

Université Cadi Ayyad

École Nationale des Sciences Appliquées de Safi

Département Informatique, Réseau et Télécommunications

Filière Génie Aéronautique et Technologies de l'Espace

2ème Année du Cycle Ingénieur

Maintenance Prédictive Aéronautique par IA

Réalisé par :
MEZOUARI Saad
BOUKIOUD Hamza

Encadré par :
M. Ahmed Bentajer

Année Universitaire : 2025/2026

Remerciements

Nous tenons à exprimer notre reconnaissance envers **Monsieur Bentajer** pour son encadrement de qualité et son professionnalisme tout au long de notre formation. Son enseignement nous a permis d'approfondir notre compréhension des principes fondamentaux de la **cybersécurité**, ainsi que des enjeux liés à la protection des données et à la résilience des systèmes, éléments cruciaux dans l'environnement numérique actuel.

Nous lui sommes reconnaissants pour sa disponibilité constante, ses précieux conseils, sa patience et l'attention personnalisée qu'il accorde à chaque étudiant. Son engagement dans le module de cybersécurité renforce considérablement notre maîtrise des risques numériques et nous prépare efficacement aux défis du milieu professionnel.

Table des matières

RÉSUMÉ	3
1 INTRODUCTION	4
1.1 Contexte industriel	4
1.2 Problématique	4
1.3 Objectifs du projet	4
2 DONNÉES ET MÉTHODOLOGIE	5
2.1 Jeu de données NASA C-MAPSS	5
2.2 Description des variables	5
2.3 Méthodologie adoptée	5
3 ANALYSE EXPLORATOIRE ET PRÉTRAITEMENT	6
3.1 Analyse initiale des données	6
3.2 Identification des capteurs non-informatifs	6
3.3 Calcul de la variable cible (RUL)	6
3.4 Normalisation des données	6
4 MODÉLISATION	7
4.1 Régression Linéaire (modèle de référence)	7
4.2 Random Forest (modèle avancé)	7
4.3 Métriques d'évaluation	7
5 RÉSULTATS ET DISCUSSION	8
5.1 Comparaison des performances	8
5.2 Analyse des variables les plus importantes	8

5.3	Limites identifiées et validation	8
6	CONCLUSION ET PERSPECTIVES	9
6.1	Bilan général du projet	9
6.2	Perspectives d'intégration industrielle	9
6.3	Recommandations pour un déploiement réel	9
	RÉFÉRENCES	10

RÉSUMÉ

Ce projet a pour objectif de concevoir un système de maintenance prédictive destiné à l'industrie aéronautique en s'appuyant sur l'intelligence artificielle. L'ambition est de prédire la Durée de Vie Restante (RUL - Remaining Useful Life) des turboréacteurs à partir des données fournies par leurs capteurs embarqués.

Méthodologie : Nous exploitons le jeu de données NASA C-MAPSS, qui contient des données simulées de dégradation. Après une analyse exploratoire approfondie et un prétraitement rigoureux, nous comparons deux approches de modélisation : une régression linéaire (servant de référence) et un modèle Random Forest.

Résultats : Le modèle Random Forest obtient une erreur quadratique moyenne (RMSE) de 32,11 cycles et une erreur absolue moyenne (MAE) de 24,07 cycles sur l'ensemble de test, avec un coefficient de détermination R^2 de 0,403.

Conclusion : Bien qu'il puisse être amélioré, ce système prouve la faisabilité de la maintenance prédictive dans le domaine aéronautique et présente des perspectives prometteuses pour diminuer les coûts de maintenance et renforcer la sécurité.

Mots-clés : Maintenance prédictive, Intelligence Artificielle, Aéronautique, Random Forest, RUL, NASA C-MAPSS

1 INTRODUCTION

1.1 Contexte industriel

L'industrie aéronautique contemporaine doit relever d'importants défis en matière de maintenance. Les compagnies aériennes allouent jusqu'à 30% de leurs dépenses opérationnelles à l'entretien de leurs flottes. Les méthodes conventionnelles montrent des limites notables :

- **Maintenance corrective** : Intervention réalisée après une défaillance, ce qui est inacceptable du point de vue de la sécurité.
- **Maintenance préventive** : Remplacement programmé à intervalles réguliers, entraînant des coûts inutiles due au remplacement de composants encore opérationnels.



FIGURE 1 – Maintenance prédictive

La maintenance prédictive se positionne comme la solution optimale : elle consiste à exploiter les données des capteurs pour anticiper les défaillances potentielles et planifier les interventions au moment le plus approprié.

1.2 Problématique

Comment prédire avec exactitude la durée de vie restante d'un turboréacteur en se basant uniquement sur les données fournies par ses capteurs ? Cette prédiction doit être suffisamment fiable pour permettre une planification optimale des opérations de maintenance, tout en garantissant un niveau de sécurité maximal pour les vols.

1.3 Objectifs du projet

- Développer un algorithme d'intelligence artificielle capable d'estimer la Durée de Vie Restante (RUL) des turboréacteurs
- Comparer différentes approches de modélisation pour identifier la plus performante
- Évaluer la faisabilité technique d'un tel système dans un contexte industriel
- Formuler des recommandations concrètes en vue d'un déploiement opérationnel

2 DONNÉES ET MÉTHODOLOGIE

2.1 Jeu de données NASA C-MAPSS

Le projet s'appuie sur le jeu de données NASA C-MAPSS (Commercial Modular Aero-Propulsion System Simulation), reconnu comme une référence académique dans le domaine de la maintenance prédictive. Ses principales caractéristiques sont :

- Type : Données simulées représentant la dégradation progressive
- Moteurs : 100 turboréacteurs simulés
- Capteurs : 21 capteurs de surveillance par moteur
- Variables opérationnelles : 3 paramètres distincts
- Cycles : Suivi complet depuis le début de vie jusqu'à la panne finale

2.2 Description des variables

Chaque enregistrement comporte 26 variables distinctes :

1. Identifiant unique du moteur et numéro du cycle temporel
2. 3 paramètres opérationnels (altitude, nombre de Mach, position de la manette)
3. 21 mesures de capteurs de surveillance :
 - Températures (T2, T24, T30, T50)
 - Pressions (P2, P15, P30)
 - Vitesses de rotation (Nf, Nc)
 - Niveaux de vibrations et autres indicateurs de performance

2.3 Méthodologie adoptée

Le projet suit la méthodologie standard des projets de science des données :

1. Analyse Exploratoire des Données (EDA) pour comprendre leur structure
2. Prétraitement et création de nouvelles variables significatives
3. Modélisation avec comparaison de deux algorithmes différents
4. Évaluation rigoureuse et validation des performances
5. Interprétation des résultats et formulation de conclusions

3 ANALYSE EXPLORATOIRE ET PRÉTRAITEMENT

3.1 Analyse initiale des données

L'analyse exploratoire a permis de révéler plusieurs caractéristiques importantes :

- Les données ne présentent aucune valeur manquante
- Les différentes mesures de capteurs utilisent des échelles variées (degrés Celsius, psi, etc.)
- Certains capteurs affichent une évolution marquée au fur et à mesure de la dégradation
- D'autres capteurs maintiennent des valeurs constantes tout au long du cycle de vie

3.2 Identification des capteurs non-informatifs

Quatre capteurs présentent un écart-type nul, indiquant qu'ils ne varient pas du tout :

```
sensor_01, sensor_05, sensor_10, sensor_16
```

Ces capteurs ont été éliminés de l'analyse, car ils n'apportent aucune information utile pour la prédiction de la durée de vie restante.

3.3 Calcul de la variable cible (RUL)

La Durée de Vie Restante est calculée selon la formule suivante :

$$\text{RUL} = \text{Cycle_maximum_du_moteur} - \text{Cycle_actuel}$$

Par exemple : si un moteur connaît une défaillance complète au cycle 192, au cycle 100 son RUL est estimé à 92 cycles.

3.4 Normalisation des données

Nous avons appliqué une normalisation StandardScaler (Z-score) :

- Centre les données autour de zéro (moyenne = 0)
- Réduit l'échelle des valeurs (écart-type = 1)
- Cette étape est essentielle car les algorithmes d'IA sont sensibles aux différences d'échelle entre les variables

4 MODÉLISATION

4.1 Régression Linéaire (modèle de référence)

Ce modèle a été choisi pour plusieurs raisons :

- Sa simplicité et sa facilité d'interprétation
- Ses bonnes performances lorsque les relations sont linéaires
- Son utilité comme point de comparaison pour évaluer des modèles plus complexes

4.2 Random Forest (modèle avancé)

Le choix du Random Forest se justifie par ses caractéristiques :

- Capacité à modéliser des relations non-linéaires complexes
- Robustesse face aux valeurs aberrantes (outliers)
- Capacité à estimer l'importance relative des différentes variables
- Moindre sensibilité au surapprentissage comparé aux arbres de décision simples

Les hyperparamètres utilisés pour le Random Forest sont :

- Nombre d'arbres dans la forêt : 100
- Profondeur maximale des arbres : 10 niveaux
- Échantillons minimum requis pour diviser un nœud : 5
- Échantillons minimum requis dans une feuille : 2

4.3 Métriques d'évaluation

Trois métriques principales ont été utilisées pour évaluer les performances :

1. RMSE (Root Mean Square Error) : Erreur quadratique moyenne, qui pénalise davantage les grandes erreurs
2. MAE (Mean Absolute Error) : Erreur absolue moyenne, plus simple à interpréter
3. R² (Coefficient de détermination) : Proportion de la variance totale expliquée par le modèle

La validation des modèles a utilisé 20% des données d'entraînement, tandis que l'évaluation finale a été réalisée sur un ensemble de test distinct.

5 RÉSULTATS ET DISCUSSION

5.1 Comparaison des performances

Modèle	RMSE (Test)	MAE (Test)	R ² Score
Régression Linéaire	35,42 cycles	27,15 cycles	0,275
Random Forest	32,11 cycles	24,07 cycles	0,403

TABLE 1 – Comparaison détaillée des performances des deux modèles

L'amélioration apportée par le modèle Random Forest est significative :

- RMSE : réduction de 9,3%
- MAE : réduction de 11,3%
- R² : augmentation de 46,5%

5.2 Analyse des variables les plus importantes

Les cinq capteurs identifiés comme les plus prédictifs sont :

1. sensor_02 (pression à l'entrée du compresseur)
2. sensor_11 (température des gaz d'échappement)
3. sensor_07 (pression à la sortie du compresseur)
4. sensor_20 (consommation de carburant)
5. sensor_21 (efficacité du compresseur)

Cette hiérarchie correspond à une interprétation physique cohérente : la température et la pression apparaissent effectivement comme les indicateurs les plus sensibles au processus de dégradation.

5.3 Limites identifiées et validation

Plusieurs limites ont été constatées :

- La valeur de R² (0,403) reste modeste, indiquant que le modèle explique environ 40% de la variance totale
- Les données étant simulées, une validation sur des données réelles serait nécessaire
- L'approche adoptée est statique et ne prend pas en compte la dimension temporelle de la dégradation
- La complexité des phénomènes physiques sous-jacents dépasse parfois la capacité du modèle à les capturer

6 CONCLUSION ET PERSPECTIVES

6.1 Bilan général du projet

Le projet a rempli ses objectifs principaux :

- Développement d'un algorithme fonctionnel pour la prédition de la Durée de Vie Restante
- Comparaison méthodique de deux approches de modélisation distinctes
- Évaluation quantitative et objective des performances obtenues
- Interprétation pertinente des résultats en lien avec la réalité physique

Le modèle Random Forest démontre des performances nettement supérieures à la régression linéaire, avec une erreur moyenne réduite à 24 cycles.

6.2 Perspectives d'intégration industrielle

L'intégration potentielle dans un système de maintenance prédictive opérationnel pourrait suivre trois étapes :

1. Phase prototype : développement d'un tableau de bord de supervision avec système d'alertes
2. Phase de validation : tests approfondis sur données réelles en collaboration avec des partenaires industriels
3. Phase de déploiement : mise en œuvre pilote sur une flotte réduite d'avions

Les bénéfices attendus d'une telle intégration sont substantiels :

- Réduction des coûts de maintenance : estimée entre 15% et 20%
- Augmentation de la disponibilité des avions : entre 5% et 10%
- Amélioration significative de la sécurité : grâce à la détection précoce des anomalies

6.3 Recommandations pour un déploiement réel

Pour garantir le succès d'un déploiement industriel, nous recommandons :

1. Ajouter systématiquement une marge de sécurité de 10 à 15 cycles aux prédictions du modèle
2. Mettre en place un système de surveillance continue pour suivre les performances du modèle dans le temps
3. Prévoir un plan de secours opérationnel en cas de défaillance temporaire du système de prédition
4. Former spécifiquement le personnel de maintenance à l'utilisation et à l'interprétation des résultats
5. Valider et ré-entraîner régulièrement le modèle avec de nouvelles données pour maintenir sa pertinence

RÉFÉRENCES

1. Saxena, A., & Goebel, K. (2008). Turbofan Engine Degradation Simulation Data Set. NASA Ames Prognostics Data Repository.
2. Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing.
3. Breiman, L. (2001). Random Forests. Machine Learning.
4. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.
5. Airbus. (2022). Maintenance Planning Document. Airbus Technical Publications.