

Université Cadi Ayyad
École Nationale des Sciences Appliquées de Safi
Département Informatique, Réseaux et Télécommunications (IRT)

Smart manufacturing IoT, CyberSécurité

En Génie Aéronautique et Technologies de l'Espace

Maintenance Prédictive Aéronautique par IA



Réalisé par :

HAJAR BELKADI
IMANE OURRH
ASMA AIT ATTA

Encadré par :

Mr. Bentajer

I. Introduction et Contexte Industriel:

Dans l'industrie aéronautique moderne (Airbus, Boeing, Safran, GE), la maintenance représente une part majeure des coûts d'exploitation. Trois stratégies de maintenance existent : la maintenance corrective, la maintenance préventive et la maintenance prédictive. La maintenance corrective, qui consiste à réparer après la panne, est inacceptable en aéronautique pour des raisons de sécurité. La maintenance préventive, basée sur des intervalles fixes, est coûteuse car elle conduit souvent au remplacement de pièces encore fonctionnelles. La maintenance prédictive vise à anticiper les pannes grâce aux données issues des capteurs embarqués, afin d'intervenir au moment optimal, en réduisant les coûts tout en garantissant la sécurité.

II. Objectif du Projet:

L'objectif de ce projet est de développer un modèle d'intelligence artificielle capable d'estimer la durée de vie restante (Remaining Useful Life – RUL) d'un turboréacteur aéronautique. Cette estimation est réalisée à partir des mesures issues des capteurs embarqués, telles que la température, la pression et les vitesses de rotation. Le projet vise à transformer des données brutes en une information exploitable pour la prise de décision en maintenance.

III. Description des Données:

Les données utilisées proviennent du jeu de données NASA C-MAPSS, qui simule la dégradation progressive de turboréacteurs. Le fichier d'entraînement contient plusieurs moteurs suivis depuis le début de leur fonctionnement jusqu'à la panne. Chaque ligne correspond à un cycle de fonctionnement d'un moteur et contient 26 variables, incluant l'identifiant du moteur, le numéro du cycle, trois paramètres opérationnels : altitude, nombre de Mach, et position manette et 21 capteurs physiques mesurant l'état du moteur.

IV. Phase A – Analyse Exploratoire des Données (EDA):

a- Visualisation des capteurs :

Le jeu de données NASA C-MAPSS ne fournit pas d'information explicite sur la nature physique de chaque capteur. Les capteurs sont donc identifiés par des noms génériques (sensor_1 à sensor_21). Par conséquent, **l'analyse de la Phase 1 repose sur une observation qualitative des**

valeurs numériques, de leur ordre de grandeur, de leur variabilité et de leur évolution au cours des cycles de fonctionnement.

unit_nr	time_cycles	op_setting_1	op_setting_2	op_setting_3	sensor_1	sensor_2	sensor_3	sensor_4	sensor_5	...	sensor_12	sensor_13	sensor_14	sensor_15	
0	1	1	-0.0007	-0.0004	100.0	518.67	641.82	1589.70	1400.60	14.62	...	521.66	2388.02	8138.62	8.4195
1	1	2	0.0019	-0.0003	100.0	518.67	642.15	1591.82	1403.14	14.62	...	522.28	2388.07	8131.49	8.4318
2	1	3	-0.0043	0.0003	100.0	518.67	642.35	1587.99	1404.20	14.62	...	522.42	2388.03	8133.23	8.4178
3	1	4	0.0007	0.0000	100.0	518.67	642.35	1582.79	1401.87	14.62	...	522.86	2388.08	8133.83	8.3682
4	1	5	-0.0019	-0.0002	100.0	518.67	642.37	1582.85	1406.22	14.62	...	522.19	2388.04	8133.80	8.4294

sensor_15	sensor_16	sensor_17	sensor_18	sensor_19	sensor_20	sensor_21
8.4195	0.03	392	2388	100.0	39.06	23.4190
8.4318	0.03	392	2388	100.0	39.00	23.4236
8.4178	0.03	390	2388	100.0	38.95	23.3442
8.3682	0.03	392	2388	100.0	38.88	23.3739
8.4294	0.03	393	2388	100.0	38.90	23.4044

L'analyse des données montre que certains capteurs présentent des valeurs élevées et continues, avec de légères variations au fil des cycles (par exemple: sensor_2, sensor_3, sensor_4 et sensor_12). **Ce comportement est cohérent avec des grandeurs thermiques**, car l'usure progressive du moteur entraîne une augmentation des pertes internes, ce qui se traduit généralement par une élévation des températures nécessaires pour maintenir les performances.

D'autres capteurs présentent des valeurs caractéristiques de paramètres dynamiques, avec des variations faibles mais régulières (par exemple sensor_14, sensor_15 et sensor_17). **Ces variations peuvent être associées à des ajustements des vitesses de rotation ou des pressions internes du moteur**, qui s'adaptent progressivement afin de compenser la dégradation mécanique.

b- Identification des capteurs utiles :

À l'inverse, plusieurs capteurs affichent des valeurs constantes ou quasi constantes sur l'ensemble des cycles (notamment sensor_1, sensor_5, sensor_16, sensor_18 et sensor_19). L'absence totale de variation indique que ces capteurs ne contiennent aucune information sur l'évolution de l'état du moteur et peuvent être considérés comme **non informatifs** pour l'analyse de la dégradation.

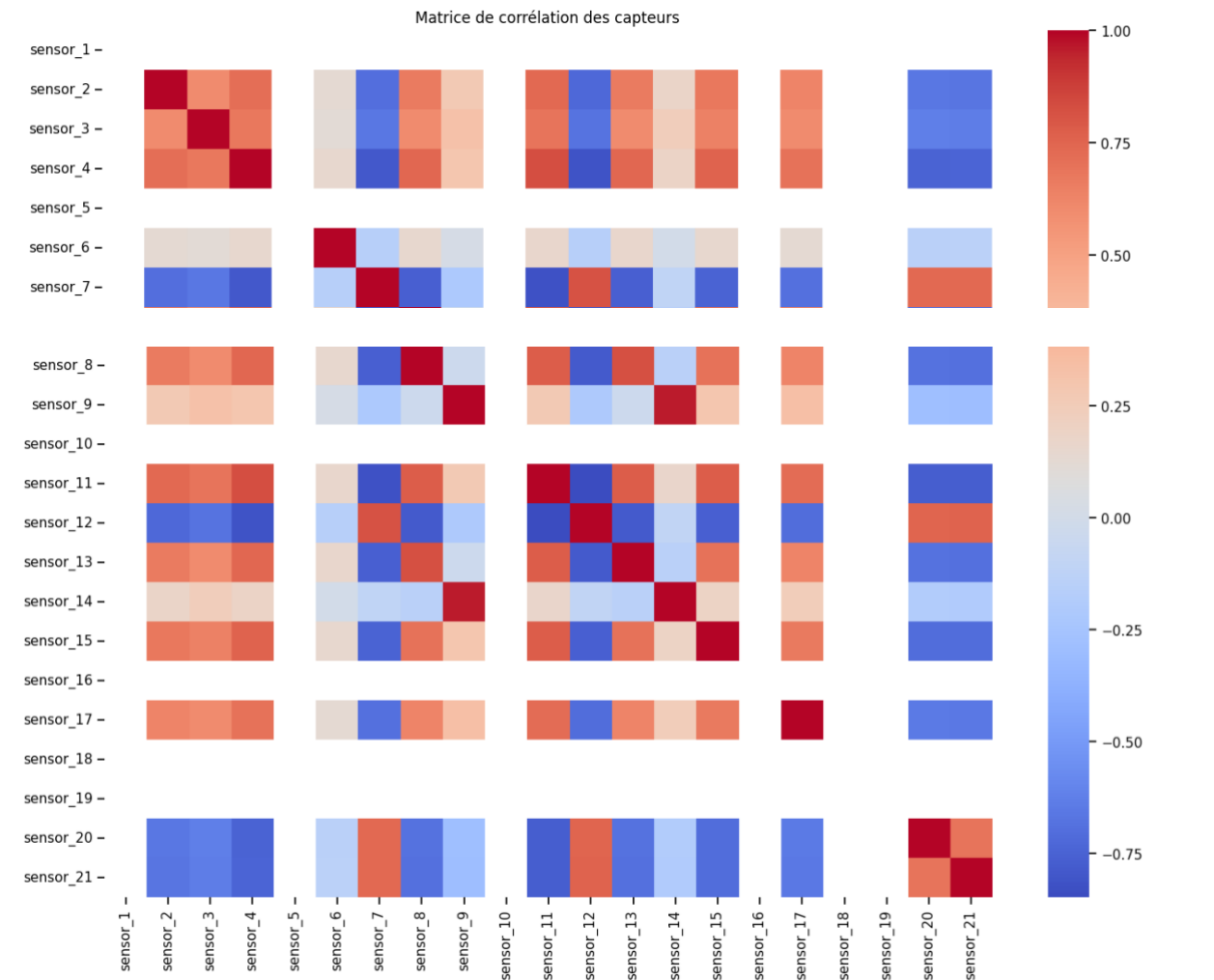
En résumé, la dégradation du moteur ne se manifeste pas par une variation brutale d'un seul capteur, mais par des dérives progressives et conjointes de plusieurs capteurs, observables uniquement lorsque l'on suit l'évolution temporelle d'un moteur donné.

C- Analyse les corrélations :

L'analyse des corrélations a pour objectif d'étudier les relations entre les capteurs et l'évolution de l'usure des moteurs. Elle permet d'identifier les capteurs dont les variations sont significativement liées au nombre de cycles ou à la durée de vie restante (RUL).

Les capteurs présentant une faible corrélation ou une variance nulle n'apportent pas d'information pertinente et peuvent être supprimés. Cette étape permet de réduire la

dimension des données, d'améliorer la stabilité du modèle et d'augmenter la précision des prédictions dans le cadre de la maintenance prédictive.



V. Phase B – Prétraitement des Données:

Le prétraitement constitue une étape essentielle dans tout projet de science des données. Les données brutes issues du fichier `train_FD001.txt` ne sont pas directement exploitables par un algorithme d'intelligence artificielle, car elles ne contiennent ni la variable cible à prédire, ni un format adapté à l'apprentissage. Cette phase a pour objectif de nettoyer, structurer et transformer les données afin de les rendre prêtes pour la modélisation.

a- Génération de la variable cible (RUL) :

Le fichier d'entraînement fourni ne contient pas explicitement la Durée de Vie Restante (RUL). Or, c'est la cible que l'IA doit apprendre à prédire. Nous l'avons donc calculée à partir de la logique suivante :

- Pour chaque moteur, le RUL à un cycle donné est égal au **nombre total de cycles de sa vie moins le cycle actuel**.

Mathématiquement, cela s'exprime par:

$$RUL = Cycle_{max} - Cycle_{actuel}$$

Par exemple, si un moteur tombe en panne au cycle 192, alors à son cycle 100, son RUL est de $192 - 100 = 92$ cycles.

Cette opération a été réalisée à l'aide de la fonction `groupby().transform('max')` en Python, qui attribue à chaque ligne le cycle maximum de son moteur, puis soustrait le cycle courant. Le résultat est une **colonne RUL décroissante**, allant de la durée totale de vie jusqu'à 0 (panne).

➤ Troncature du RUL (Clipping)

Une observation importante dans les études de maintenance prédictive est que **les moteurs ne présentent aucun signe de dégradation pendant la majorité de leur vie utile**. Si l'IA apprend à prédire des RUL très élevés (ex. : 300 cycles), elle perdra en précision **là où cela compte le plus : dans les derniers cycles avant la panne**.

Pour remédier à ce problème, nous avons appliqué une **troncature (clipping)** du RUL à **125 cycles**. Cela signifie que **tout RUL supérieur à 125 est remplacé par 125**. Ce choix permet au modèle de se concentrer sur la **phase critique de la dégradation**, où les capteurs commencent réellement à dévier de leur comportement nominal.

b- Tableau comparatif des performances des modèles testés :

L'analyse exploratoire (Phase A) a révélé que certains capteurs sont constants sur l'ensemble des cycles (ex. : `sensor_1`, `sensor_5`, `sensor_16`, `sensor_18`, `sensor_19`). Ces capteurs, n'affichant aucune variation, n'apportent aucune information sur l'état de dégradation du moteur. Leur présence pourrait même nuire à l'apprentissage en introduisant du bruit inutile. Nous les avons donc supprimés du jeu de données, ainsi que le paramètre opérationnel `op_setting_3`, souvent constant dans ce sous-ensemble.

c- [Normalisation des capteurs (Min-Max Scaling):

Les capteurs mesurent des grandeurs physiques très différentes (température, pression, vitesse), avec des échelles numériques disparates (ex. : température ≈ 600 , pression ≈ 14). Si ces valeurs sont transmises telles quelles à l'IA, celle-ci accordera plus d'importance aux capteurs ayant des valeurs numériques plus élevées, ce qui introduirait un biais artificiel.

Pour résoudre ce problème, nous avons appliqué une **normalisation Min-Max**, qui transforme toutes les valeurs pour qu'elles soient comprises entre **0 et 1**, selon la formule :

$$X_{\text{normalisé}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Cette transformation a été appliquée uniquement aux capteurs utiles, à l'exclusion des colonnes `unit_nr`, `time_cycles` et `RUL`, qui ne doivent pas être normalisées. Le même scaler, entraîné sur les données d'entraînement, sera réutilisé sans réapprentissage sur les données de test, afin de garantir une cohérence entre les deux jeux de données.

VI. Phase C – Modélisation (Machine Learning):

a- Comparer deux modèles : Régression Linéaire et Random Forest Regressor :

Deux approches de modélisation ont été évaluées pour la prédiction de la durée de vie restante (RUL) des moteurs. La première est une régression linéaire, utilisée comme modèle de référence (baseline), et la seconde est un modèle avancé de type Random Forest Regressor.

Les performances ont été mesurées à l'aide de la métrique RMSE (Root Mean Square Error). Le modèle de régression linéaire obtient un RMSE de 21,68, tandis que le modèle Random Forest atteint un RMSE plus faible de 18,72. Cette diminution de l'erreur indique que le modèle Random Forest fournit des prédictions plus précises de la durée de vie restante des moteurs.

```
*** RMSE Baseline (Linéaire) : 21.68
    RMSE Avancé (Random Forest) : 18.72
```

Ainsi, bien que la régression linéaire constitue une référence simple et interprétable, **le modèle Random Forest apparaît comme la méthode la plus appropriée pour ce problème de maintenance prédictive**, en raison de sa meilleure performance et de sa capacité à capturer la complexité des mécanismes de dégradation.

VII. Phase D – Évaluation des Performances:

a- Utilisation du RMSE :

Après l'étape de modélisation et d'optimisation, le modèle d'intelligence artificielle a été évalué sur le jeu de données de test. La performance finale obtenue est un **RMSE égal à 17,75**, ce qui représente l'erreur moyenne entre la durée de vie restante réelle (RUL réel) et celle prédite par le modèle.

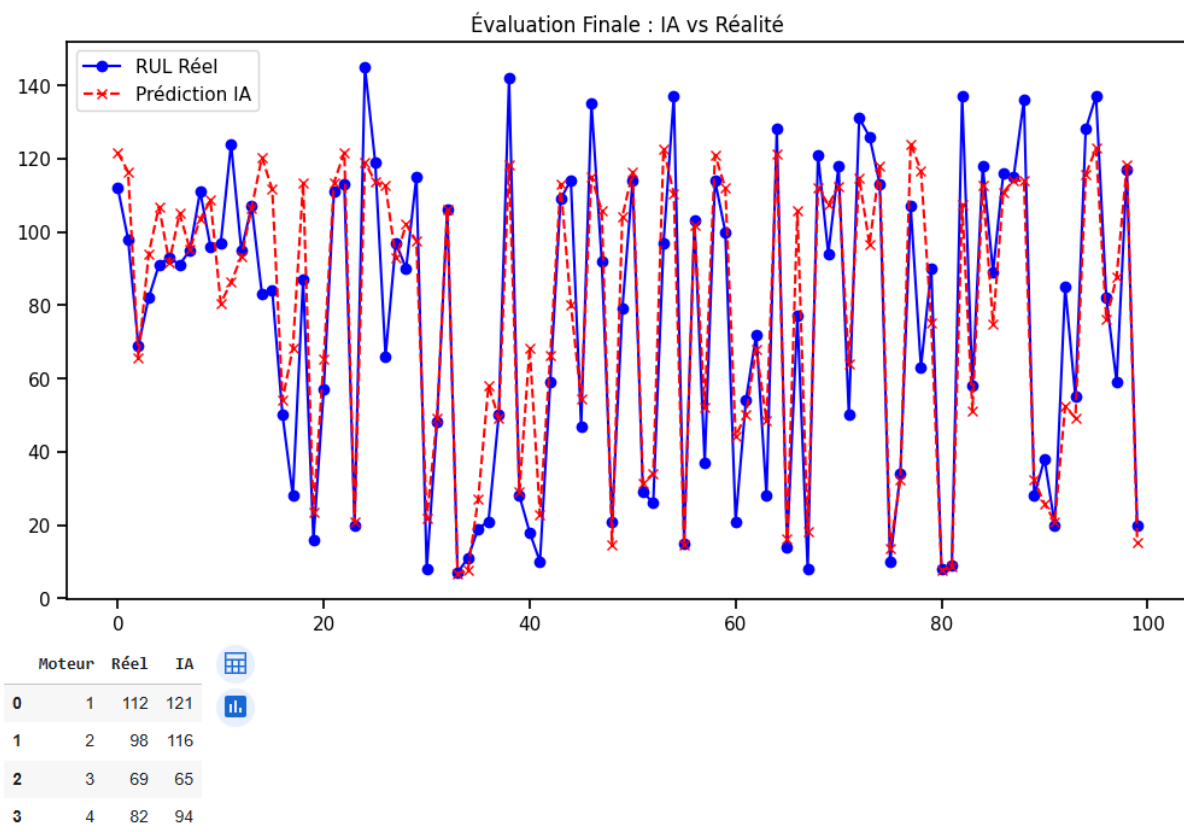
```
=====
RÉSULTAT FINAL RMSE SUR TEST : 17.75
=====
```

Cette valeur de RMSE est inférieure à celles obtenues lors des phases précédentes (modèle linéaire et Random Forest de base), ce qui montre une **amélioration progressive de la précision des prédictions**. Un RMSE plus faible signifie que les prédictions du modèle sont globalement

plus proches des valeurs réelles de RUL, ce qui est essentiel dans un contexte de maintenance prédictive.

b -Tableau comparatif des performances des modèles testés:

Le tableau comparatif entre les valeurs réelles et les valeurs prédites par l'IA pour plusieurs moteurs confirme cette observation. Par exemple, pour le moteur 3, la valeur réelle du RUL est de 69 cycles, tandis que le modèle prédit 65 cycles, ce qui représente une erreur faible. De manière générale, les écarts observés restent raisonnables et cohérents avec la valeur globale du RMSE.



Les différences restantes entre les valeurs réelles et prédites peuvent s'expliquer par la complexité du phénomène de dégradation des moteurs, qui dépend de multiples capteurs, de relations non linéaires et de comportements variables d'un moteur à un autre. Néanmoins, le modèle parvient à capter la tendance globale de l'usure et à fournir des estimations fiables du RUL.

Ainsi, avec un RMSE final de 17,75, le modèle retenu peut être considéré comme **performant et adapté à une application de maintenance prédictive aéronautique**, car il permet d'anticiper la défaillance des moteurs avec une précision satisfaisante.

VIII. Conclusion:

Ce projet montre que **l'intelligence artificielle peut être utilisée efficacement pour la maintenance prédictive en aéronautique**. Les résultats obtenus sont prometteurs, mais l'intégration dans un système réel nécessite une validation approfondie, des données réelles et une prise en compte des contraintes de sécurité. Les principaux risques concernent **la qualité des données et la généralisation du modèle**. En effet, les données utilisées sont simulées et ne reflètent pas toujours les imperfections du monde réel, telles que **les capteurs défectueux, le bruit de mesure ou les conditions d'usage extrêmes**. De plus, le modèle a été entraîné et **évalué uniquement sur un scénario spécifique (FD001)**, ce qui limite sa capacité à bien prédire sur d'autres types de moteurs ou d'autres profils de vol ; c'est ce qu'on appelle un manque de généralisation. Ces deux risques peuvent conduire à des erreurs de prédiction imprévues, ce qui est inacceptable dans un contexte aéronautique où la sécurité est primordiale.