\* NLP & Maintenance Prédictive Titre : Prédiction des défaillances dans les installations solaires/éoliennes via l'analyse sémantique Objectifs du Projet Objectifs principaux: • Analyser automatiquement les rapports techniques des installations solaires et éoliennes. • Détecter précocement les signaux faibles (anomalies récurrentes, pannes critiques) grâce aux techniques NLP. • Aider les équipes techniques à identifier rapidement les équipements à risque afin d'améliorer la réactivité face aux défaillances potentielles.

Ces données synthétiques simulent de manière réaliste les descriptions des pannes et anomalies couramment rencontrées dans les équipements industriels spécifiques aux secteurs solaire et

Objectifs secondaires:

• Explorer les performances d'un modèle NLP supervisé pour prédire la criticité des incidents. • Développer une application web intuitive permettant la consultation, la saisie et la gestion proactive des rapports techniques. • Visualiser les résultats de prédiction et les performances du modèle via des tableaux de bord interactifs (Power BI).

El Sources de Données

Données utilisées : • Structurées : Dataset synthétique CSV (rapport\_maintenance\_synthetique.csv)

éolien.

♦ Saisie des Rapports :

M-Report

• Possibilité d'indiquer si la panne a été corrigée lors de la saisie initiale.

Colonnes: rapport (texte), panne\_critique (0 ou 1).

• Non structurées : Rapports textuels simulés, inspirés des incidents fréquents dans les installations solaires/éoliennes réelles. Description :

Traitements NLP effectués

Prétraitement des textes : • Nettoyage avancé des textes (minuscules, suppression ponctuations et chiffres). • Lemmatisation et suppression des stopwords français avec spaCy (fr\_core\_news\_lg).

• Génération d'une colonne nettoyée rapport\_clean.

Vectorisation & Modélisation :

Transformation des textes via TF-IDF (Term Frequency-Inverse Document Frequency).

• Classification supervisée par régression logistique (scikit-learn).

• Séparation entraînement/test : 70/30.

Évaluation & Interprétabilité :

• Mesure des performances : Accuracy, Precision, Recall, F1-score, ROC-AUC.

• Courbes de performance (ROC, calibration, gain cumulée). • Matrice de confusion.

• Analyse LIME (contribution des mots aux prédictions).

Présentation de l'Application Web Une plateforme web interactive (Django) a été développée afin d'exploiter efficacement le modèle NLP et fournir aux utilisateurs un outil pratique pour :

• Formulaire intuitif permettant la rédaction de nouveaux rapports techniques (auteur, dispositif, élément concerné, texte libre).

Nom du rédacteur: Alex Lefèvre

Nom du dispositif (ex. Éolienne Nord #2):

Rapport d'incident ou de maintenance:

Pellicule de rouille sur l'ossature.

Les pannes ont-elles été corrigées ?

**■** Enregistrer le rapport

Station solaire de Perpignan

Élément ou pièce concerné(e):

Corrosion / Humidité

ECRIRE RAPPORT D'INCIDENCE OU DE MAINTENANCE

← Retour à la liste Consultation Dynamique : • Page principale affichant deux catégories : o Incidents à traiter : Rapports non résolus, avec criticité prédite affichée sous forme de barre colorée (vert-jaune-rouge). o Incidents résolus / archivés : Historique des rapports clôturés. • Tri rapide et interactif des rapports par date, dispositif ou criticité. **M-Report** LISTE DES RAPPORTS + Écrire un rapport Incidents à traiter 💙 DATE -ÉLÉMENT CONCERNÉ CRITICITÉ \$ DISPOSITIF \$ STATUT 2025-05-24 16:40 Station photovoltaïque de Strasbourg Réducteur De Vitesse Flender À traiter Centrale solaire de Bordeaux 2025-05-24 16:39 Analyse/Logs À traiter

Capteur Uv

Charge / Impédance

Capteur Irradiance

ÉLÉMENT CONCERNÉ

Charge / Impédance

Endpoints/Api

LISTE DES RAPPORTS

ÉLÉMENT CONCERNÉ

Endpoints/Api

Charge / Impédance

Analyse/Logs

Comportement Sous Charge / Vent

**Auteur:** Antoine Giraud

Créé le : 2025-05-24 16:37

Des tableaux de bord interactifs ont été créés pour explorer les données, évaluer les performances du modèle et fournir une interprétation claire des résultats :

Rapports de maintenance technique : Visualisations Exploratoires

55

2024

2025

Total

Nombre de pannes

0,871

Courbe de calibration

1.0 -

0.0

Top 40 mots favorisant la prédiction

-1,569

-1,273

-1,256

-1,187

-1,150

-1,146

-1,146

-1,137

-1,100

-1,091

-1,047

-1,022

-1,016

-0,988

-0,923

-0,921

-0,908

-0,904

-0,859

-0,851

-0,851

-0,834

-0,834

"panne non critique"

surface

fichier

dérive

sortie

horloge

échouer

oscillation

relais principal

guide

bras

altérer

mesurer

bruit

api

log

traitement

modbu

ancrage

horloge inter...

oscillation ho...

encroûtement

encroûtemen...

precision

-- Parfaitement calibré

83

71

208

205

203

186

64

62

Pannes critiques mensuelles

53

53

51

Condensateurs

fixations structurelles

Frein de rotor Svendborg

microcontroleur interne

Pannes critiques vs non critiques

641

Critique

Rapports de maintenance technique : Evaluation du modèle

Probabilité prédite

Rapports de maintenance technique : Interprétabilité et Explicabilité du modèle

contact

entraîner

lubrifiant -

absence -

0

0,2

0,3

0,4

0,5

proba\_critique

0,6

0,7

0,8

0,9

pales

359

Non critique

1,000

recall

Enveloppe externe nacelle

Guindeau industriel Leroy-Somer

13

53 56

54

54

Top 5 des équipements ayant connu le plus de pannes critiques

January February March April May June July August September October November December Total

52

**Dispositif :** Station photovoltaïque de Strasbourg

Réducteur De Vitesse Flender

**Boîtier Capteur** 

Comportement Sous Charge / Vent

À traiter

À traiter

À traiter

À traiter

À traiter

+ Écrire un rapport

À traiter

52 415

52 641

54

13

13

13

constater

Nuage de mots des rapports de pannes critiques

Edevoir blocage pr

excessif • base

0,974

Courbe de gain cumulée

1.0

0.0

Analyse LIME : contribution des mots au score de criticité

0.02

Rapport: "absence film lubrifiant entraîner contact direct"

0.04

Distribution des probabilités de prédiction par classe réelle

1.0

Modèle

--- Modèle aléatoire

roc\_auc

Proportion d'échantillons (classés par proba)

0.10

226

STATUT

STATUT

Résolu

Résolu

CRITICITÉ \$

CRITICITÉ \$

Incidents à traiter 💙 DATE -2025-05-24 16:40

♦ Visualisation détaillée :

2025-05-24 16:39

2025-05-24 16:38

2025-05-24 16:37

2025-05-24 15:28

2025-05-24 15:24

2025-05-24 15:31

2025-05-24 15:29

2025-05-24 16:39

2025-05-24 16:39

2025-05-24 16:38

2025-05-24 16:37

2025-05-24 15:28

2025-05-24 15:31

2025-05-24 15:29

Style dark-mode, ergonomique et facile à lire.

Tableaux de bord Power BI

Page 1 — Exploration des données :

• Répartition des types d'anomalies

• Nombre de rapports par site d'installation

• Analyse temporelle des incidents critiques

• Équipements les plus souvent sujets aux pannes critiques

Unité solaire d'Avignon

Site éolien de Reims

Parc éolien d'Amiens

Station solaire de Perpignan

Ferme solaire de Toulouse

Plateforme éolienne de Brest

Centrale solaire de Bordeaux

Parc hybride de Grenoble

Station mixte de Marseille

Nombre de rappors par type d'anomalie

Ferme solaire de Nîmes

Éolienne offshore de Saint-Nazaire

• Nuage de mots représentant les mots-clés associés aux incidents critiques

• Utilisation de Tailwind CSS pour un design réactif et agréable.

Interface claire et intuitive :

DATE -

Rapports résolus / archivés ^

Parc hybride de Grenoble

Parc éolien d'Amiens

Station solaire de Perpignan

Centrale solaire de Bordeaux

Station photovoltaïque de Strasbourg

DISPOSITIF \$

Unité solaire d'Avignon

Unité solaire d'Avignon

2025-05-24 15:24 Centrale solaire de Bordeaux Mouvement anormal lors des rafales de vent. Rapports résolus / archivés ^ ✓ Marquer comme corrigé

DISPOSITIF \$

Unité solaire d'Avignon

Unité solaire d'Avignon

Parc éolien d'Amiens

• Modal interactif affichant les détails complets du rapport (texte intégral, auteur, dispositif concerné).

DISPOSITIF \$

Station photovoltaïque de Strasbourg

Centrale solaire de Bordeaux

Parc hybride de Grenoble

Station solaire de Perpignan

Station photovoltaïque de Sti

• Option pour marquer directement depuis le modal qu'un incident est résolu.

• Icônes interactives (FontAwesome) pour améliorer la navigation et l'interaction utilisateur.

Nombre de rapports par site Station photovoltaïque de Strasbo... Parc éolien de Lille Site hybride de Montpellier

mécanique

structurelle

électrique

capteur

logiciel

• Courbe ROC et matrice de confusion

0,910

Critique

Non critique

Courbe ROC

1.0

0.8

accuracy

**Confusion Matrix** 

Classe\_Predite Critique Non critique

641

• Courbes de calibration et de gain cumulée

type\_anomalie

📌 Page 2 — Évaluation du modèle :

• Affichage des métriques (Accuracy, Precision, Recall, F1-score, ROC AUC)

305

0,931

f1\_score

0.2

Page 3 — Interprétabilité & Explainabilité :

"panne critique"

mécanique

blocage

système

devoir

structure

défaillance

onduleur

provoquer

roulement

température

anormal

constater

valeur

contact

mémoire

coupure

fissure

excessif

boucle

Limites Actuelles du Projet

présente

comportement

sou

• Explications locales des prédictions par la méthode LIME

Top 40 mots favorisant la prédic → 🖂 …

• Distribution des scores de prédiction par classe réelle

False Positive Rate

• Identification des termes influençant fortement les prédictions critiques/non critiques

1,458

1,101

1,080

1,019

0,984

0,980

0,903

0,895

0,883

0,830

0,819

0,754

0,753

0,750

0,677

0,673

0,665

0,665

0,663

0,656

0,655

0,650

0,649

• Données synthétiques : absence de nuances et diversité réelles.

• V Développement d'une plateforme complète d'analyse prédictive :

• 🗁 Fichier Python(Notebook) de traitement NLP et création du modèle :

model\_creation/Maintenance\_predictive.ipynb

o Tableau de bord en temps réel pour prédire les incidents avant leur apparition.

Perspectives et Améliorations Futures

• V Utilisation de modèles NLP avancés :

Annexes et Ressources du Projet

model\_creation/rapport\_BI.pbix

- LIST\_REPORTS.png

screen\_shots/

ROC curve (AUC = 0.99)

1.0

0,736 -0,985 axe amortisseur 0,723 -0,984 structurel json 0,717 -0,954 microfissure plateforme 0,712 -0,944 fusible parasite 0,693 -0,937 logiciel session

Résultats et Analyses • Le modèle NLP permet une identification précise des rapports critiques grâce à une bonne performance générale. • Certains équipements (turbines éoliennes spécifiques, onduleurs solaires) montrent une fréquence élevée de rapports critiques, clairement identifiés par le modèle. • L'analyse sémantique a permis d'isoler des mots-clés (ex. « fuite », « usure », « température élevée », « bruit ») associés aux défaillances critiques.

• Intégration d'un jeu de données réel : o Collecte de données historiques de maintenance avec horodatages précis. o Analyse des successions d'événements pour anticiper les pannes futures.

o Application de modèles Transformers (BERT, RoBERTa) pour une meilleure compréhension sémantique contextuelle.

Afin d'étendre la portée prédictive du projet, les points suivants seront explorés dans les développements futurs :

• Absence de données temporelles réelles : impossibilité actuelle d'analyser la séquence d'événements et la temporalité des pannes.

• Modèle simple : la régression logistique reste efficace mais limitée face à des modèles NLP plus complexes tels que BERT ou Transformers.

o Système d'alerte préventif intégré à l'application web. • Mise en place d'une validation croisée robuste pour améliorer la généralisation du modèle.

• Dataset synthétique (CSV) : model\_creation/rapport\_maintenance\_synthetique.csv

• Tapport interactif Power BI:

• 🗁 Captures d'écran du projet :

— MODAL\_REPORT.png — CREATE\_REPORT.png ├── BI\_Visual\_Exploratory.png — BI\_Model\_Evaluation.png BI\_Interpretability\_Explainability.png

Auteur du Projet Tombo H. ANDRIAMAHATANA is tomixbo@gmail.com @ GitHub : Tomixbo

• 🌐 Dépôt GitHub complet : https://github.com/<votre-utilisateur>/NLP\_Maintenance\_Predictive.git