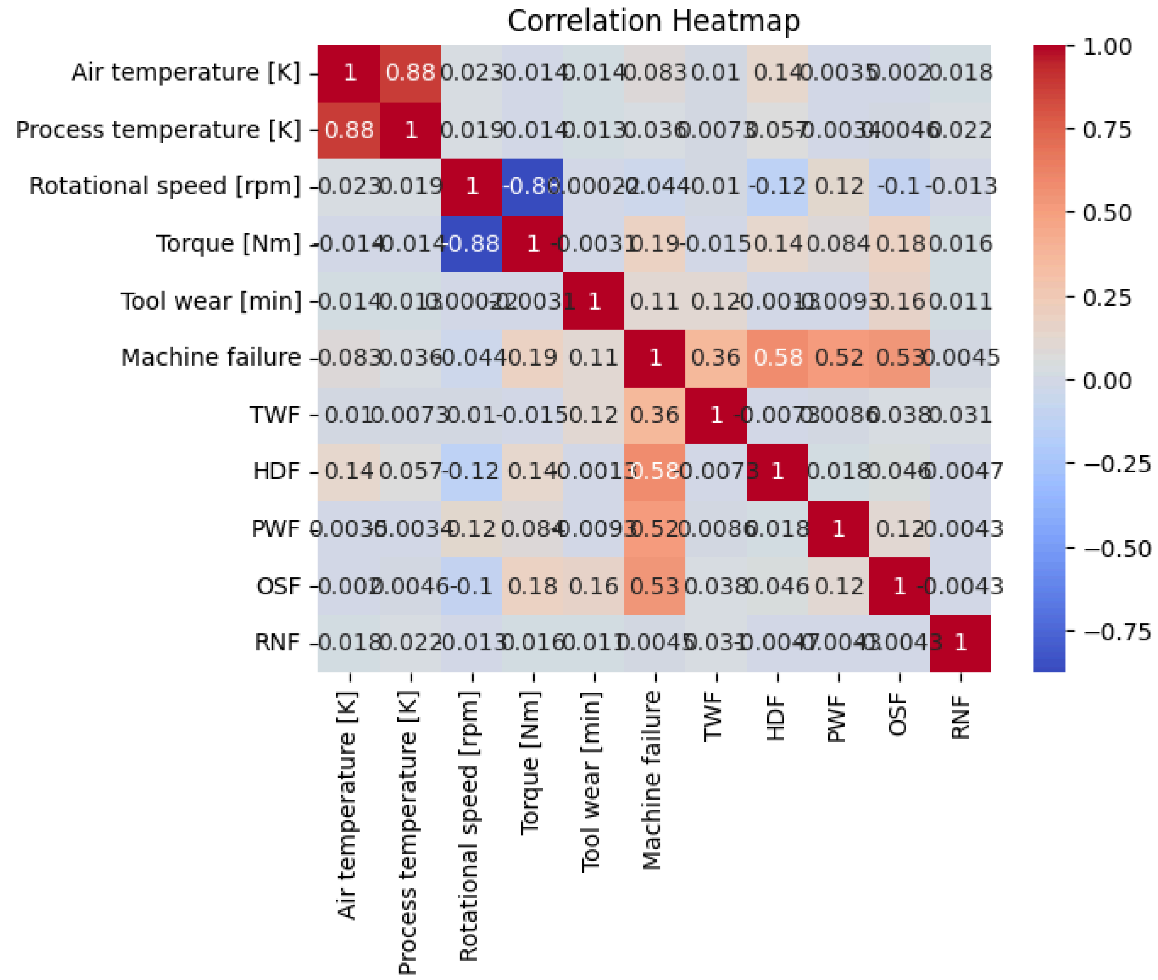
Distribution of RNF




```
sns.countplot(x='Machine failure', data=df)
plt.show()
```

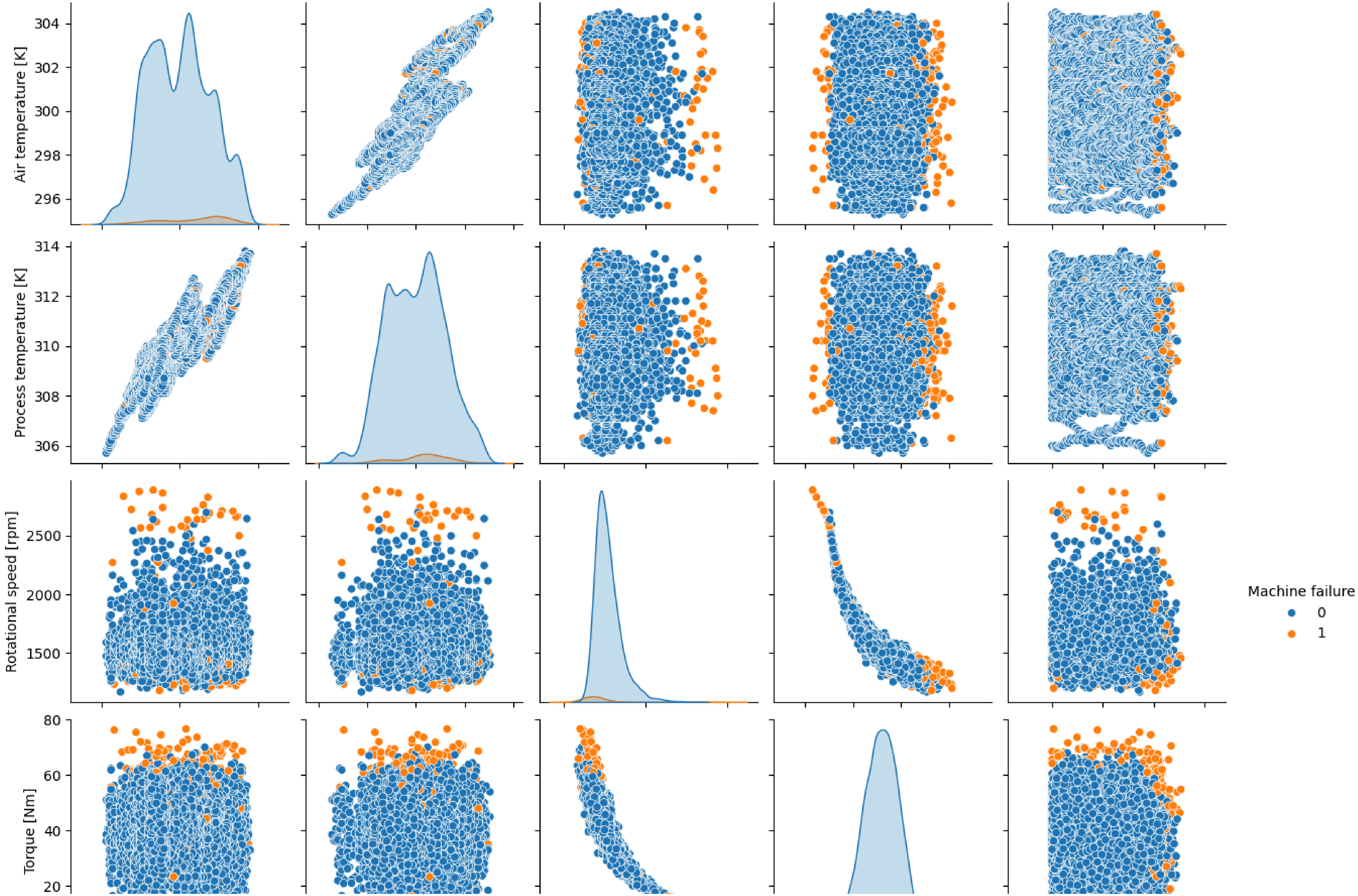


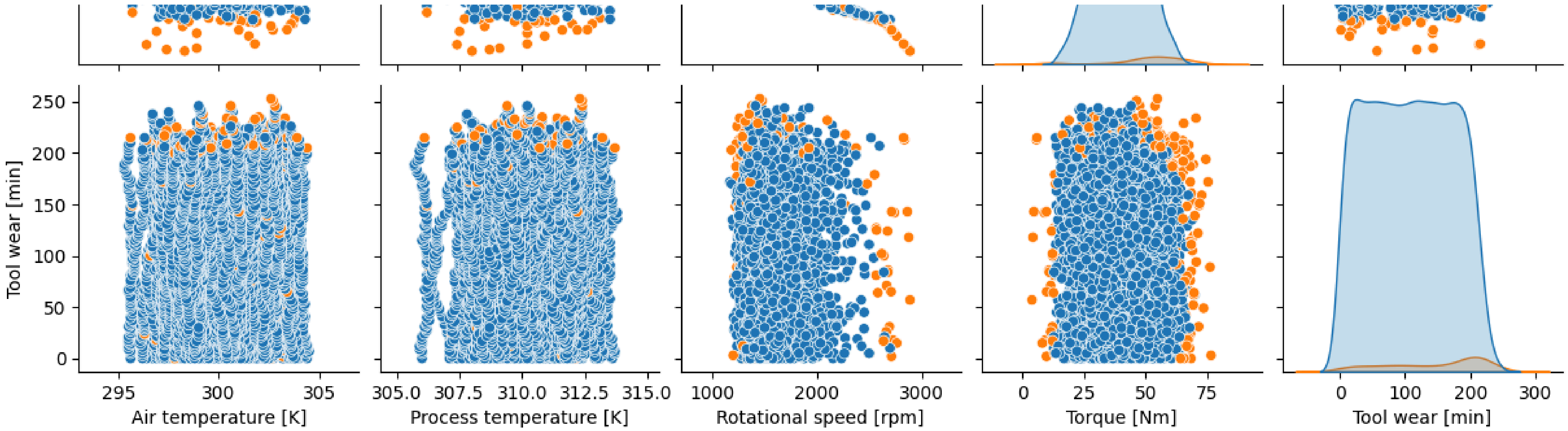

```
cor_matrix = df[num_features].corr()
sns.heatmap(cor_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



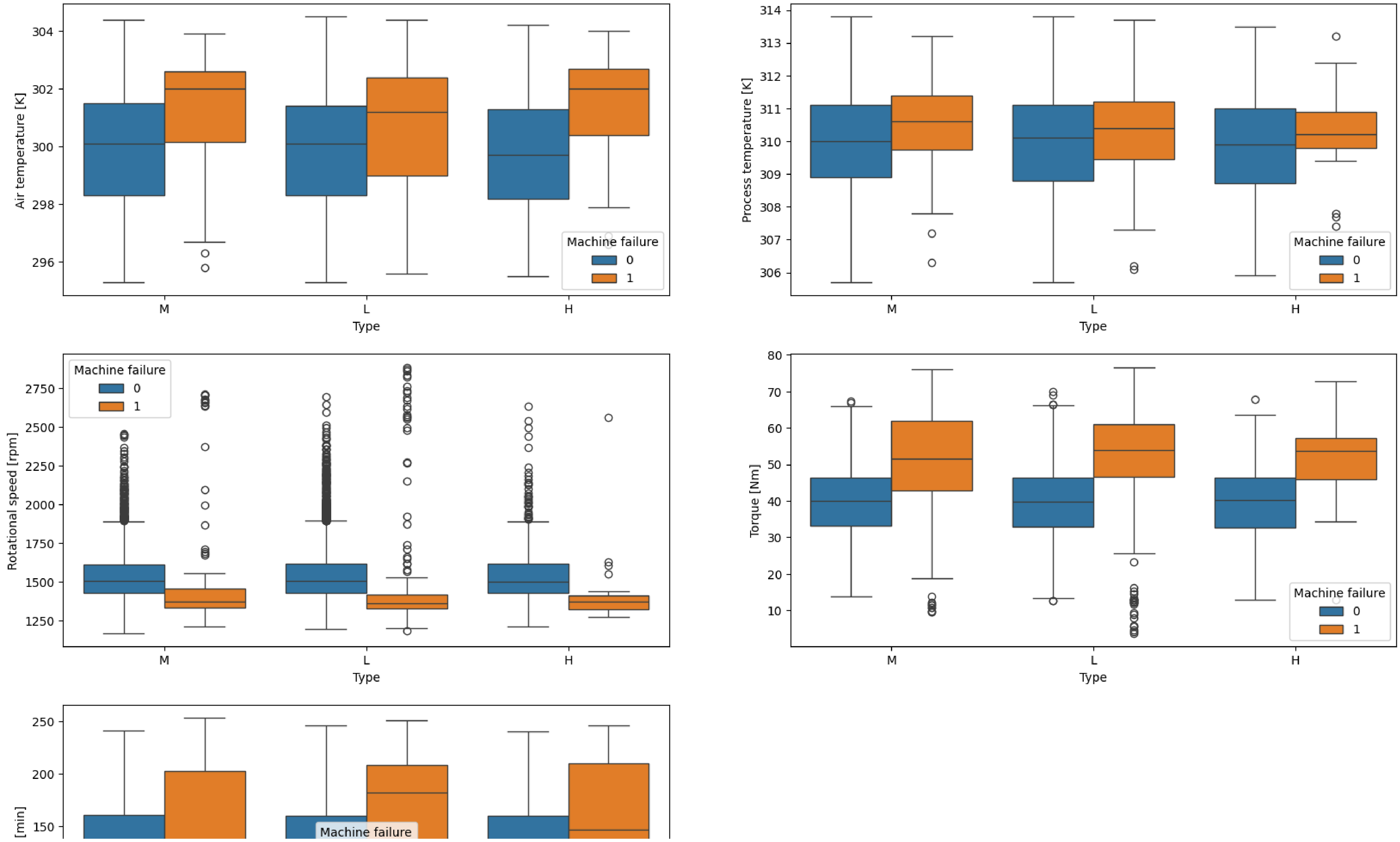

```
sns.pairplot(df.drop(['TWF','HDF','PWF','OSF','RNF'], axis=1),hue='Machine failure')
```

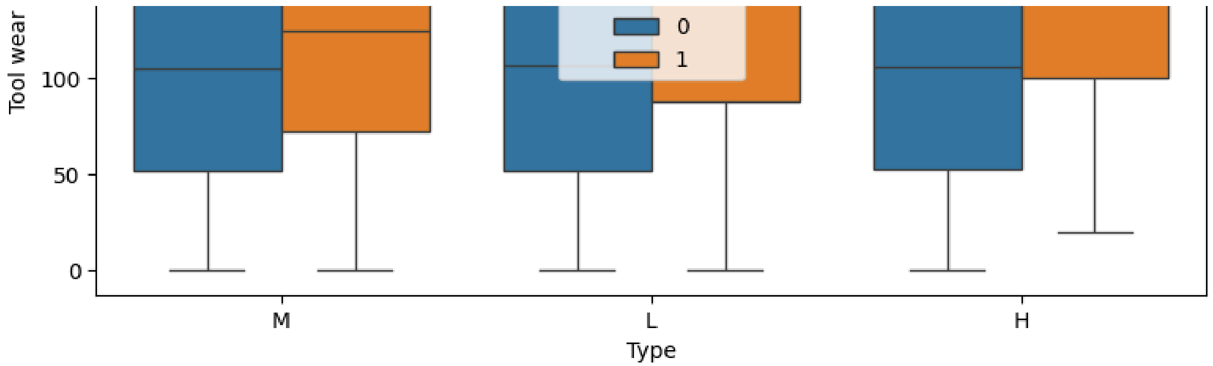
```
<seaborn.axisgrid.PairGrid at 0x7cd706943eb0>
```





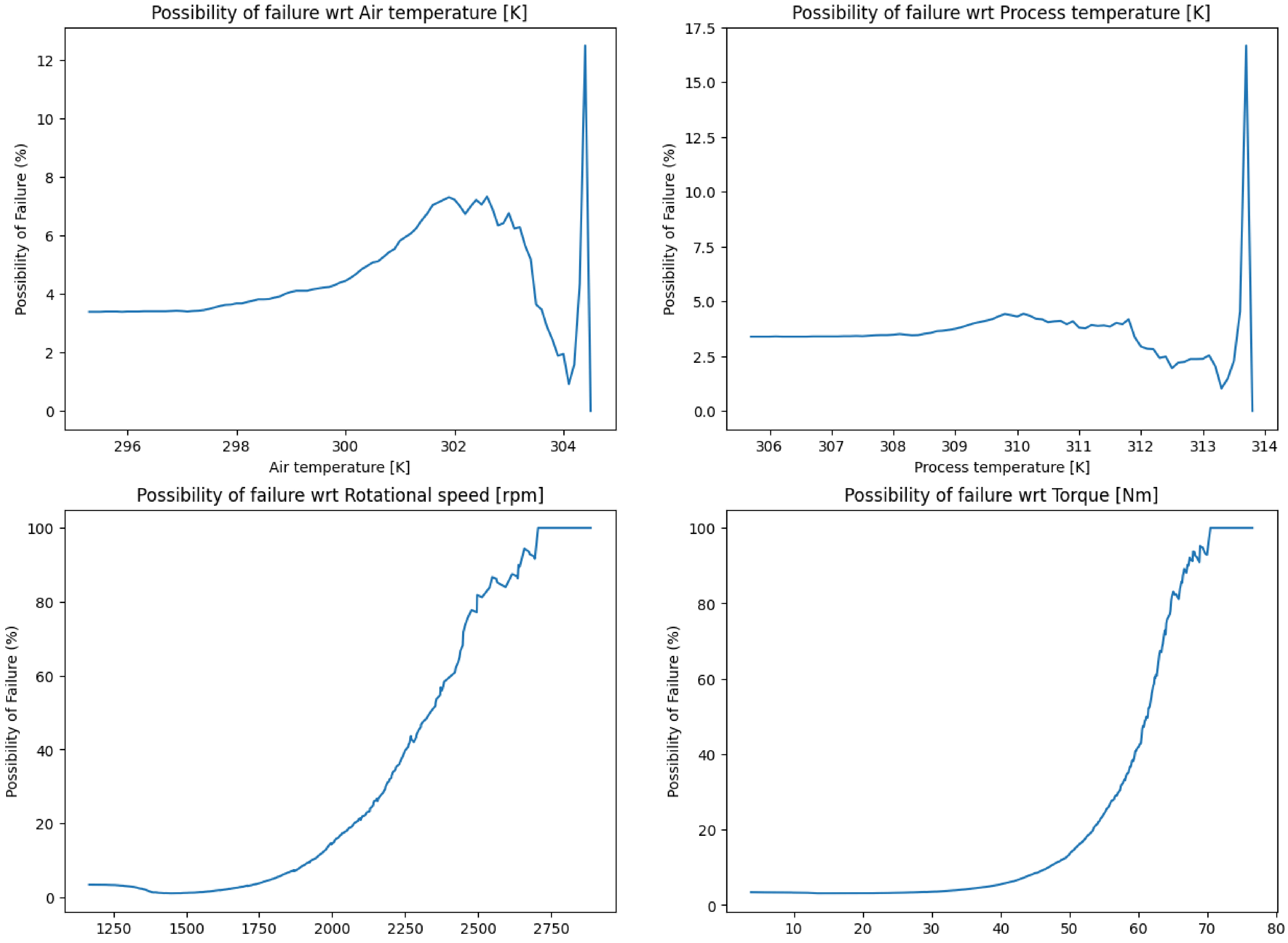

```
plt.figure(figsize = (20,15))
m=1
for i in ['Air temperature [K]', 'Process temperature [K]','Rotational speed [rpm]', 'Torque [Nm]', 'Tool wear [min]'] :
    plt.subplot(3,2,m)
    sns.boxplot(data=df, y = i, x="Type", hue="Machine failure")
    m+=1
```


Air temperature [K]
304
302
300
298
296
Machine failure
0
1
M
L
H
Type
Process temperature [K]
314
313
312
311
310
309
308
307
306
Machine failure
0
1
M
L
H
Type
Rotational speed [rpm]
2750
2500
2250
2000
1750
1500
1250
Machine failure
0
1
M
L
H
Type
Torque [Nm]
80
70
60
50
40
30
20
10
Machine failure
0
1
M
L
H
Type
[min]
250
200
150
Machine failure




```
def feat_prob(feature, data):
    x,y = [],[]
    for j in df[feature].unique():
        temp = data
        temp = temp[temp[feature]>=j]
        y.append(round((temp['Machine failure'].mean()*100),2))
        x.append(j)
    return(x,y)
```

```
plt.figure(figsize=(15,17))
m=1
for i in ['Air temperature [K]', 'Process temperature [K]','Rotational speed [rpm]', 'Torque [Nm]', 'Tool wear [min]'] :
    plt.subplot(3,2,m).set_title(label=("Possibility of failure wrt "+i))

    x,y = feat_prob(i,df)  #function call
    plt.xlabel(i)

    plt.ylabel("Possibility of Failure (%)")
    sns.lineplot(y=y,x=x)
    m+=1
```



Possibility of failure wrt Air temperature [K]
Possibility of Failure (%)
0
2
4
6
8
10
12
296
298
300
302
304
Air temperature [K]
Possibility of failure wrt Process temperature [K]
Possibility of Failure (%)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
306
307
308
309
310
311
312
313
314
Process temperature [K]
Possibility of failure wrt Rotational speed [rpm]
Possibility of Failure (%)
0
20
40
60
80
100
1250
1500
1750
2000
2250
2500
2750
Possibility of failure wrt Torque [Nm]
Possibility of Failure (%)
0
20
40
60
80
100
10
20
30
40
50
60
70
80