



# *Maintenance prédictive*

04 Janvier, 2025

Réalisé par : SGHIR Marwa

Supervisé par : Pr.KHAMJANE Aziz



# Plan:

- **Problématique/Objectif**
- **Analyse du projet (Data preparation/Modeling)**
- **Evaluation**
- **Demonstration**



# *Problématique :*



- Comment prédire les défaillances des machines à partir des données historiques afin de réduire les arrêts imprévus et les coûts associés ?
- Comment utiliser un ensemble de données pour former un modèle qui peut anticiper les pannes futures et aider à planifier la maintenance de manière proactive ?

## *Objectif :*

- Développer un modèle de classification binaire qui prédit si une machine tombera en panne ou non.



# Dataset overview:



	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



# *Data preparation:*



## *Nettoyage et Formatage des Données:*

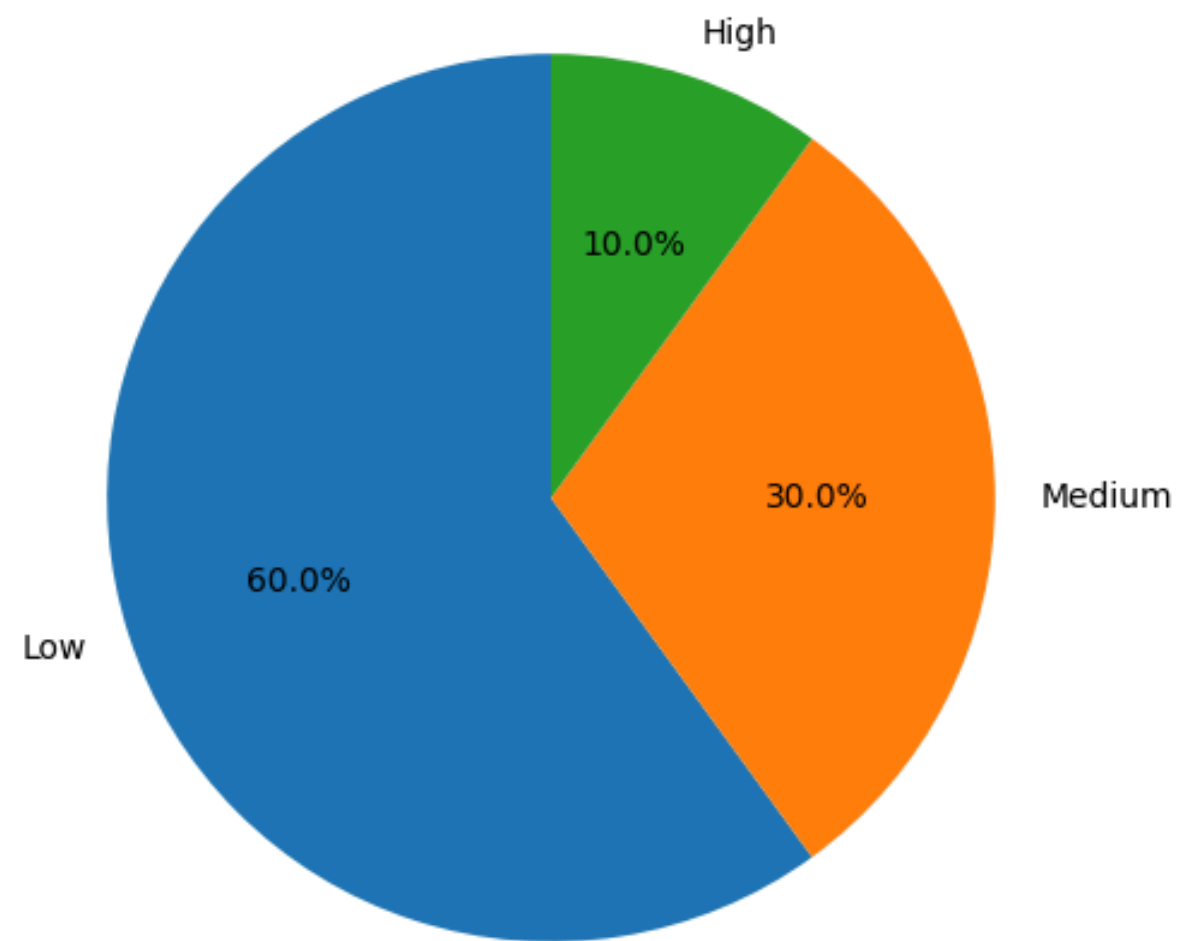
- Supprimer les colonnes inutiles, telles que les identifiants ou celles non pertinentes pour la prédiction, comme 'UDI', 'Product ID' .
- Remplir les valeurs manquantes avec la moyenne de chaque colonne .
- Data conditioning
- Data type conversion
- Handling outliers: Density-Based Anomaly Detection (LOF)



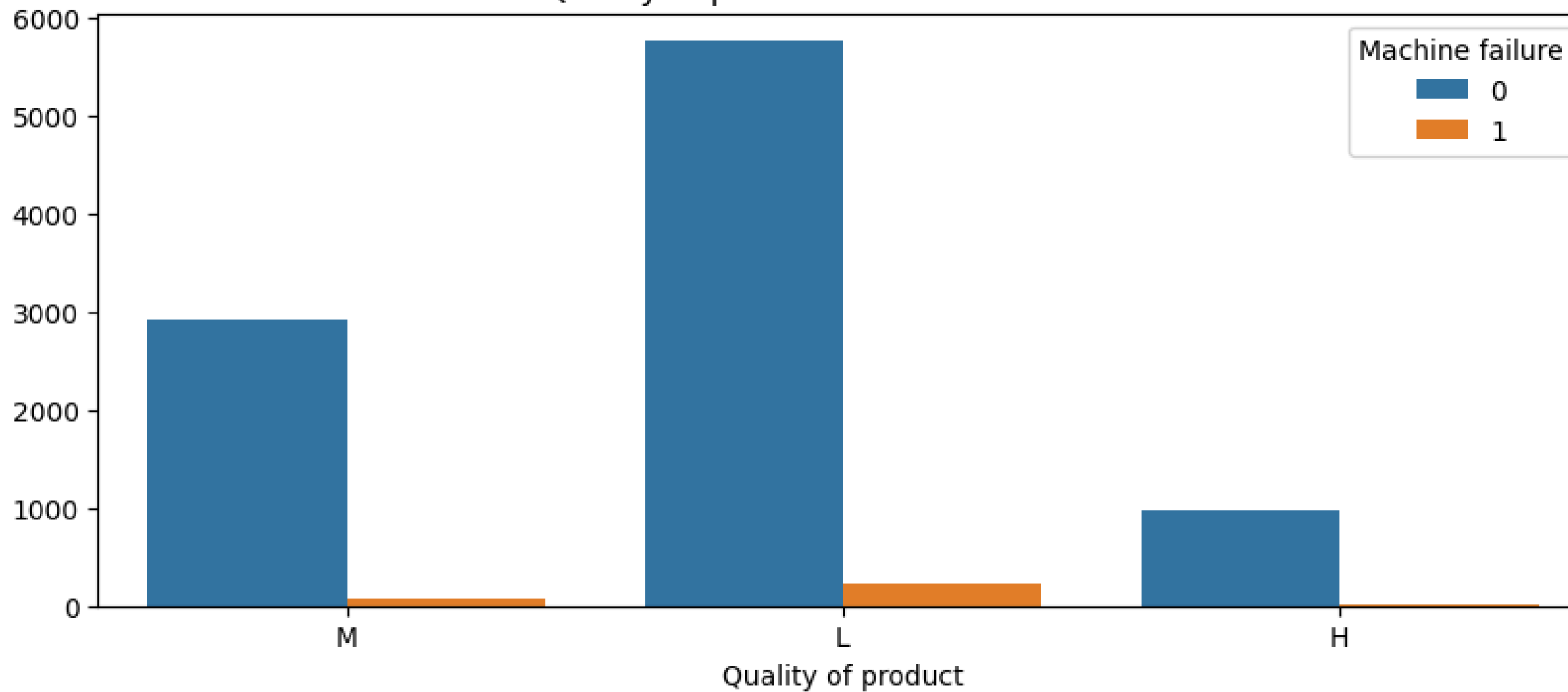
# *Exploratory data analysis (EDA):*

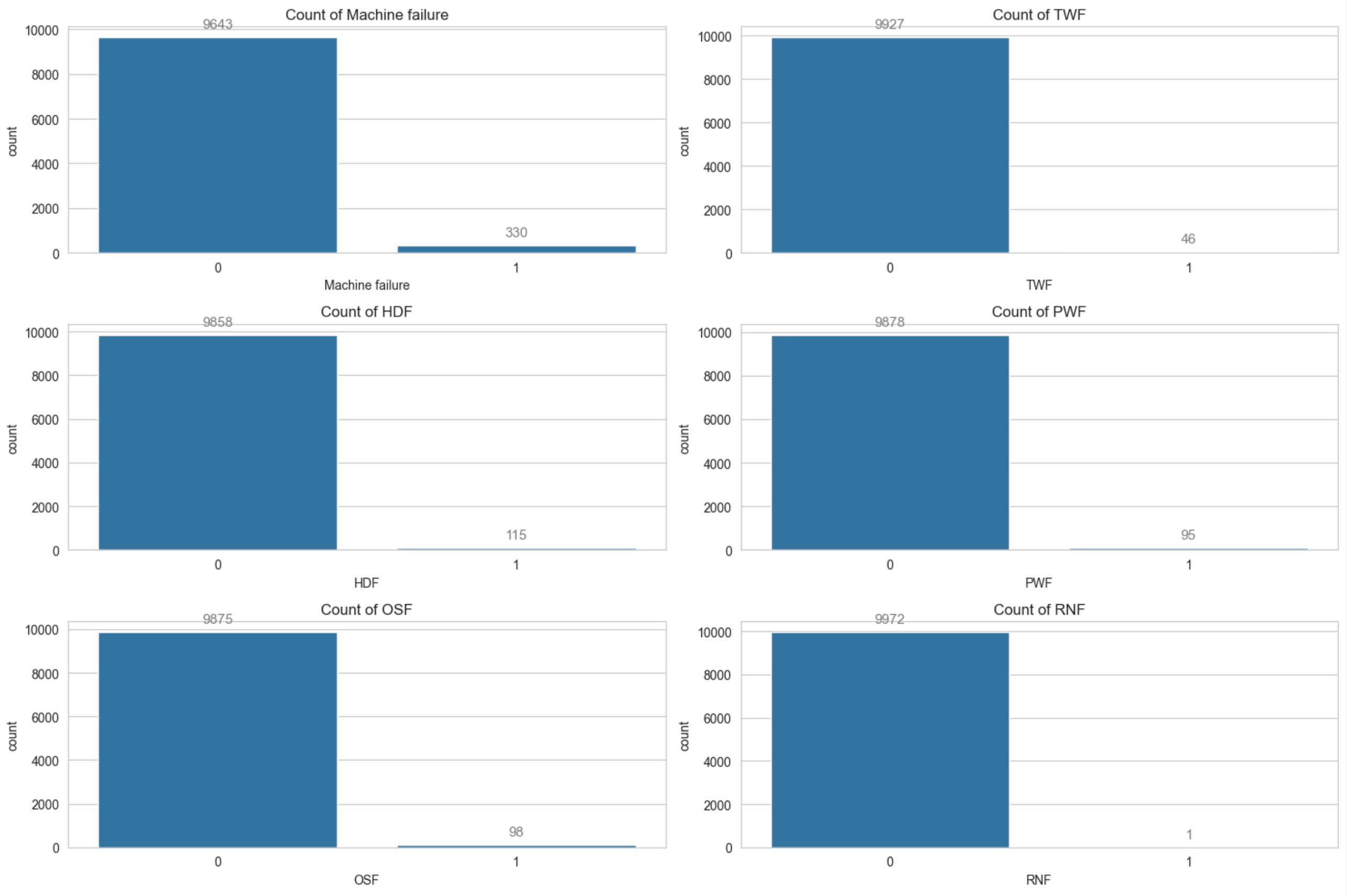


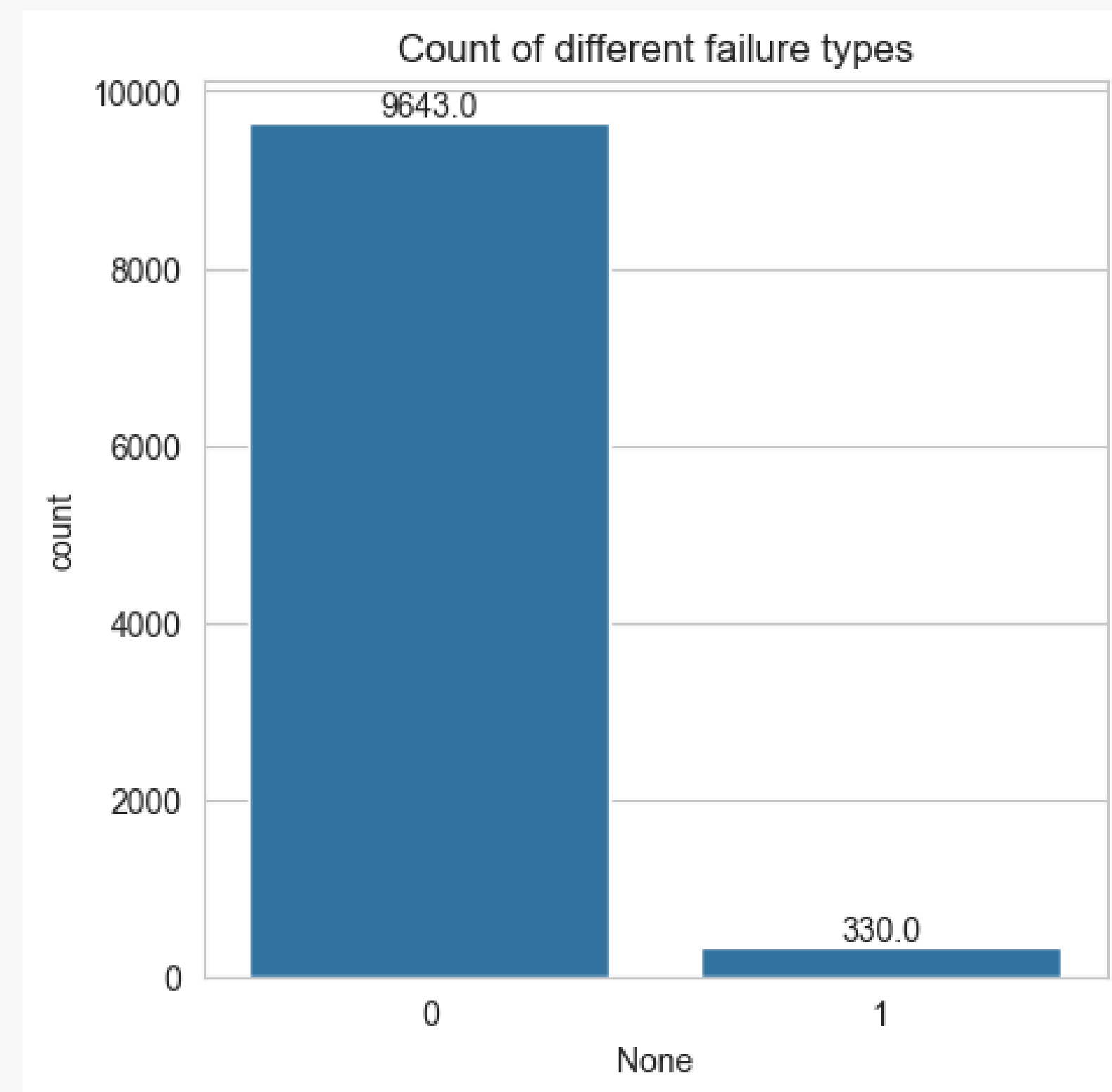
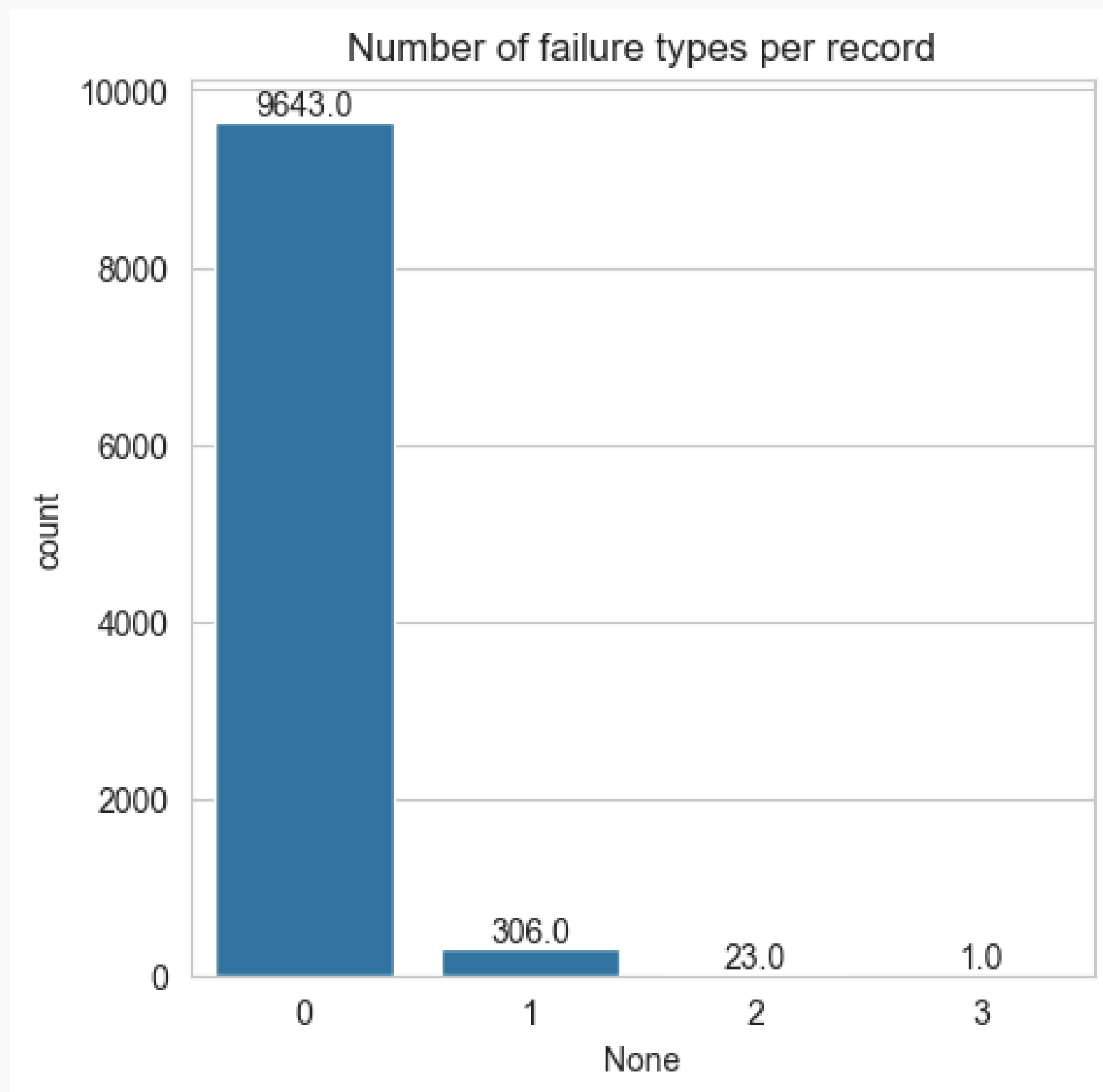
Distribution of Quality Types



Quality of product VS Machine failure

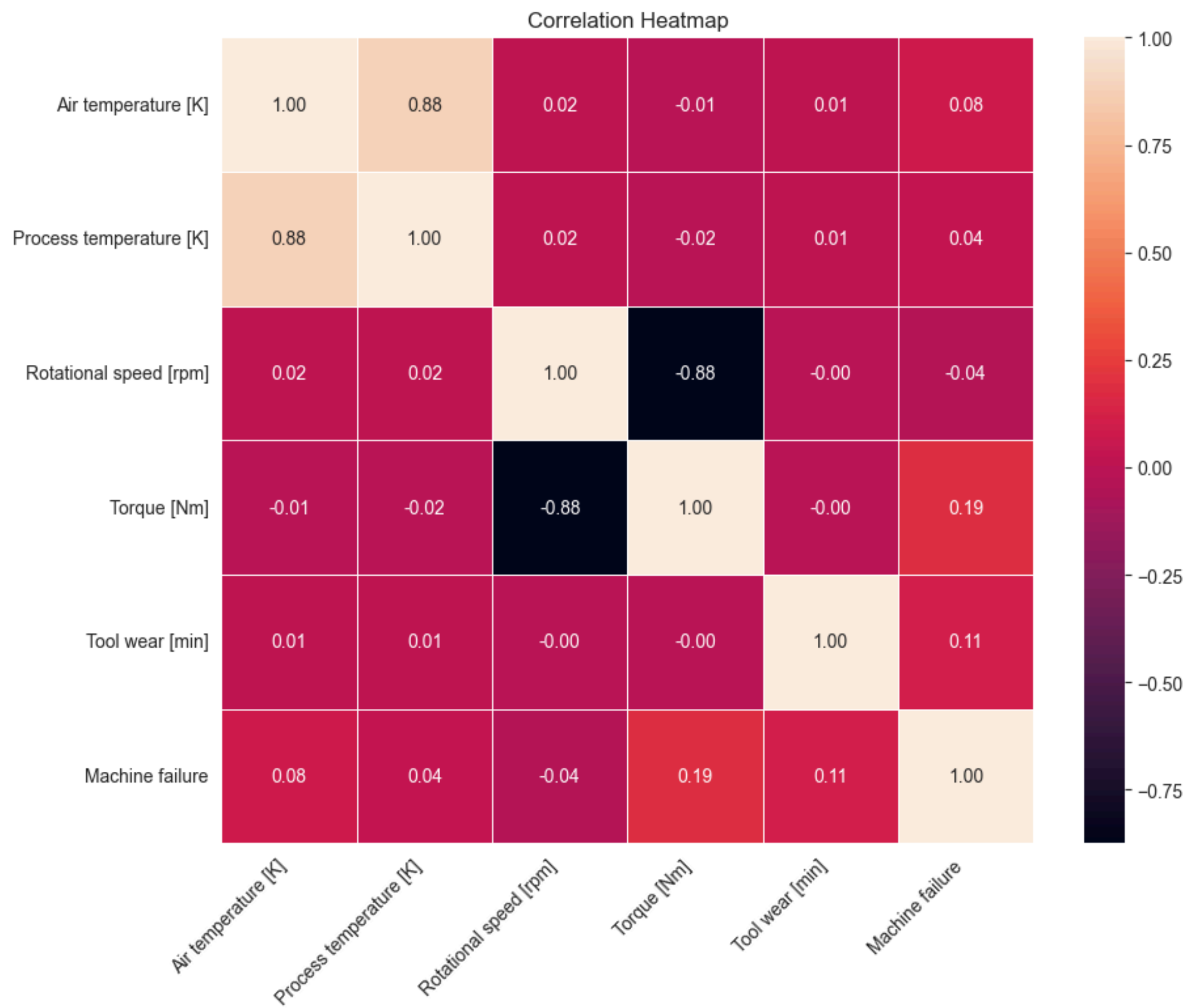








# Correlation Heatmap:



# *Feature Engineering:*

**Power** = Rotational speed \* Torque

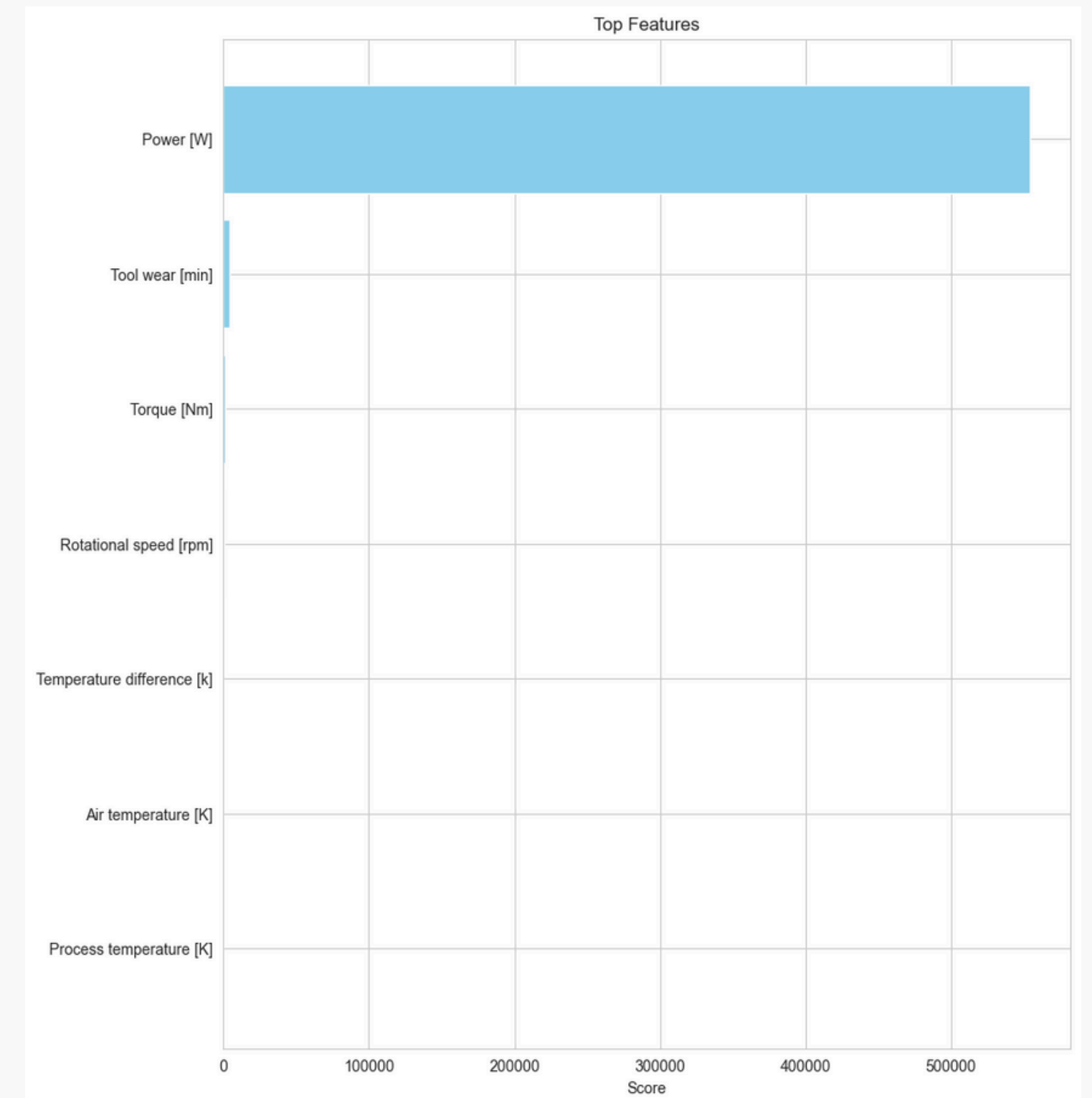
**Temperature difference** = Process temperature - Air temperature

	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	Power [W]	Temperature difference [k]
0	298.1	308.6	1551	42.8	0	0	66382.8	10.5
1	298.2	308.7	1408	46.3	3	0	65190.4	10.5
2	298.1	308.5	1498	49.4	5	0	74001.2	10.4
3	298.2	308.6	1433	39.5	7	0	56603.5	10.4
4	298.2	308.7	1408	40.0	9	0	56320.0	10.5

# *Feature selection:*

## Sélection univariée des caractéristiques

- Utilisation de **SelectKBest** avec la **statistique chi2**.
- Évaluation de l'association de chaque caractéristique avec la variable cible.
- Sélection des caractéristiques les plus significatives, particulièrement pour les variables catégorielles.



# *Feature selection:*

## la régression pas-à-pas en arrière (backward stepwise regression)

- Commence avec toutes les caractéristiques candidates.
- Supprime itérativement la caractéristique la moins significative .
- Continue jusqu'à ce que toutes les caractéristiques restantes aient une relation significative avec la variable cible.

```
backward_regression(X, y)
```

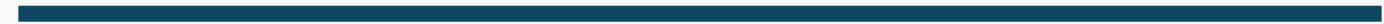
Selected Features:

```
['Air temperature [K]', 'Process temperature [K]', 'Rotational speed [rpm]', 'Torque [Nm]', 'Tool wear [min]', 'Power [W]', 'Temperature difference [k]']
```

.....



# *Modeling*

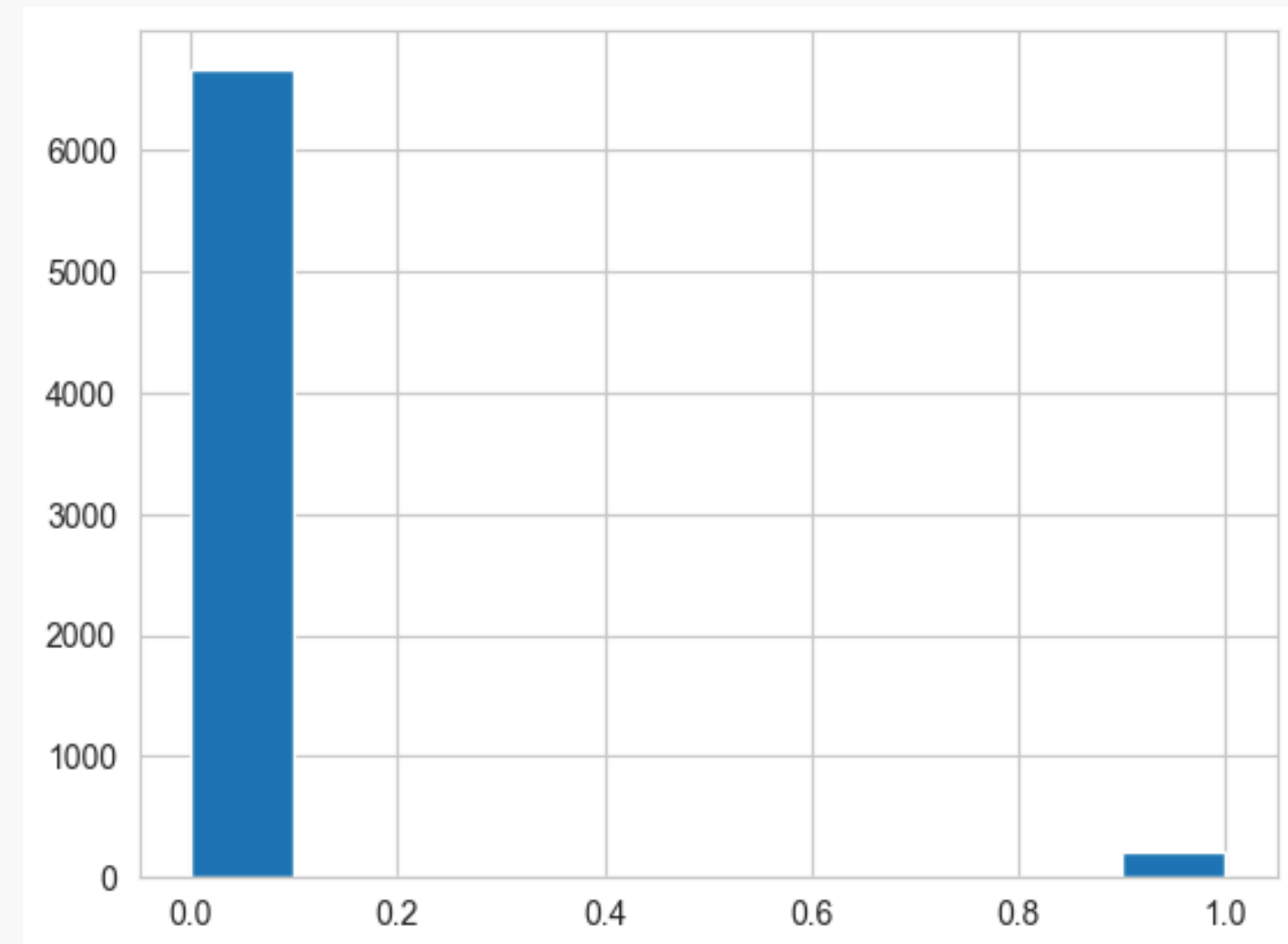
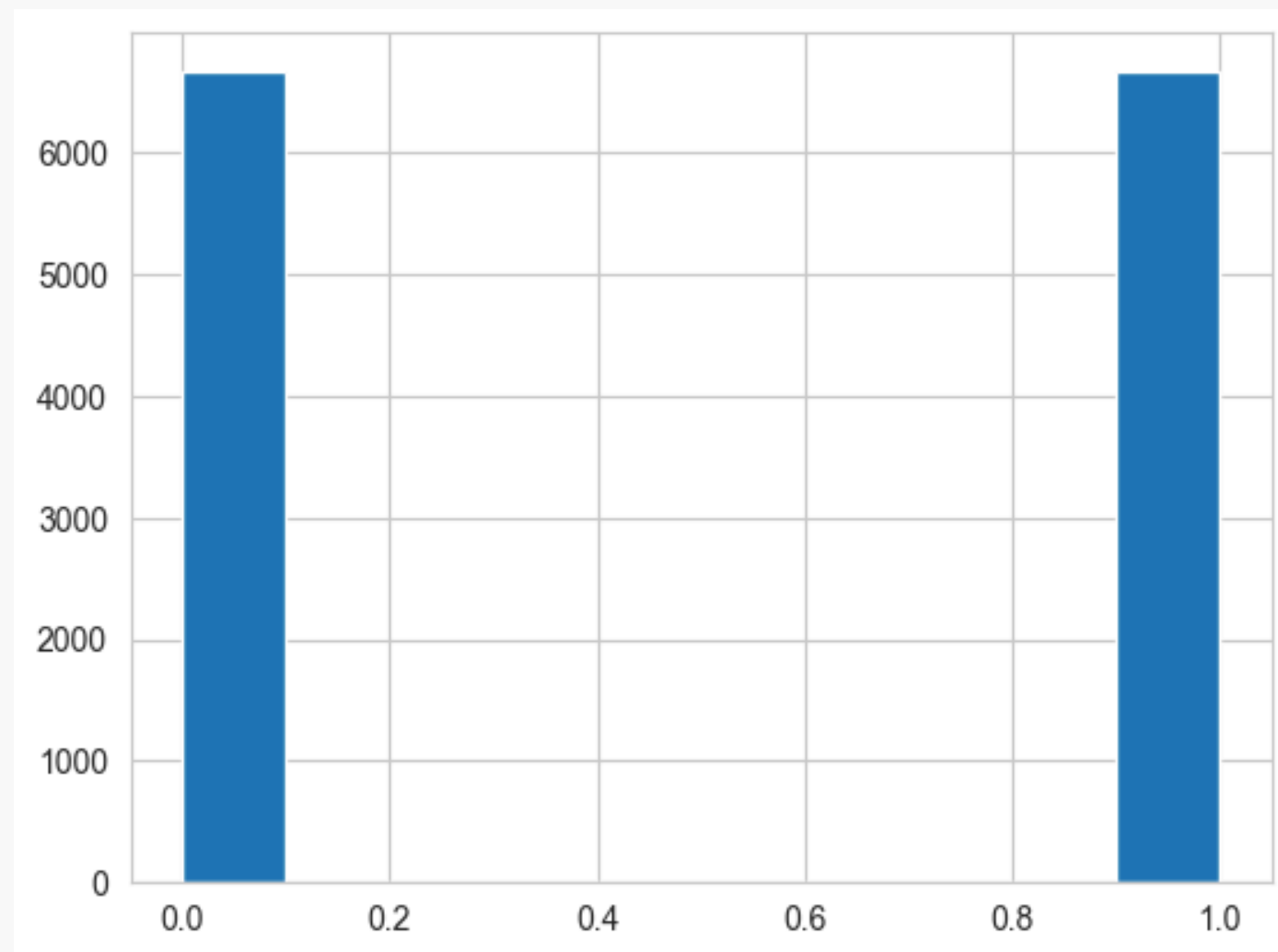


.....



# Traitement du Déséquilibre des Classes

**SMOTE (Synthetic Minority Over-sampling Technique)** a été utilisé pour équilibrer les classes en générant des échantillons synthétiques pour la classe minoritaire. Cela permet d'éviter le biais vers la classe majoritaire et d'améliorer la précision des prédictions pour les deux classes

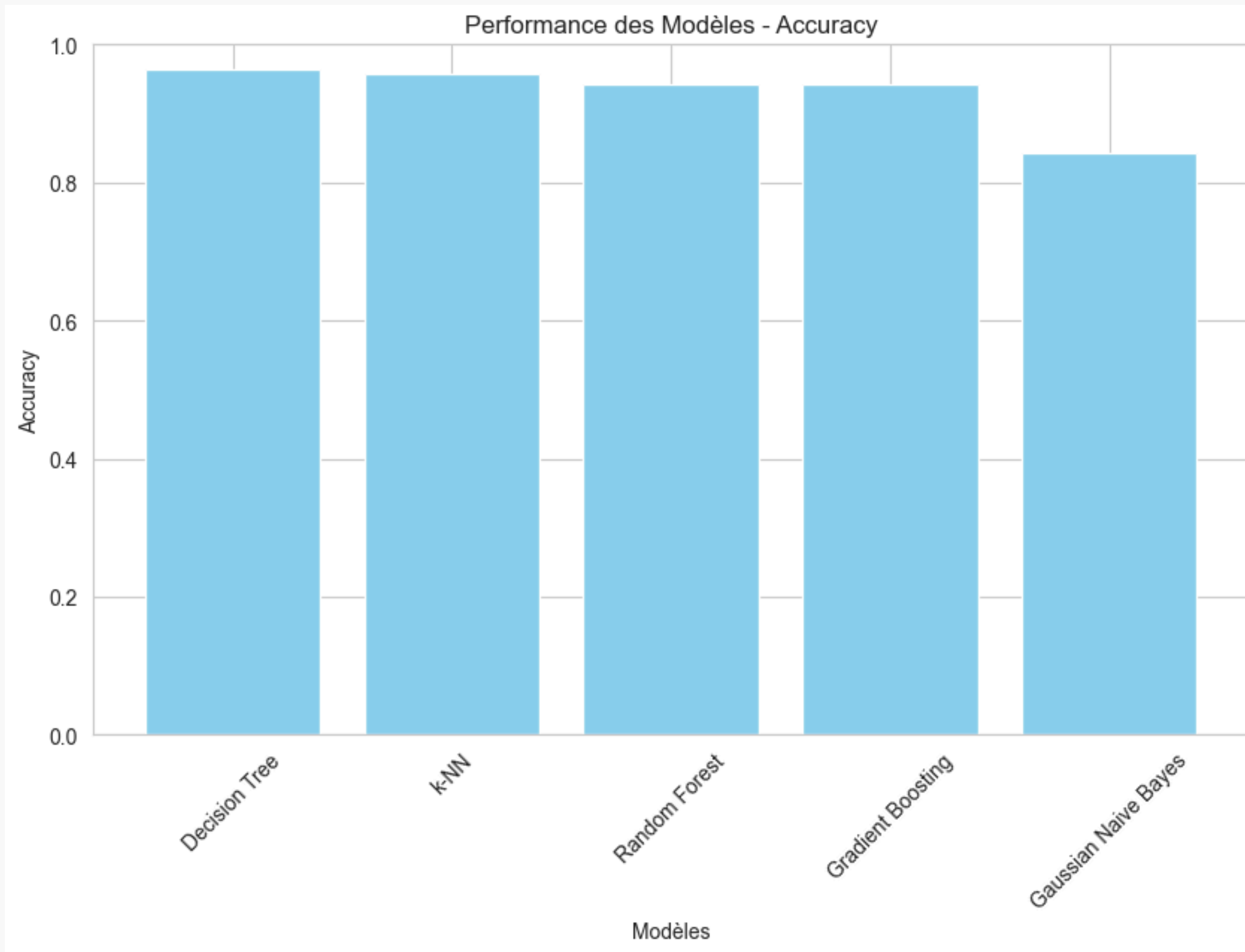




# Modèles utilisés

- Decision Tree
- k-NN (K-nearest neighbors)
- Random Forest
- Gradient Boosting
- Gaussian Naive Bayes

# *Evaluation des modèles:*





Gaussian Naive Bayes				
	precision	recall	f1-score	
0.0	0.99	0.84	0.91	
1.0	0.15	0.81	0.25	
accuracy			0.84	
macro avg	0.57	0.83	0.58	
weighted avg	0.97	0.84	0.89	

Random Forest Model				
	precision	recall	f1-score	
0.0	1.00	0.94	0.97	
1.0	0.36	0.92	0.51	
accuracy			0.94	
macro avg	0.68	0.93	0.74	
weighted avg	0.98	0.94	0.96	

k-NN Model				
	precision	recall	f1-score	
0.0	0.99	0.96	0.98	
1.0	0.41	0.74	0.53	
accuracy			0.96	
macro avg	0.70	0.85	0.75	
weighted avg	0.97	0.96	0.96	

Decision Tree				
	precision	recall	f1-score	
0.0	0.99	0.97	0.98	
1.0	0.47	0.77	0.58	
accuracy			0.96	
macro avg	0.73	0.87	0.78	
weighted avg	0.98	0.96	0.97	

Gradient Boosting				
	precision	recall	f1-score	
0.0	1.00	0.94	0.97	
1.0	0.35	0.92	0.51	
accuracy			0.94	
macro avg	0.68	0.93	0.74	
weighted avg	0.98	0.94	0.95	

.....



# *Démonstration technique*



.....



*Merci pour votre attention*

