

Projet de prédition des pannes

Introduction

La prévention des pannes devient de plus en plus indispensable dans les opérations industrielles. Elle permet d'éviter les arrêts non planifiés, de réduire les retards de livraison aux clients et de limiter les pertes financières.

Dans ce contexte, l'analyse des données industrielles permet d'anticiper les défaillances, de réduire les coûts et d'améliorer la performance globale.

Problématique

Comment peut-on anticiper et prévenir les pannes des machines industrielles à l'aide des données ?

Réponse

La maintenance prédictive permet d'analyser les données issues des capteurs afin de prévoir les défaillances avant qu'elles ne se produisent.

L'objectif de ce projet est de développer un modèle simple de prédition des pannes basé sur le machine learning afin d'aider à la prise de décision.

Dataset

Nom : ai4i2020

Ce dataset contient des mesures issues de capteurs, l'état des machines ainsi que des indicateurs opérationnels dans un environnement industriel semi-réel.

Chaque observation est unique, c'est-à-dire qu'il n'y a pas de doublons.

Variables principales :

- UDI : identifiant unique.
- Product ID : identifiant du produit.
- Air temperature [K] : température de l'air.
- Process temperature [K] : température du processus.
- Rotational speed [rpm] : vitesse de rotation.
- Torque [Nm] : couple.
- Tool wear [min] : usure de l'outil.

- Machine failure : présence (1) ou non (0) d'une panne.
 - TWF, HDF, PWF, OSF, RNF : types de pannes.
 - Type_L, Type_M, Type_H : types de machines.
 - Temp_diff [K] : différence de température.
 - power_factor [W] : puissance estimée.
 - Tool_wear_norm [min] : usure normalisée.
-

Technologies utilisées

- Python : analyse descriptive et développement du modèle.
 - Machine Learning : régression logistique.
 - Power BI Desktop : visualisation et tableaux de bord.
 - MySQL : stockage des données dans la table machine_data.
-

Plan de l'étude

1. Analyse des données

- Vérification des valeurs manquantes (aucune détectée).
- Encodage des variables catégorielles.
- Normalisation des variables numériques.
- Analyse des distributions.
- Analyse des corrélations.
- Identification des tendances.

2. Développement du modèle

- Séparation des données : 80 % entraînement et 20 % test.
- Modèle utilisé : régression logistique.
- Évaluation : accuracy et matrice de confusion.

3. Visualisation Power BI

Vue globale :

- 10 000 observations.
- 339 pannes.
- Usure moyenne : 107,95 minutes.
- Répartition des pannes par type de machine.

Vue détaillée :

- Analyse de l'usure par type.
 - Analyse des types de pannes.
 - Comparaison des performances selon le type de machine.
-

Résultats

Accuracy : 86,15 %

Matrice de confusion :

- 1675 machines sans panne correctement détectées.
 - 48 pannes correctement détectées.
 - 264 fausses alertes (machines prédictes en panne alors qu'elles fonctionnent).
 - 13 pannes non détectées.
-

Interprétation

Le modèle atteint une exactitude de 86,15 %, ce qui indique une bonne performance globale.

Il détecte correctement la majorité des machines sans panne.

Pour les machines en panne :

- Il détecte environ 79 % des pannes réelles.
- Mais il génère un nombre important de fausses alertes.

Les faux positifs entraînent des maintenances inutiles.

Les faux négatifs sont plus critiques car ils représentent des pannes non détectées.

Le dataset est déséquilibré (peu de pannes par rapport aux machines normales), ce qui influence les performances du modèle.

L'analyse Power BI montre que :

- Le type L présente le taux de panne le plus élevé.
 - Le type H présente le taux le plus faible.
 - Les pannes les plus fréquentes sont HDF, suivies de OSF et PWF.
 - Une usure élevée des outils augmente significativement la probabilité de panne.
-

Conclusion

Ce projet démontre l'intérêt de la maintenance prédictive pour l'optimisation des opérations industrielles. La combinaison de l'analyse des données, du machine learning et des tableaux de

bord interactifs permet une meilleure planification de la maintenance, réduit les arrêts non planifiés et améliore la performance industrielle .