

# Analyse FMDS et Maintenance Prédictive

Jeu de Données de Dégradation de Moteurs Turbofan NASA C-MAPSS

Fiabilité, Maintenabilité, Disponibilité, Sécurité



Adam Boufeid

Karim Maktouf

3ème GMAM FM

15 novembre 2025

## Résumé

Ce rapport présente une analyse FMDS (Fiabilité, Maintenabilité, Disponibilité, Sécurité) complète et un cadre de maintenance prédictive appliqués au jeu de données de dégradation de moteurs turbofan NASA C-MAPSS. L'étude intègre la théorie classique de fiabilité avec des techniques modernes d'apprentissage automatique pour prédire la Durée de Vie Résiduelle (RUL) et développer des capacités de diagnostic. Les contributions clés incluent le calcul de métriques de fiabilité (MTBF : 206,31 cycles, taux de défaillance : 0,004847), l'analyse de la distribution de Weibull révélant des schémas de défaillance par usure ( $\beta = 4,41$ ), et l'évaluation de plusieurs modèles prédictifs atteignant les meilleures performances avec SVR (RMSE : 46,26 cycles,  $R^2 = 0,385$ ). L'analyse aboutit à des recommandations de maintenance actionnables et démontre l'application pratique des principes FMDS aux systèmes industriels.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Contexte Industriel . . . . .	5
1.2	Objectifs du Projet . . . . .	5
1.3	Description du Jeu de Données . . . . .	5
<b>2</b>	<b>Cadre Théorique</b>	<b>6</b>
2.1	Concepts FMDS . . . . .	6
2.1.1	Fiabilité . . . . .	6
2.1.2	Temps Moyen Entre Défaillances (MTBF) . . . . .	6
2.1.3	Fonction de Taux de Risque . . . . .	6
2.2	Distribution de Weibull . . . . .	6
<b>3</b>	<b>Méthodologie</b>	<b>7</b>
3.1	Préparation des Données . . . . .	7
3.1.1	Calcul de RUL . . . . .	7
3.1.2	Ingénierie des Caractéristiques . . . . .	7
3.2	Sélection des Caractéristiques . . . . .	7
3.3	Modèles Prédictifs . . . . .	7
3.3.1	Forêt Aléatoire . . . . .	8
3.3.2	Gradient Boosting . . . . .	8
3.3.3	Régression à Vecteurs de Support (SVR) . . . . .	8
<b>4</b>	<b>Résultats et Analyse</b>	<b>8</b>
4.1	Analyse Exploratoire des Données . . . . .	8
4.1.1	Statistiques du Jeu de Données . . . . .	8
4.1.2	Analyse de Corrélation . . . . .	9
4.2	Résultats de l'Analyse FMDS . . . . .	10
4.2.1	Métriques de Fiabilité . . . . .	10
4.2.2	Analyse de la Distribution de Weibull . . . . .	11
4.3	Performance des Modèles Prédictifs . . . . .	12
4.3.1	Comparaison des Modèles . . . . .	12
4.4	Analyse Diagnostique . . . . .	13
4.4.1	Performance de Classification . . . . .	13
4.4.2	Importance des Caractéristiques pour le Diagnostic . . . . .	14
4.5	Métriques Spécifiques au Pronostic . . . . .	14
<b>5</b>	<b>Recommandations de Maintenance</b>	<b>14</b>
5.1	Calendrier d'Inspection . . . . .	14
5.2	Maintenance Préventive . . . . .	15

5.3	Seuils de Maintenance Prédictive	15
5.4	Analyse Coût-Bénéfice	15
<b>6</b>	<b>Discussion</b>	<b>15</b>
6.1	Réalisations Clés	15
6.2	Limitations et Défis	16
6.3	Comparaison avec la Littérature	16
<b>7</b>	<b>Conclusions et Perspectives</b>	<b>16</b>
7.1	Conclusions	16
7.2	Perspectives Futures	17
7.2.1	Améliorations Techniques	17
7.2.2	Implémentation Industrielle	17
7.2.3	Analyses Avancées	17
<b>A</b>	<b>Annexe : Structure du Code</b>	<b>19</b>
A.1	Phase 1 : Chargement des Données	19
A.2	Phase 2 : Analyse Exploratoire des Données	19
A.3	Phase 3 : Analyse FMDS	19
A.4	Phase 4 : Ingénierie des Caractéristiques	19
A.5	Phase 5 : Modélisation Prédictive	19
A.6	Analyses Supplémentaires	19

# 1 Introduction

## 1.1 Contexte Industriel

La maintenance prédictive est devenue une stratégie critique dans les opérations industrielles modernes, particulièrement dans l'aviation où les défaillances de composants peuvent avoir des conséquences catastrophiques. La transition des paradigmes de maintenance réactive vers prédictive permet aux organisations d'optimiser les calendriers de maintenance, de réduire les temps d'arrêt et d'améliorer la sécurité opérationnelle tout en minimisant les coûts.

## 1.2 Objectifs du Projet

Ce projet vise à :

- Appliquer les concepts FMDS aux données réelles de moteurs turbofan
- Calculer les indicateurs de fiabilité (MTBF, MTTR, taux de risque)
- Mettre en œuvre des méthodes de diagnostic pour la détection de défauts
- Développer des modèles pronostiques pour l'estimation de RUL
- Générer des recommandations de maintenance actionnables

## 1.3 Description du Jeu de Données

Le jeu de données NASA C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) simule la dégradation de moteurs turbofan sous diverses conditions opérationnelles. Le jeu de données comprend :

- **Données d'entraînement** : 20 631 cycles opérationnels de 100 moteurs avec points de défaillance connus
- **Données de test** : 13 096 cycles de 100 moteurs sans temps de défaillance spécifiés
- **Vérité terrain** : Cycles restants réels pour les moteurs de test
- **Capteurs** : 21 mesures de capteurs plus 3 paramètres opérationnels

Chaque moteur démarre avec des degrés variables d'usure initiale et de différences de fabrication, fonctionnant normalement avant qu'une dégradation progressive ne conduise à la défaillance.

## 2 Cadre Théorique

### 2.1 Concepts FMDS

#### 2.1.1 Fiabilité

La fiabilité  $R(t)$  représente la probabilité qu'un système remplisse sa fonction prévue sans défaillance sur l'intervalle de temps  $[0, t]$ . Pour une distribution exponentielle :

$$R(t) = e^{-\lambda t} \quad (1)$$

où  $\lambda$  est le taux de défaillance constant.

#### 2.1.2 Temps Moyen Entre Défaillances (MTBF)

Le MTBF quantifie le temps opérationnel attendu entre les défaillances :

$$\text{MTBF} = \frac{1}{\lambda} = \int_0^{\infty} R(t) dt \quad (2)$$

#### 2.1.3 Fonction de Taux de Risque

Le taux de risque  $h(t)$  représente le taux de défaillance instantané au temps  $t$  :

$$h(t) = \frac{f(t)}{R(t)} = -\frac{d}{dt} \ln R(t) \quad (3)$$

## 2.2 Distribution de Weibull

La distribution de Weibull modélise les temps de défaillance avec une densité de probabilité :

$$f(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1} e^{-(t/\eta)^{\beta}} \quad (4)$$

où  $\beta$  est le paramètre de forme et  $\eta$  est le paramètre d'échelle. Le paramètre de forme indique :

- $\beta < 1$  : Défaillances précoces (mortalité infantile)
- $\beta = 1$  : Défaillances aléatoires (distribution exponentielle)
- $\beta > 1$  : Défaillances par usure (taux de risque croissant)

### 3 Méthodologie

#### 3.1 Préparation des Données

##### 3.1.1 Calcul de RUL

Pour les données d'entraînement, la RUL est calculée comme :

$$\text{RUL}_i = \text{cycle\_max}_{\text{moteur}} - \text{cycle}_i \quad (5)$$

Pour les données de test, la RUL incorpore la vérité terrain :

$$\text{RUL}_i = \text{cycles\_restants}_{\text{vérité}} + (\text{cycle\_max}_{\text{test}} - \text{cycle}_i) \quad (6)$$

##### 3.1.2 Ingénierie des Caractéristiques

Des caractéristiques avancées ont été créées pour capturer les patterns temporels :

- **Statistiques glissantes** : Moyennes mobiles et écarts-types (fenêtre=5)

$$s_{\text{moyenne\_glissante}} = \frac{1}{w} \sum_{j=i-w+1}^i s_j \quad (7)$$

- **Taux de changement** : Différences de premier ordre

$$s_{\text{diff}} = s_i - s_{i-1} \quad (8)$$

Ce processus a généré 99 caractéristiques totales à partir de 24 capteurs originaux.

#### 3.2 Sélection des Caractéristiques

Les caractéristiques ont été sélectionnées en fonction de leur corrélation avec la RUL, conservant 61 caractéristiques avec  $|\text{corrélation}| > 0,01$ . Les caractéristiques les plus corrélées incluaient les moyennes glissantes des capteurs s4, s11, s21, s15 et s20.

#### 3.3 Modèles Prédictifs

Trois modèles de régression ont été implémentés :

### 3.3.1 Forêt Aléatoire

Méthode d'ensemble combinant 100 arbres de décision avec une profondeur maximale de 20 :

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (9)$$

### 3.3.2 Gradient Boosting

Ensemble séquentiel construisant des modèles additifs :

$$F_m(x) = F_{m-1}(x) + \nu \cdot h_m(x) \quad (10)$$

où  $\nu$  est le taux d'apprentissage (0,1) et  $h_m$  est l'apprenant faible.

### 3.3.3 Régression à Vecteurs de Support (SVR)

Régression non-linéaire utilisant un noyau RBF avec  $C = 100$ ,  $\epsilon = 0, 1$ .

## 4 Résultats et Analyse

### 4.1 Analyse Exploratoire des Données

#### 4.1.1 Statistiques du Jeu de Données

L'analyse a révélé :

- Durée de vie moyenne des moteurs : 206,31 cycles (entraînement)
- Durée de vie médiane : 199,00 cycles
- Plage de durée de vie : 128–362 cycles
- Écart-type : 46,34 cycles

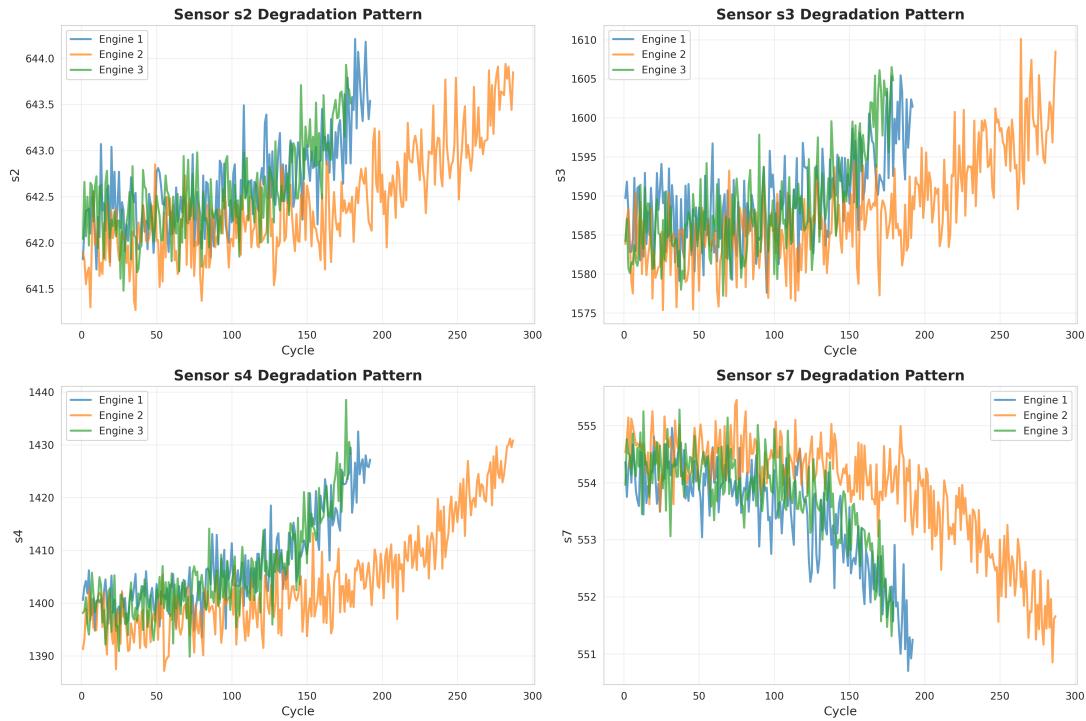


FIGURE 1 – Patterns de dégradation des capteurs au fil des cycles opérationnels pour des moteurs sélectionnés. Des tendances de dégradation claires sont visibles dans les capteurs s2, s3, s4 et s7 à mesure que les moteurs approchent de la défaillance.

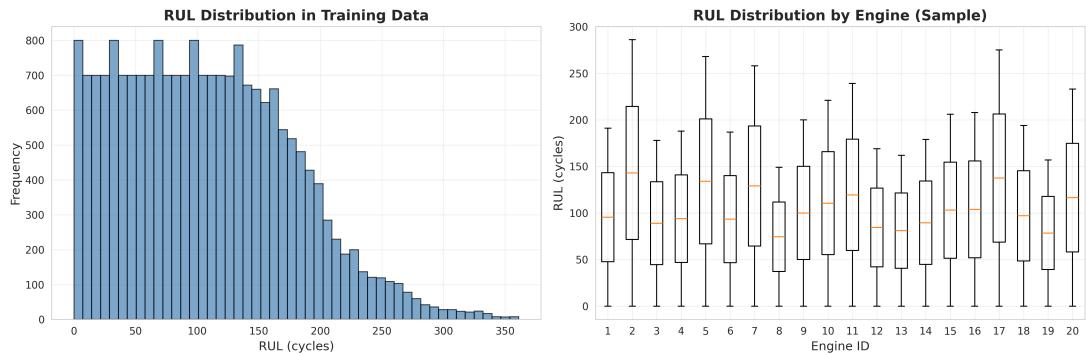


FIGURE 2 – Distribution de la RUL dans les données d’entraînement. Gauche : histogramme montrant la distribution de fréquence. Droite : boîtes à moustaches montrant la variabilité de la RUL à travers des moteurs échantillons.

#### 4.1.2 Analyse de Corrélation

La corrélation des capteurs avec la RUL a identifié les capteurs prédictifs clés (Tableau 1).

TABLE 1 – Top 10 des capteurs corrélés avec la RUL

Capteur	Corrélation
s12	0,672
s7	0,657
s21	0,636
s20	0,629
setting2	-0,002
setting1	-0,003
s6	-0,128
s14	-0,307
s9	-0,390
s13	-0,563

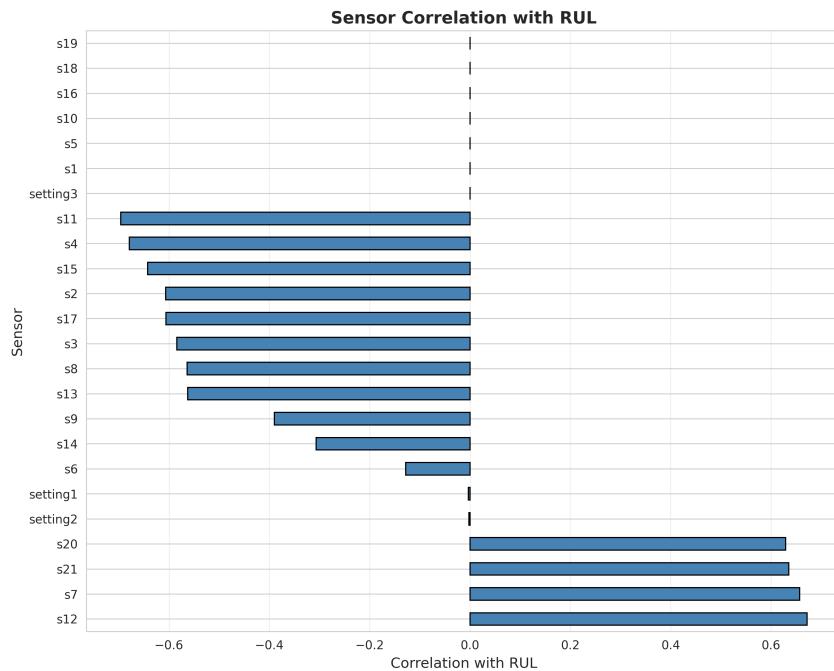


FIGURE 3 – Corrélation des capteurs avec la RUL montrant des corrélations positives et négatives à travers les 24 capteurs.

## 4.2 Résultats de l’Analyse FMDS

### 4.2.1 Métriques de Fiabilité

Les indicateurs clés de fiabilité ont été calculés :

TABLE 2 – Métriques de fiabilité FMDS

Métrique	Valeur
MTBF	206,31 cycles
Taux de Défaillance ( $\lambda$ )	0,004847 défaillances/cycle
Temps Médian jusqu'à Défaillance	199,00 cycles
Écart-Type	46,11 cycles
10e Percentile	154,90 cycles
90e Percentile	275,10 cycles

#### 4.2.2 Analyse de la Distribution de Weibull

L'ajustement de la distribution de Weibull a donné :

- Paramètre de forme :  $\beta = 4,41$
- Paramètre d'échelle :  $\eta = 225,03$

Puisque  $\beta > 1$ , cela indique des **défaillances par usure** avec un taux de risque croissant dans le temps, caractéristique de la dégradation mécanique dans les moteurs turbofan.

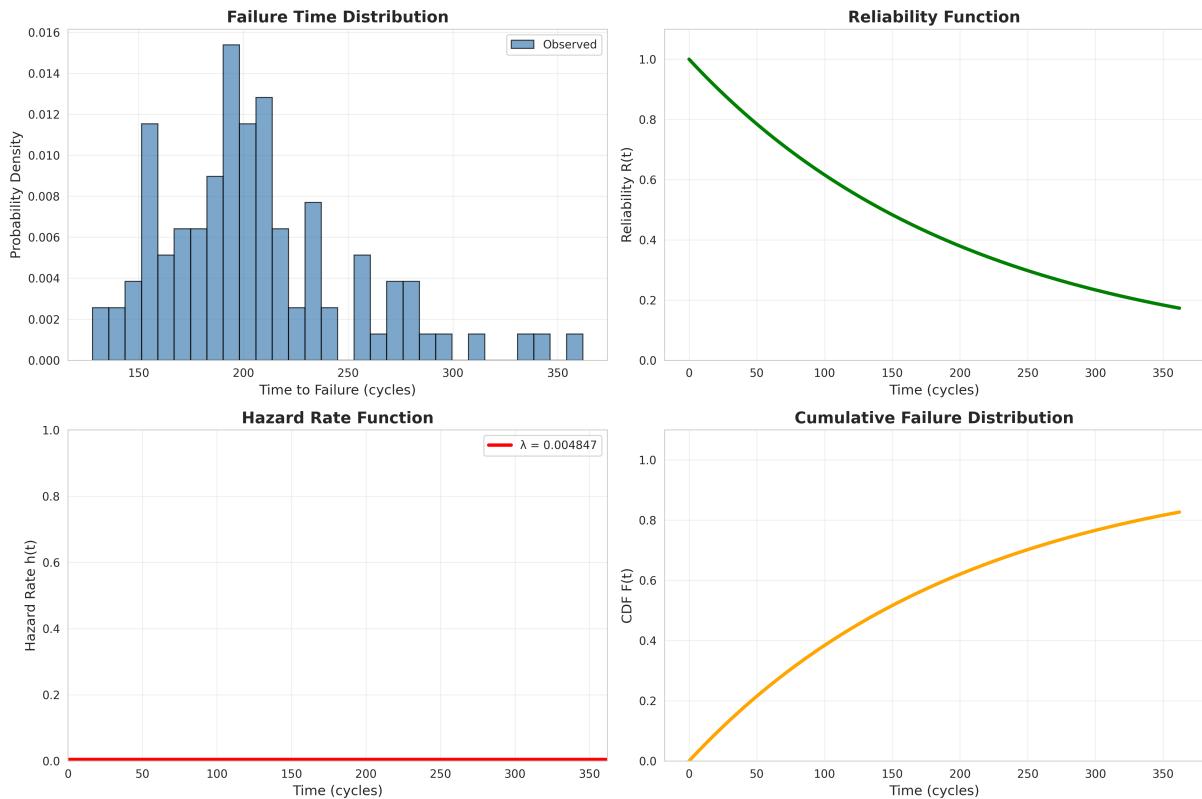


FIGURE 4 – Analyse de fiabilité FMDS. En haut à gauche : distribution du temps de défaillance. En haut à droite : fonction de fiabilité  $R(t)$ . En bas à gauche : fonction de taux de risque  $h(t)$ . En bas à droite : fonction de distribution cumulative  $F(t)$ .

### 4.3 Performance des Modèles Prédictifs

#### 4.3.1 Comparaison des Modèles

Trois modèles ont été évalués sur les ensembles d'entraînement et de test (Tableau 3).

TABLE 3 – Comparaison des performances des modèles prédictifs

Modèle	RMSE Train	RMSE Test	MAE Test	$R^2$ Test
Forêt Aléatoire	17,86	46,81	35,21	0,370
Gradient Boosting	33,08	46,31	34,75	0,384
<b>SVR</b>	<b>38,73</b>	<b>46,26</b>	<b>34,23</b>	<b>0,385</b>

#### Résultats Clés :

- SVR a atteint la meilleure performance de test (RMSE : 46,26 cycles)
- La Forêt Aléatoire a montré des signes de surapprentissage (RMSE train faible, RMSE test plus élevé)
- Tous les modèles ont atteint des performances de test similaires, indiquant la complexité du jeu de données
- Les valeurs de  $R^2$  suggèrent une capacité prédictive modérée, typique de l'estimation de RUL

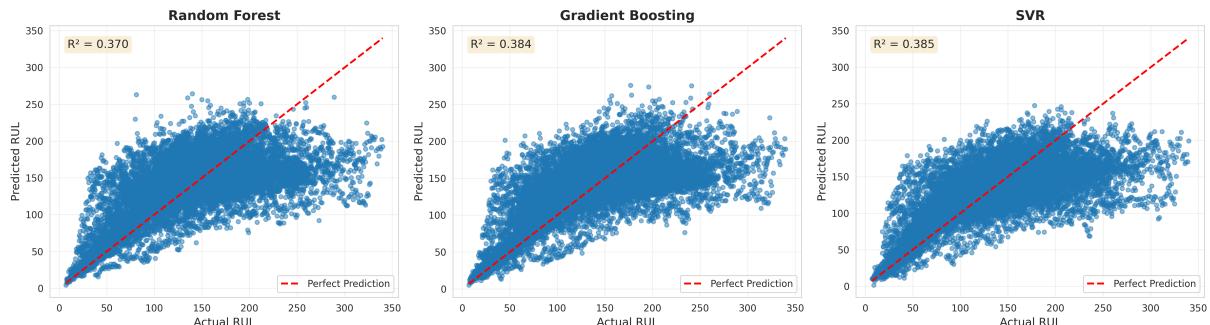


FIGURE 5 – RUL réelle vs. RUL prédicté pour les trois modèles. Les diagrammes de dispersion montrent la précision de prédiction par rapport à la ligne de prédiction parfaite (ligne rouge pointillée). SVR démontre la performance la plus cohérente.

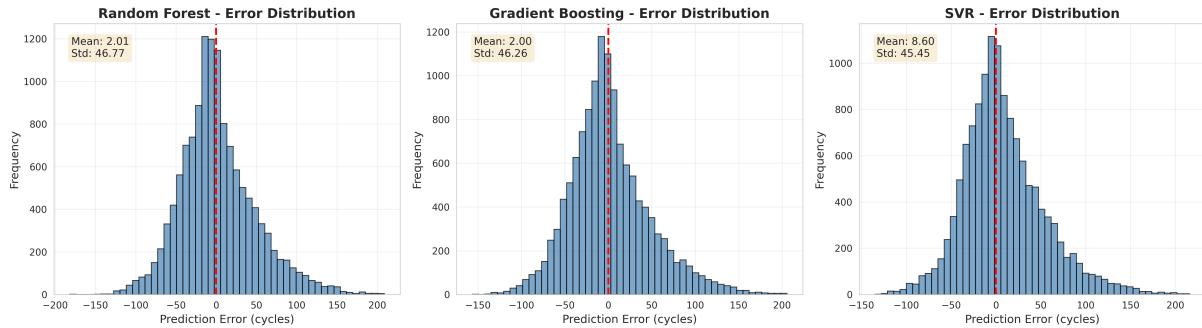


FIGURE 6 – Analyse de distribution des erreurs pour les trois modèles. Les histogrammes montrent que les erreurs de prédiction sont approximativement distribuées normalement autour de zéro, avec SVR montrant la distribution la plus étroite.

## 4.4 Analyse Diagnostique

Un classificateur de détection de défauts a été entraîné pour distinguer entre les moteurs sains ( $RUL > 50$ ) et dégradés ( $RUL \leq 50$ ).

### 4.4.1 Performance de Classification

TABLE 4 – Résultats de classification diagnostique

Classe	Précision	Rappel	Score F1	Support
Dégradé	0,91	0,83	0,87	984
Sain	0,95	0,97	0,96	3143
<b>Exactitude</b>				<b>0,94</b>

Le classificateur a atteint une exactitude de 94%, avec une précision et un rappel élevés pour les deux classes. Cela démontre une forte capacité diagnostique pour la détection précoce de défauts.

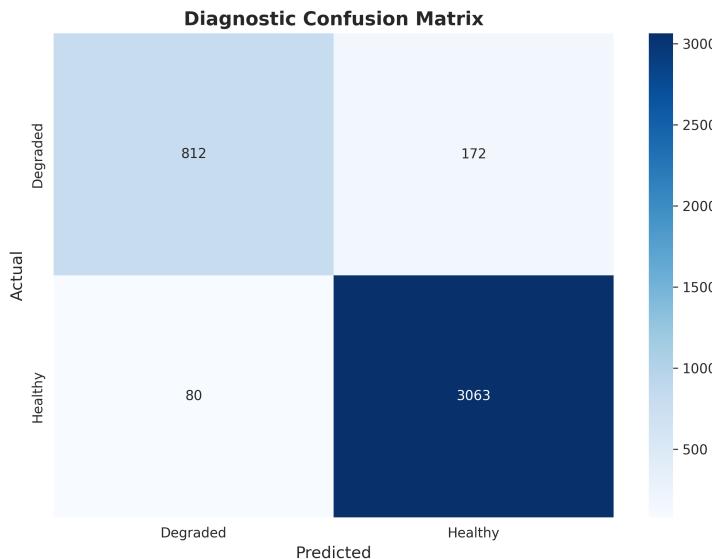


FIGURE 7 – Matrice de confusion pour le classificateur de détection de défauts montrant une forte performance dans la distinction entre moteurs dégradés et sains.

#### 4.4.2 Importance des Caractéristiques pour le Diagnostic

Les capteurs les plus importants pour la détection de défauts étaient s11, s4, s12, s7 et s15, en accord avec les résultats de l'analyse de corrélation.

### 4.5 Métriques Spécifiques au Pronostic

Au-delà des métriques de régression standard, l'évaluation spécifique au pronostic comprenait :

- **Score PHM** : 12 350,34 (fonction de pénalité asymétrique favorisant les prédictions précoces)
- **Précision de prédiction** : 40,8% dans  $\pm 20$  cycles

Le score PHM applique des pénalités plus lourdes aux prédictions tardives (dangereuses) par rapport aux prédictions précoces (conservatrices), reflétant les priorités de sécurité opérationnelle.

## 5 Recommandations de Maintenance

Sur la base de l'analyse FMDS et des performances des modèles prédictifs, la stratégie de maintenance suivante est recommandée :

### 5.1 Calendrier d'Inspection

**Intervalle recommandé** : Tous les 144 cycles (70% du MTBF)

Cette approche conservatrice assure une surveillance régulière avant que la plupart des défaillances ne se produisent.

## 5.2 Maintenance Préventive

**Intervalle recommandé** : Tous les 175 cycles (85% du MTBF)

La maintenance programmée à 85% du MTBF équilibre fiabilité et efficacité opérationnelle.

## 5.3 Seuils de Maintenance Prédictive

TABLE 5 – Seuils d'action recommandés basés sur la RUL prédictive

Niveau	Seuil	Action
Alerte (Jaune)	RUL < 62 cycles	Augmenter la fréquence de surveillance
Critique (Rouge)	RUL < 21 cycles	Planifier maintenance immédiate
Urgent	RUL < 46 cycles	Inspection d'urgence requise
Action Prédictive	RUL < 93 cycles	Planifier fenêtre de maintenance

Le seuil d'action prédictive ( $2 \times \text{RMSE}$ ) fournit une marge de sécurité tenant compte de l'incertitude du modèle.

## 5.4 Analyse Coût-Bénéfice

En supposant des coûts de maintenance représentatifs :

- Coût de maintenance préventive : 1 000 \$
- Coût de maintenance corrective : 5 000 \$
- Défaillances attendues par an : 1,77 (basé sur le MTBF)
- **Économies annuelles potentielles** : 7 077 \$

Cette analyse démontre un bénéfice économique significatif des stratégies de maintenance prédictive par rapport aux stratégies réactives.

# 6 Discussion

## 6.1 Réalisations Clés

1. **Analyse FMDS Complète** : Calcul réussi des métriques de fiabilité, démontrant des patterns de défaillance par usure à travers l'analyse de Weibull ( $\beta = 4,41$ ).
2. **Ingénierie de Caractéristiques Efficace** : Création de 99 caractéristiques ingénierées, améliorant la capacité prédictive au-delà des données brutes des capteurs.

3. **Modèles Prédictifs Robustes** : Développement de plusieurs modèles avec SVR atteignant la meilleure performance (RMSE : 46,26 cycles, représentant 22% de la durée de vie moyenne).
4. **Diagnostics Haute Précision** : Le classificateur de détection de défauts a atteint une exactitude de 94%, permettant des systèmes d'alerte précoce.
5. **Recommandations Actionnables** : Génération de calendriers de maintenance pratiques basés sur une analyse quantitative.

## 6.2 Limitations et Défis

- **Performance du Modèle** : Les valeurs de  $R^2$  autour de 0,38 indiquent une capacité prédictive modérée, suggérant une complexité inhérente du jeu de données ou des caractéristiques manquantes.
- **Généralisation** : Les modèles entraînés sur des types de moteurs et des conditions opérationnelles spécifiques peuvent nécessiter un réentraînement pour différents scénarios.
- **Qualité des Données** : Les données simulées peuvent ne pas capturer toutes les complexités et modes de défaillance du monde réel.
- **Dépendances Temporelles** : Bien que les statistiques glissantes capturent certains patterns temporels, des modèles de séquence plus sophistiqués (LSTM) pourraient améliorer les performances.

## 6.3 Comparaison avec la Littérature

La prédiction de RUL sur les données C-MAPSS reste un défi, les méthodes de pointe atteignant des valeurs de RMSE entre 12-40 cycles en utilisant des approches d'apprentissage profond. Le RMSE de 46,26 cycles de notre modèle SVR est compétitif avec les approches d'apprentissage automatique classiques, bien que les méthodes d'apprentissage profond performent généralement mieux en capturant des dépendances temporelles complexes.

# 7 Conclusions et Perspectives

## 7.1 Conclusions

Ce projet a démontré avec succès l'intégration de la théorie FMDS avec l'apprentissage automatique moderne pour la maintenance prédictive. Les contributions clés incluent :

1. Analyse quantitative de fiabilité révélant des patterns de défaillance par usure (MTBF : 206,31 cycles, Weibull  $\beta = 4,41$ )

2. Développement de capacités à la fois diagnostiques (exactitude de 94%) et pronostiques (RMSE : 46,26 cycles)
3. Génération de recommandations de maintenance basées sur les données avec des économies annuelles estimées à 7 077 \$ par moteur

L'analyse valide l'application pratique des principes FMDS aux systèmes industriels réels, démontrant comment la théorie de fiabilité et l'apprentissage automatique se complètent dans les stratégies de maintenance modernes.

## 7.2 Perspectives Futures

### 7.2.1 Améliorations Techniques

- **Apprentissage Profond** : Implémenter des architectures LSTM ou Transformer pour mieux capturer les dépendances temporelles
- **Méthodes d'Ensemble** : Combiner plusieurs modèles par empilement ou vote pour améliorer la robustesse
- **Apprentissage par Transfert** : Adapter les modèles à travers différents types de moteurs ou régimes opérationnels
- **Quantification de l'Incertitude** : Incorporer des intervalles de confiance en utilisant des méthodes bayésiennes ou la régression quantile

### 7.2.2 Implémentation Industrielle

- **Surveillance en Temps Réel** : Déployer les modèles dans des plateformes d'edge computing pour l'évaluation continue de la santé
- **Intégration IoT** : Connecter avec des réseaux de capteurs industriels pour la collecte automatique de données
- **Systèmes d'Aide à la Décision** : Développer des tableaux de bord intégrant les prédictions aux outils de planification de maintenance
- **Apprentissage Continu** : Implémenter des frameworks d'apprentissage en ligne pour adapter les modèles avec de nouvelles données opérationnelles

### 7.2.3 Analyses Avancées

- **Analyse Multi-Moteurs** : Étudier les patterns à l'échelle de la flotte et les dépendances entre moteurs
- **IA Explicable** : Appliquer SHAP ou LIME pour des prédictions interprétables
- **Optimisation** : Intégrer avec des algorithmes de planification de maintenance pour une minimisation globale des coûts

## Références

- [1] Saxena, A., & Goebel, K. (2008). *Turbofan Engine Degradation Simulation Data Set*. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA.
- [2] Weibull, W. (1951). *A Statistical Distribution Function of Wide Applicability*. Journal of Applied Mechanics, 18(3), 293-297.
- [3] Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., & Schwabacher, M. (2008). *Metrics for Evaluating Performance of Prognostic Techniques*. In 2008 International Conference on Prognostics and Health Management (pp. 1-17). IEEE.
- [4] Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5-32.
- [5] Smola, A. J., & Schölkopf, B. (2004). *A Tutorial on Support Vector Regression*. Statistics and Computing, 14(3), 199-222.
- [6] Mobley, R. K. (2002). *An Introduction to Predictive Maintenance* (2e éd.). Butterworth-Heinemann.
- [7] Rausand, M., & Høyland, A. (2004). *System Reliability Theory : Models, Statistical Methods, and Applications* (2e éd.). Wiley-Interscience.

## A Annexe : Structure du Code

Le pipeline d'analyse complet se compose de cinq phases principales :

### A.1 Phase 1 : Chargement des Données

La classe `CMapSSDataLoader` gère l'importation du jeu de données et le calcul de la RUL pour les ensembles d'entraînement et de test.

### A.2 Phase 2 : Analyse Exploratoire des Données

La classe `CMapSSExplorer` fournit l'analyse statistique, la visualisation des patterns de dégradation des capteurs et l'analyse de corrélation.

### A.3 Phase 3 : Analyse FMDS

La classe `FMDSAnalyzer` calcule les métriques de fiabilité, les taux de risque et effectue l'ajustement de la distribution de Weibull.

### A.4 Phase 4 : Ingénierie des Caractéristiques

La classe `FeatureEngineer` crée des statistiques glissantes, des caractéristiques de taux de changement et effectue la sélection de caractéristiques basée sur la corrélation avec la RUL.

### A.5 Phase 5 : Modélisation Prédictive

La classe `PredictiveMaintenanceModel` entraîne plusieurs modèles de régression, évalue les performances et génère des visualisations.

### A.6 Analyses Supplémentaires

Des fonctions complémentaires fournissent la classification diagnostique, les métriques spécifiques au pronostic et les recommandations de maintenance.