

UNIVERSITE KASDI MERBAH OUARGLA

Faculté des Sciences Appliquées

Département de Génie Electrique



Mémoire MAS-
TER ACADEMIQUE

Domaine : Sciences et technologies Fi-

lière : Génie électrique

Spécialité : Machines électrique

Présenté par :

Ghaia Abdeslem

Della Abdelmoula

Thème :

**Modélisation Prédictive de la Durée de Vie Résiduelle (RUL) des
Systèmes de Stockage Electrochimique Lithium-ion via des Ap-
proches d'Apprentissage Automatique**

Soumis au jury composé de :

Larouci.B	Prof	Président	UKM Ouargla
Rezoug Med Redha	MCA	Examineur	UKM Ouargla
Djarah Djalal	MCA	Encadreur/rapporteur	UKM Ouargla

Année universitaire 2024/2025

Remerciement

Avant tout, nous exprimons notre profonde gratitude à Allah, source de toute connaissance et sagesse, qui nous a guidés tout au long de ce parcours.

Nous tenons à adresser nos plus sincères remerciements à nos chers parents, pour leurs sacrifices inestimables, leur soutien inconditionnel et leurs encouragements constants tout au long de notre formation.

Nos remerciements les plus chaleureux vont à notre encadrant, **Dr. Djalal Djarah**, pour son accompagnement exceptionnel, sa disponibilité constante, ses conseils précieux et l'ambiance de confiance et de bienveillance qu'il a su instaurer durant la préparation de ce mémoire.

Nous exprimons également notre reconnaissance aux membres du jury pour l'honneur qu'ils nous font en acceptant d'évaluer ce travail, ainsi qu'à tous les enseignants qui ont contribué, de près ou de loin, à notre formation.

Enfin, nous remercions toutes les personnes qui nous ont soutenus, encouragés ou aidés, de quelque manière que ce soit, dans l'élaboration de ce mémoire.

DÉDICACE

À mes chers parents,

En reconnaissance de vos efforts, de votre soutien et de vos sacrifices tout au long de mon parcours. Que Dieu vous récompense généreusement.

À mes frères et sœurs,

Merci pour votre soutien et vos encouragements constants.

À mes amis et collègues,

Je vous exprime toute ma gratitude pour votre présence et votre aide.

À mes enseignants,

Merci pour votre encadrement et la qualité de l'enseignement reçu.

À mon encadrant,

Je vous adresse mes sincères remerciements pour vos conseils et votre accompagnement tout au long de ce travail.

ملخص

يتناول هذا العمل البحثي تطوير نظام تنبؤي يهدف إلى تقدير العمر المتبقي الممكن (RUL) لبطاريات الليثيوم-أيون، بالاعتماد على منصة تحليل البيانات. Orange تبدأ المنهجية المعتمدة بجمع بيانات تتعلق بأداء البطاريات، يليها مرحلة تمهيدية تشمل تنظيف البيانات، واختيار الخصائص الملائمة، وتصميم المتغير RUL. يعتمد النظام على تطبيق نماذج التعلم الآلي، منها الغابات العشوائية (Random Forest) والشبكات العصبية (Neural Networks). تُدرَّب هذه النماذج وتُختبر باستخدام مجموعة بيانات مُقسمة مسبقًا إلى بيانات تدريب واختبار. ويتم تقييم أداء النماذج استنادًا إلى دقة التنبؤات وباستخدام مقاييس كمية مثل الجذر التربيعي لمتوسط الخطأ (RMSE). أدَّت الاستفادة من منصة Orange إلى تمكين تصور بصري واضح للنتائج، مما سهَّل مقارنة منهجية صارمة بين النماذج وتحليل كفاءتها بشكل فعال.

الكلمات المفتاحية

بطاريات الليثيوم-أيون العمر المتبقي الممكن (RUL) التعلم الآلي (Machine Learning) الغابات العشوائية (Random Forest) الشبكات العصبية (Neural Networks) منصة Orange لتحليل البيانات (Orange Data Mining)

ABSTRACT

This research work presents the development of a predictive system designed to estimate the Remaining Useful Life (RUL) of lithium-ion batteries, leveraging the Orange data analysis platform. The adopted methodology begins with the collection of data related to battery performance, followed by a preprocessing phase comprising data cleaning, relevant feature selection, and target variable engineering (RUL).

The system relies on the implementation of machine learning models, including Random Forest and Neural Networks. These models are trained and validated on a dataset previously partitioned into training and test sets. Model performance evaluation is carried out based on the accuracy of their predictions and quantitative metrics such as the Root Mean Squared Error (RMSE). The use of the Orange platform allowed for intuitive visualization of the results, thereby facilitating rigorous model comparison and the efficient analysis of their respective performances.

KEYWORDS

Lithium-ion Batteries, Remaining Useful Life (RUL), Machine Learning (ML), Random Forest, Neural Networks, Orange Platform (Orange Data Mining)

Sommaire

Remerciement	II
DÉDICACE	III
ملخص	V
Introduction générale.....	2
 Chapitre 1 : Les Batteries Lithium-ion : Principes et Importance de la Durée de Vie Utile Restante (RUL)	
1.1 Introduction	5
1.2 Historique et évolution des batteries lithium-ion	5
1.2.1 Origines des batteries	5
1.2.2 Développement des batteries lithium-ion	5
1.2.3 Émergence des batteries 18650	6
1.3 Caractéristiques électrotechniques des batteries 18650	6
1.3.1 Structure physique	6
1.3.2 Chimies principales	7
1.3.3 Paramètres clés	8
1.3.4 Comparaison avec d'autres formats	8
1.4 Principes de fonctionnement	8
1.4.1 Réactions électrochimiques	8
1.4.2 Facteurs de dégradation	9
1.4.3 Illustration	9
1.5 Applications des batteries 18650	10
1.5.1 Véhicules électriques et hybrides	10
1.5.2 Stockage d'énergie renouvelable	10
1.5.3 Électronique portable	11

Sommaire

1.6 Importance de la durée de vie utile restante (RUL)	11
1.6.1 Définitions	11
1.6.2 Méthodes d'estimation	11
1.6.3 Enjeux	11
1.7 Présentation de la base de données B0005	12
1.7.1 Origine	12
1.7.2 Caractéristiques	12
1.7.3 Rôle dans l'étude	12
1.8 Conclusion	12
 Chapitre 2 : Application du Machine Learning dans le Domaine des Batte- ries	
2.1 Introduction.....	15
2.2 Problématiques dans le Domaine des Batteries.....	15
2.3 Forêts Aléatoires (Random Forest).....	16
2.3.1 Présentation du Modèle.....	16
2.3.2 Fonctionnement des forêts aléatoires.....	16
2.3.3 Avantages des forêts aléatoires.....	17
2.3.4 Applications dans le Domaine des Batteries.....	17
2.3.5 Avantages et Limites.....	18
2.4 Réseaux de Neurones.....	19
2.4.1 Présentation du Modèle.....	19

Sommaire

2.4.2 Composants des réseaux de neurones artificiels.....	19
2.4.3 Applications dans le Domaine des Batteries.....	20
2.4.4 Avantages et Limites.....	20
2.5 Comparaison des Modèles.....	21
2.6 Conclusion.....	22

Chapitre 3 : Méthodologie, Résultats et Analyses

3.1 Introduction.....	24
3.1.1 Base de données.....	25
3.2 Workflow widget.....	25
3.3 Importation de Données (Fichier).....	26
3.3.1 Fonctionnalités et Rôle.....	26
3.4 Widget d'Exécution de Script Python.....	28
3.4.1 Fonctionnalités et Rationale.....	29
3.4.2 Description du Script Utilisé.....	29
3.5 Widget de Visualisation de Table de Données.....	30
3.5.1 Fonctionnalités et Rôle.....	30
3.6 Observations Issues du Widget.....	32
3.6.1 Élément RUL.....	34
3.6.2 Importance de l'Analyse des Corrélations.....	35
3.7 Widget de Prétraitement des Données (Preprocess).....	36
3.7.1 Fonctionnalités et Rôle.....	36
3.8 Widget de Sélection de Colonnes (Select Columns).....	37
3.8.1 Configuration et Rôle.....	37

Sommaire

3.9 Widget de Forêt Aléatoire (Random Forest).....	38
3.10 Widget Réseau Neuronal (Neural Network).....	39
3.11 Widget Test et Score (Test and Score).....	40
3.12 Widget Prédications (Predictions).....	41
3.12.1 Fonction.....	41
3.12.2 Valeurs réelles et prédites.....	43
3.13 Visualisation des Résultats.....	44
3.13.1 Comparaison graphique des prédictions.....	44
3.13.2 Comparaison des erreurs.....	44
3.14 Conclusion.....	45
Conclusion Générale.....	48

Liste des figures

Figure1. 1 : Coupe transversale d'une batterie lithium-ion 18650	6
Figure 1.2 : Schéma du fonctionnement d'une batterie lithium-ion (charge/décharge)	9
Figure 1.3 : Graphique de dégradation typique de la capacité d'une batterie lithium-ion en fonction des cycles	10
Figure 3.1: Organigramme du Processus de Prédiction de la RUL.....	24
Figure3.2 : Workflow d'Apprentissage Automatique pour la Prédiction de la RUL des Batteries avec Comparaison de Modèles dans Orange.....	26
Figure3.3: Une image montrant le fichier de données.....	27
Figure3.4: Image montrant la sortie de l'outil Fichier après l'ajout de données.....	27
Figure3.5 : les sorties et les entrées de l'outil Python Script après le calcul du RUL.....	28
Figure 3.6: Script Python pour la Génération de la Variable Cible RUL.....	28
Figure3.7 : Génération de la RUL et Corrélations.....	33
Figure 3.8 : Transformation et Sélection des Données.....,,	36
Figure3. 9 : Flux de Données à Travers le Widget "Sélectionner les Colonnes".....	37
Figure 3. 10 : les paramètres du modèle Random Forest.....	38
Figure3.11 : Flux de Données du Widget "Random Forest".....	39
Figure 3. 12 : Hyperparamètres du Modèle de Réseau Neuronal.....	39
Figure3. 13 : Flux de Données du Widget "Réseau Neuronal".....	40
Figure3.14 : Flux de Données du Widget "Test and Score".....	41

Liste des figures

Figure3.15 : Flux d'Analyse des Prédiction.....	42
Figure3.16 : Comparaison des prédiction (RUL Réelle - RUL Prédite).....	44
Figure3.17 : Distribution des erreurs de prédiction (RUL Réelle - RUL Prédite).....	45

Liste des tableaux

Tableau 1.1 : Comparaison des principales chimies de cathodes pour batteries lithium-ion ..	7
Tableau1. 2: Comparaison entre le format 18650 et d'autres formats de cellules.....	8
Tableau 2.1 : Comparaison des Modèles Random Forest et Réseaux de Neurones	21
Tableau3.1: Tableau montrant les données de la batterie après traitement et ajout de la colonne RUL.....	31
Tableau3.2 : Rapport de Corrélation des Caractéristiques Avant l'Ajout de la Variable RUL.....	33
Tableau3.3 : Rapport de Corrélation des Caractéristiques Après l'Ajout de la Variable RUL.....	33
Tableau3.4 : Tableau présentant les valeurs réelles et attendues du RUL résultant des modèles de prévision au sein d'Orange.....	43

INTRODUCTION GENERALE

INTRODUCTION GENERALE

Dans un contexte mondial marqué par les défis majeurs liés à la sécurité énergétique et au changement climatique, la transition vers des sources d'énergie renouvelable telles que le solaire et l'éolien est devenue une priorité. Ces énergies propres et durables offrent une alternative prometteuse aux combustibles fossiles. Toutefois, leur nature intermittente impose la nécessité cruciale de disposer de systèmes de stockage d'énergie performants pour garantir la continuité de l'alimentation électrique et la stabilité des réseaux. Parmi ces systèmes, les batteries, et plus particulièrement les batteries lithium-ion, jouent un rôle central grâce à leurs caractéristiques remarquables : haute densité énergétique, légèreté et longue durée de vie.

Avec l'utilisation exponentielle de ces batteries dans des applications critiques telles que les véhicules électriques, les systèmes de stockage stationnaires ou l'électronique portable, il devient impératif de surveiller précisément leur état de santé (State of Health - SOH) et d'estimer leur durée de vie restante (Remaining Useful Life – RUL). Pourquoi cette estimation est-elle si cruciale ? Une estimation fiable du RUL permet d'anticiper les défaillances, d'optimiser les cycles de charge et de décharge, de planifier la maintenance ou le remplacement de manière proactive, garantissant ainsi la sécurité des systèmes, la minimisation des coûts opérationnels et l'amélioration de la fiabilité globale. Traditionnellement, cette tâche reposait sur des modèles physiques complexes basés sur des équations électrochimiques, nécessitant souvent une connaissance approfondie des paramètres internes de la batterie et un effort de modélisation conséquent. Face aux limites des approches purement physiques, les approches basées sur les données, issues du domaine de l'Intelligence Artificielle et du Machine Learning, apparaissent comme des alternatives particulièrement prometteuses. En exploitant de vastes quantités de données opérationnelles réelles (telles que la tension, le courant, la température, la capacité mesurée au cours du temps...), il devient possible de découvrir des relations complexes et non linéaires et de développer des modèles prédictifs robustes pour le suivi de l'état de santé et l'estimation du RUL des batteries sans nécessiter une modélisation physique préalable détaillée.

Ce travail s'inscrit dans cette optique et vise à développer un modèle prédictif précis pour l'estimation du RUL des batteries lithium-ion. Nous explorons et combinons plusieurs algorithmes de Machine Learning performants, notamment la Forêt Aléatoire (Random Forest) et les Réseaux de Neurones Artificiels (Neural Network). Pour faciliter le développement et l'expérimentation de ces modèles, nous utilisons l'environnement visuel Orange. Cette plateforme open-source est particulièrement adaptée au domaine du Data Mining et de la Data

Science grâce à son interface intuitive basée sur des workflows visuels. Son principal avantage est de permettre la construction et l'analyse de modèles complexes sans nécessiter de compétences avancées en programmation, rendant l'accès à ces techniques d'IA plus aisé pour un large public. Le modèle développé est testé et validé sur une base de données réelle provenant du centre de recherche de la NASA, reconnue comme une référence dans le domaine de la surveillance des batteries. À travers cette étude, nous espérons apporter une contribution significative au domaine de la maintenance prédictive des systèmes de stockage d'énergie, en démontrant l'efficacité des techniques de Machine Learning appliquées à l'estimation du RUL et en proposant une méthodologie accessible via des outils comme Orange. Nous souhaitons ainsi ouvrir la voie à des solutions plus intelligentes, adaptables et applicables dans des environnements industriels réels. Le présent travail est organisé en plusieurs parties : la première partie est consacrée à une revue générale sur les batteries, leur fonctionnement et leurs applications. La deuxième partie aborde spécifiquement l'Application du Machine Learning dans le Domaine des Batteries, en présentant les algorithmes clés utilisés. Enfin, la troisième partie détaille la méthodologie de notre approche, les résultats obtenus et leur analyse, avant de conclure et de proposer des perspectives pour de futurs travaux

Chapitre I

***Les Batteries Lithium-ion 18650 :
Principes, Caractéristiques et
Importance de la Durée de Vie Utile
Restante (RUL)***

1.1 Introduction

Dans le contexte actuel de transition énergétique, le stockage de l'énergie joue un rôle crucial pour l'intégration des énergies renouvelables intermittentes et le développement des véhicules électriques. Parmi les diverses technologies de stockage, les batteries lithium-ion se sont imposées comme une solution prédominante en raison de leur haute densité énergétique et de leur longue durée de vie. Au sein de cette famille, la batterie lithium-ion de format 18650 est particulièrement répandue dans de nombreuses applications, allant de l'électronique portable aux systèmes de stockage à grande échelle.

Ce chapitre vise à fournir une base théorique et technique approfondie sur les batteries lithium-ion 18650. Nous aborderons leur historique, leurs caractéristiques électrotechniques, leurs principes de fonctionnement et leurs applications. Une attention particulière sera portée sur l'importance de l'estimation de la durée de vie utile restante (RUL) pour la maintenance prédictive et la gestion optimale de ces systèmes énergétiques. Enfin, la base de données B0005 de la NASA, utilisée dans les chapitres 2 et 3 pour la prédiction de la RUL via des modèles d'apprentissage automatique, sera présentée.

1.2 Historique et évolution des batteries lithium-ion

1.2.1 Origines des batteries

L'histoire des batteries remonte à la pile voltaïque inventée par Alessandro Volta en 1800, marquant le début de la conversion chimique en énergie électrique [1]. Depuis lors, diverses technologies ont vu le jour, telles que les batteries plomb-acide, développées au milieu du XIXe siècle, et les batteries nickel-cadmium, apparues au début du XXe siècle, chacune répondant à des besoins spécifiques en termes de puissance et de capacité.

1.2.2 Développement des batteries lithium-ion

Les recherches sur les batteries lithium-ion ont débuté dans les années 1970, avec des travaux pionniers sur l'intercalation du lithium dans des matériaux carbonés et oxydes métalliques. La commercialisation révolutionnaire de la première batterie lithium-ion par Sony en 1991[2], utilisant une anode en carbone et une cathode en oxyde de lithium et de cobalt (LiCoO_2), a ouvert la voie à une nouvelle ère pour l'électronique portable en offrant une densité énergétique significativement plus élevée que les technologies précédentes[3].

1.2.3 Émergence des batteries 18650

Le format 18650, caractérisé par des dimensions de 18 mm de diamètre et 65 mm de hauteur, a été standardisé dans les années 1990. Son adoption rapide dans l'industrie s'explique par sa polyvalence, son coût compétitif, et sa capacité à être regroupée en packs pour diverses applications, des ordinateurs portables aux véhicules électriques [3].

1.3 Caractéristiques électrotechniques des batteries 18650

1.3.1 Structure physique

Une batterie lithium-ion 18650 est composée de plusieurs éléments clés. L'anode est généralement constituée de graphite, tandis que la cathode est faite d'oxydes métalliques. Un électrolyte, souvent un sel de lithium dissous dans un solvant organique, assure le transport des ions lithium entre l'anode et la cathode. Un séparateur microporeux empêche les courts-circuits directs entre les deux électrodes tout en permettant le passage des ions [4].

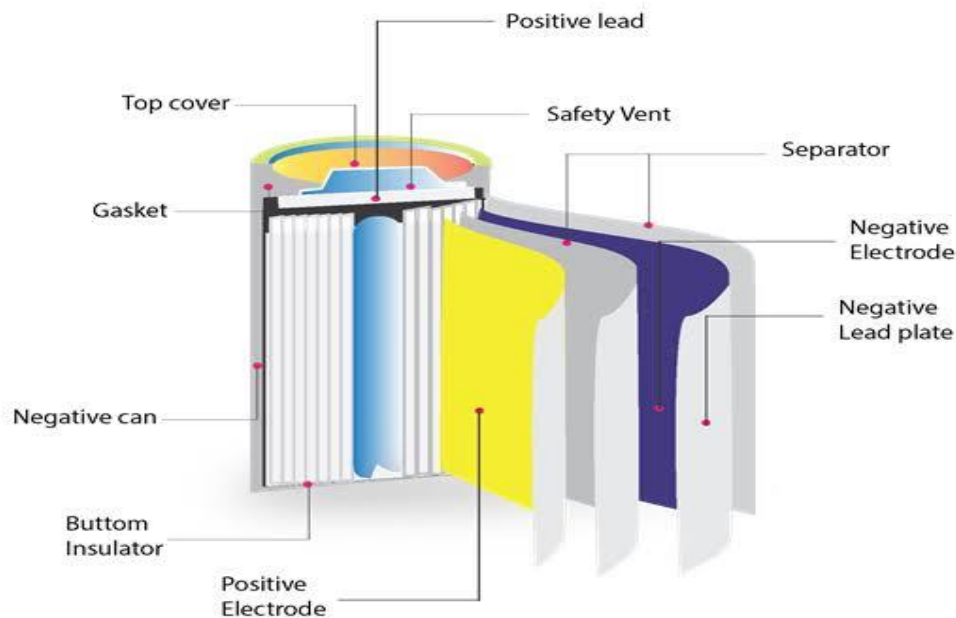


Figure1. 1 : Coupe transversale d'une batterie lithium-ion 18650 [11]

1.3.2 Chimies principales

Il existe plusieurs chimies de cathodes pour les batteries lithium-ion, chacune offrant un compromis différent entre densité énergétique, durée de vie, puissance et sécurité [4]. Le tableau ci-dessous compare les chimies les plus courantes [9].

Tableau 1.1 : Comparaison des principales chimies de cathodes pour batteries lithium ion -

Chimie	Abréviation	Densité Énergétique (Wh/kg)	Durée de Vie (Cycles)	Sécurité	Applications Typiques
Oxyde de Lithium Co-balt	LCO	150-220	500-1000	Modérée	Téléphones portables, ordinateurs portables
Nickel Manganèse Co-balt	NMC	150-220	1000-2000	Bonne	Véhicules électriques, outils
Chimie	Abréviation	Densité Énergétique (Wh/kg)	Durée de Vie (Cycles)	Sécurité	Applications Typiques
Nickel Co-balt Aluminium	NCA	200-260	1000-2000	Bonne	Véhicules électriques (Tesla)
Phosphate de Fer Lithium	LFP	90-120	2000-5000+	Excellente	Stockage stationnaire, bus électriques

1.3.3 Paramètres clés

Les batteries 18650 se caractérisent par des paramètres clés tels que leur capacité nominale, qui varie généralement de 1800 à 3500 mAh, et leur tension nominale, typiquement entre 3,6 et 3,7 V. Leur densité énergétique se situe dans la plage de 150 à 260 Wh/kg, les rendant attractives pour les applications nécessitant une grande autonomie.

1.3.4 Comparaison avec d'autres formats

Bien que les batteries 18650 soient très répandues, d'autres formats existent, tels que les cellules 21700 (21mm de diamètre, 70mm de hauteur) et les cellules prismatiques. Les 21700, par exemple, offrent une capacité et une densité énergétique supérieures aux 18650, ce qui les rend de plus en plus populaires dans les véhicules électriques. Les cellules prismatiques, quant à elles, sont privilégiées pour leur compacité et leur efficacité spatiale dans certains designs de packs batterie.

Format	Dimensions (mm)	Capacité Typique (mAh)	Applications Courantes
18650	18 × 65	2000-3500	Ordinateurs portables, Véhicules électriques, Outils électriques
21700	21 × 70	3000-5000	Véhicules électriques, Systèmes de stockage d'énergie
Cellules prismatiques	Variable	Variable	Smartphones, Tablettes, Véhicules électriques

Tableau1. 2: Comparaison entre le format 18650 et d'autres formats de cellules

1.4 Principes de fonctionnement

1.4.1 Réactions électrochimiques

Le fonctionnement d'une batterie lithium-ion repose sur le principe de l'intercalation et de la désintercalation des ions lithium au sein des matériaux d'électrodes.

Pendant la décharge, les ions lithium se désintercalent de l'anode (graphite) et traversent l'électrolyte et le séparateur pour s'intercaler dans la cathode (par exemple, LiCoO₂).

Simultanément, des électrons sont libérés par l'anode et voyagent à travers le circuit externe

pour atteindre la cathode, produisant ainsi un courant électrique.

La réaction électrochimique à la cathode (pour LiCoO_2) peut être représentée comme suit:



Pendant la charge, le processus est inversé : les ions lithium se désintercalent de la cathode et migrent vers l'anode, où ils s'intercalent dans le graphite, tandis que les électrons retournent à l'anode via le circuit externe [10].

1.4.2 Facteurs de dégradation

La performance et la durée de vie des batteries lithium-ion sont influencées par plusieurs facteurs de dégradation : [9]

- **Température** : Des températures extrêmes (trop basses ou trop élevées) accélèrent la dégradation des matériaux et réduisent la capacité.
- **Taux de charge/décharge** : Des courants élevés (C-rate) peuvent entraîner une formation excessive de la couche d'interface solide-électrolyte (SEI) et une fissuration des électrodes.
- **Cycles de charge/décharge** : Chaque cycle entraîne une certaine dégradation irréversible des matériaux d'électrodes et une perte de capacité.
- **Profondeur de décharge (DoD)** : Une décharge profonde et répétée peut réduire significativement la durée de vie de la batterie.

1.4.3 Illustration

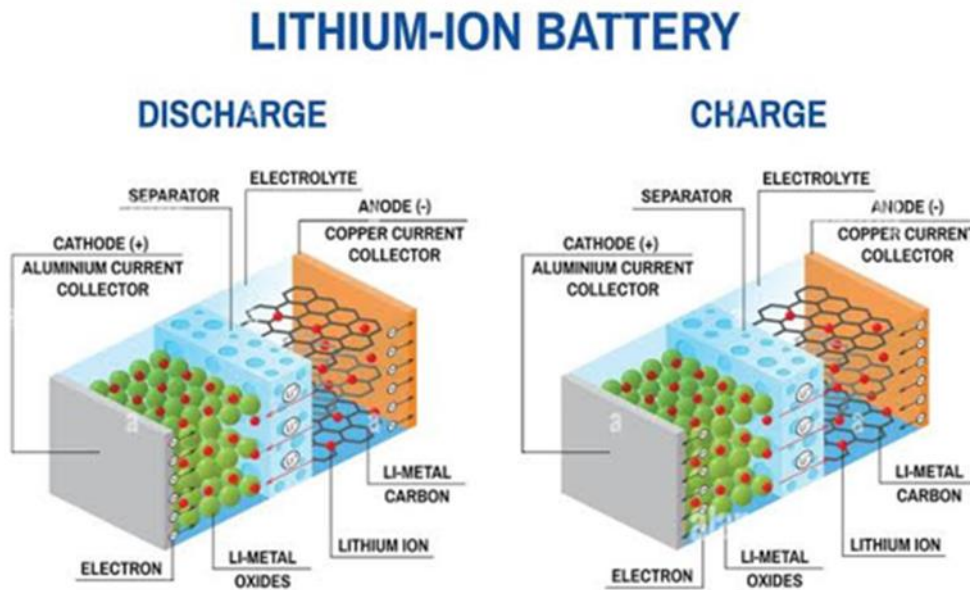


Figure1. 2 : Schéma du fonctionnement d'une batterie lithium-ion (charge/décharge)

[12]

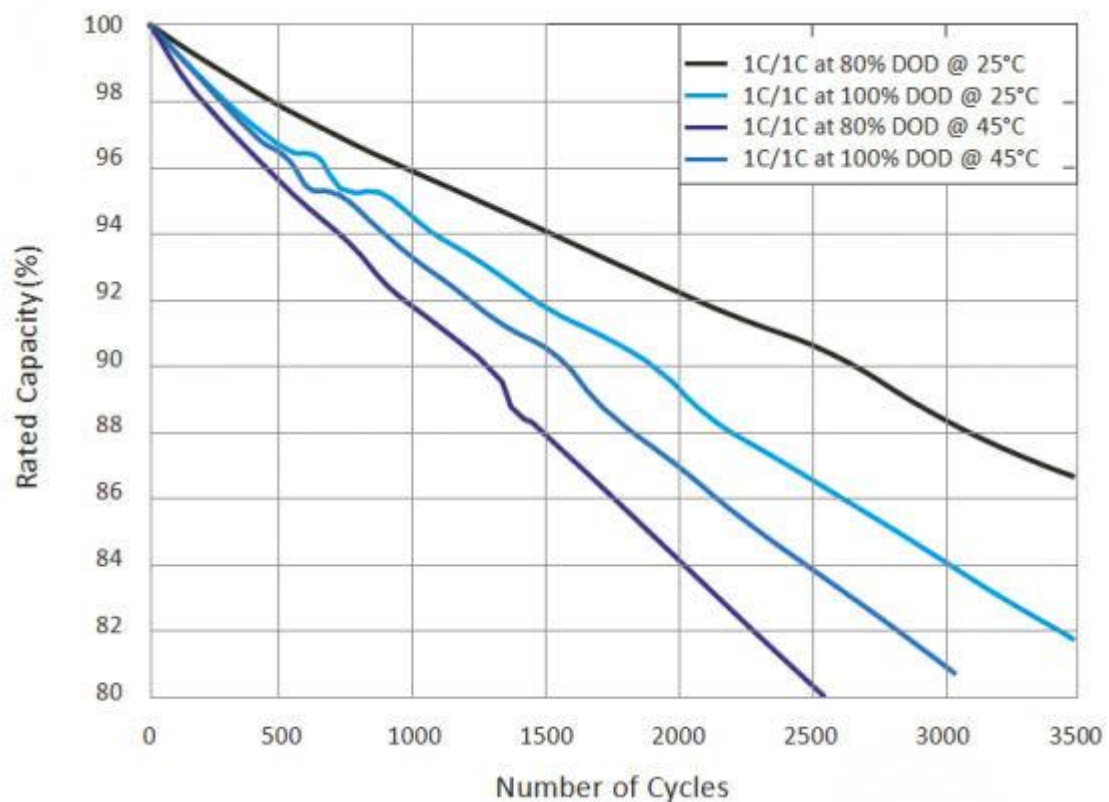


Figure 1.3 : Graphique de dégradation typique de la capacité d'une batterie lithium-ion en fonction des cycles [13]

1.5 Applications des batteries 18650

La polyvalence et les performances des batteries 18650 les ont rendues indispensables dans une multitude d'applications industrielles et grand public.

1.5.1 Véhicules électriques et hybrides

L'un des exemples les plus emblématiques est leur utilisation dans les véhicules électriques. Le constructeur Tesla, par exemple, a initialement massivement utilisé des milliers de cellules 18650 (environ 7000 cellules dans une Tesla Model S) pour constituer ses packs batterie, soulignant leur efficacité pour des applications de forte puissance et d'autonomie. [5]

1.5.2 Stockage d'énergie renouvelable

Les batteries 18650 sont également intégrées dans des systèmes de stockage d'énergie à grande échelle pour les énergies renouvelables. Des projets comme le Tesla Powerwall pour le stockage résidentiel, ou le Hornsdale Power Reserve en Australie, démontrent la capacité de ces batteries à stabiliser les réseaux électriques et à stocker l'énergie solaire ou éolienne excédentaire.

1.5.3 Électronique portable

Historiquement, les 18650 ont trouvé leur place dans une large gamme d'électronique portable, y compris les ordinateurs portables, les lampes torches haute puissance, les cigarettes électroniques, et divers appareils médicaux. Leur format standardisé facilite leur intégration dans des designs compacts et efficaces

1.6 Importance de la durée de vie utile restante (RUL)

L'estimation de la durée de vie utile restante (RUL) est un enjeu majeur pour l'optimisation des performances et la sécurité des systèmes alimentés par des batteries lithium-ion. [6]

1.6.1 Définitions

- **État de Santé (SOH) :** Le SOH est un indicateur de la performance globale d'une batterie par rapport à ses spécifications initiales. Il est souvent exprimé en pourcentage de la capacité nominale initiale.
- **Fin de Vie (EOL) :** La fin de vie est le point où la capacité d'une batterie tombe en dessous d'un seuil critique (généralement 80% de sa capacité nominale), la rendant inadaptée à son application prévue.
- **Durée de Vie Utile Restante (RUL) :** La RUL est le temps ou le nombre de cycles de fonctionnement qu'une batterie peut encore effectuer avant d'atteindre sa fin de vie (EOL). [7]

1.6.2 Méthodes d'estimation

L'estimation de la RUL peut être abordée selon deux catégories principales de méthodes :

- **Modèles physiques** : Basés sur la compréhension des mécanismes de dégradation électrochimique, ces modèles utilisent des équations différentielles et des paramètres physiques pour simuler le vieillissement de la batterie [6].
- **Modèles basés sur les données (Data-driven)** : Ces approches utilisent des algorithmes d'apprentissage automatique pour établir une relation entre les données de fonctionnement (tension, courant, température, capacité) et la RUL, sans nécessiter une connaissance approfondie des processus internes de la batterie. [7]

1.6.3 Enjeux

L'estimation précise de la RUL revêt une importance capitale pour plusieurs raisons:

- **Maintenance prédictive** : Elle permet de planifier les opérations de maintenance ou de remplacement des batteries avant une défaillance inattendue, réduisant ainsi les temps d'arrêt et les coûts.
- **Sécurité** : Une connaissance de la RUL aide à prévenir les situations dangereuses liées au vieillissement avancé des batteries, comme les courts-circuits ou les emballements thermiques.
- **Optimisation des coûts** : En maximisant l'utilisation des batteries jusqu'à leur fin de vie effective, la RUL contribue à réduire les coûts d'exploitation et d'investissement.

Ce chapitre pose les bases pour la suite du mémoire, où les approches d'apprentissage automatique seront utilisées pour prédire la RUL des batteries lithium-ion.

1.7 Présentation de la base de données B0005

Dans le cadre de cette étude, la base de données B0005 a été choisie comme référence pour le développement et la validation des modèles de prédiction de la RUL[8].

1.7.1 Origine

La base de données B0005 est issue du NASA Prognostics Center of Excellence, une institution reconnue pour ses travaux sur la gestion de la santé des systèmes et la maintenance prédictive.

1.7.2 Caractéristiques

La base de données B0005 concerne une batterie lithium-ion de format 18650, avec une capacité nominale de 2 Ah. Les données ont été collectées sur 168 cycles de charge et de décharge, sous des conditions de test spécifiques : charge à 1,5 A et décharge à 2 A. Les

variables enregistrées incluent la tension, le courant, la température et la capacité mesurée à différents points des cycles.

1.7.3 Rôle dans l'étude

Cette base de données constitue le fondement sur lequel les modèles d'apprentissage automatique présentés dans les chapitres 2 et 3 seront entraînés et évalués pour prédire la RUL des batteries.

Cependant, il est important de noter les limites de cette base de données. Sa taille est restreinte (168 cycles), et les échantillons ont été collectés sur une période spécifique, correspondant à environ 1/12 de la période de décharge totale de la batterie. Ces contraintes limitent la généralisation des résultats à d'autres conditions d'utilisation ou chimies de batteries.

1.8 Conclusion

Ce chapitre a fourni une vue d'ensemble exhaustive des batteries lithium-ion 18650, couvrant leur historique, leurs caractéristiques techniques, leurs principes de fonctionnement et leurs diverses applications. Nous avons également souligné l'importance cruciale de l'estimation de la durée de vie utile restante (RUL) pour la maintenance prédictive et la gestion efficace des systèmes de stockage d'énergie. La base de données B0005 de la NASA, malgré ses limites, a été introduite comme la source de données pour les chapitres suivants s'appuieront sur cette base théorique pour explorer les algorithmes d'apprentissage automatique (Chapitre 2) et la méthodologie de prédiction de la RUL, ainsi que les résultats obtenus (Chapitre 3). Ce mémoire vise ainsi à contribuer à l'optimisation des outils de maintenance prédictive pour des solutions énergétiques fiables, intelligentes et durables.

CHAPITRE II : Application du Machine Learning dans le Domaine des Batteries

CHAPITRE II : Application du Machine Learning dans le Domaine des Batteries

2.1 Introduction

Le domaine des batteries, et plus particulièrement celui des batteries lithium-ion, a connu une croissance exponentielle ces dernières années. Cette expansion est principalement due à leur utilisation croissante dans les véhicules électriques et les systèmes de stockage d'énergie renouvelable. Cependant, la complexité inhérente aux processus chimiques et physiques qui régissent le fonctionnement des batteries rend leur modélisation et leur optimisation particulièrement ardues. C'est dans ce contexte que l'apprentissage automatique (ML, *Machine Learning*) intervient, offrant des outils puissants pour prédire, optimiser et diagnostiquer les performances des batteries.

Dans ce chapitre, nous explorerons comment deux modèles d'apprentissage automatique populaires, à savoir les forêts aléatoires (*Random Forest*) et les réseaux de neurones (*Neural Networks*), peuvent être appliqués pour résoudre des problèmes spécifiques au domaine des batteries. Nous examinerons également les avantages et les limites respectifs de chaque modèle dans ce contexte précis.

2.2 Problématiques dans le Domaine des Batteries adressables par l'Apprentissage Automatique

Avant d'aborder en détail les modèles d'apprentissage automatique, il est essentiel de comprendre les principales problématiques auxquelles les chercheurs et les ingénieurs sont confrontés dans le domaine des batteries. Ces défis incluent notamment :

- Prédiction de la durée de vie (Cycle Life) : Estimer le nombre de cycles de charge-décharge qu'une batterie peut supporter avant que sa capacité ne se dégrade de manière significative.
- Estimation de l'état de santé (State of Health, SOH) : Évaluer l'état actuel de la batterie par rapport à son état initial, fournissant une mesure de sa dégradation.
- Détermination de l'état de charge (State of Charge, SOC) : Déterminer la quantité d'énergie restante dans la batterie à un instant donné.

- Détection de défaillances (Fault Detection) : Identifier les signes avant-coureurs de défaillances potentielles, telles que la surchauffe, le court-circuit interne ou la dégradation prématurée.
- Optimisation des matériaux (Materials Optimization) : Découvrir de nouveaux matériaux ou de nouvelles compositions chimiques permettant d'améliorer les performances globales des batteries (densité énergétique, durée de vie, sécurité).

Ces problématiques sont souvent caractérisées par leur complexité et leur nature non linéaire, ce qui les rend particulièrement adaptées à l'application des techniques d'apprentissage automatique.

2.3 Forêts Aléatoires (Random Forest) dans le Domaine des Batteries

2.3.1 Présentation du Modèle

Les forêts aléatoires constituent un modèle d'apprentissage supervisé de type ensemble. Elles sont composées d'un grand nombre d'arbres de décision individuels, où chaque arbre est entraîné sur un sous-ensemble aléatoire des données d'entraînement (échantillonnage bootstrap). Les prédictions finales sont obtenues en agrégeant les prédictions de tous les arbres : par un vote majoritaire pour les tâches de classification, ou par la moyenne des prédictions pour les tâches de régression. Ce modèle est particulièrement apprécié pour sa robustesse, sa capacité à gérer des données bruitées ou manquantes, et sa relative facilité d'interprétation en termes d'importance des variables [14].

2.3.2 Fonctionnement des Forêts Aléatoires

Le fonctionnement d'une forêt aléatoire repose sur les principes suivants :

1. Construction d'arbres multiples (Bagging) : Une forêt aléatoire est constituée de nombreux arbres de décision. Chacun est construit à partir d'un échantillon bootstrap, c'est-à-dire un échantillon obtenu par tirage avec remise à partir du jeu de données d'entraînement original.
2. Sélection aléatoire des variables (Feature Randomness) : Lors de la construction de chaque arbre, à chaque nœud, un sous-ensemble aléatoire de variables prédictives est sélectionné. La meilleure division du nœud est alors recherchée uniquement parmi ces

variables présélectionnées. Ce mécanisme introduit de la diversité entre les arbres et réduit la corrélation entre eux, améliorant ainsi la généralisation du modèle.

3. Agrégation des prédictions : Pour une nouvelle observation, chaque arbre de la forêt fournit une prédiction individuelle. En classification, la classe prédite par la forêt est celle qui obtient le plus de votes (majorité des prédictions des arbres). En régression, la prédiction finale de la forêt est la moyenne des prédictions individuelles des arbres [15].

2.3.3 Avantages des Forêts Aléatoires

Les forêts aléatoires présentent plusieurs avantages significatifs :

1. Robustesse au surapprentissage (Overfitting) : Grâce à l'agrégation des prédictions de multiples arbres (souvent décorrelés), les forêts aléatoires sont moins susceptibles de surapprendre les données d'entraînement comparativement à un arbre de décision unique.
2. Gestion des données manquantes : Elles peuvent gérer efficacement les valeurs manquantes dans les données d'entrée et maintenir une performance élevée même en présence de données incomplètes, grâce à des stratégies d'imputation ou à la nature des arbres de décision.
3. Estimation de l'importance des variables (Feature Importance) : Les forêts aléatoires fournissent des mesures intrinsèques de l'importance des variables, aidant à identifier les caractéristiques les plus influentes pour la prédiction. Ces mesures peuvent être basées, par exemple, sur la diminution moyenne de l'impureté (indice de Gini, variance) apportée par chaque variable lors des divisions, ou sur la dégradation de la performance du modèle après permutation aléatoire des valeurs d'une variable [15].

2.3.4 Applications des Forêts Aléatoires dans le Domaine des Batteries

Les forêts aléatoires trouvent plusieurs applications pertinentes dans le domaine des batteries :

1. Prédiction de la Durée de Vie (RUL/Cycle Life) : Les forêts aléatoires peuvent être utilisées pour prédire la durée de vie d'une batterie en fonction de ses caractéristiques initiales (par exemple, la composition chimique, la capacité nominale) et/ou de son historique d'utilisation (profils de charge/décharge, température). En utilisant des don-

nées historiques sur des batteries similaires, le modèle peut apprendre à associer ces caractéristiques à la longévité de la batterie [16].

2. Estimation de l'État de Santé (SOH) : Les forêts aléatoires peuvent également être employées pour estimer l'état de santé d'une batterie en fonction de ses performances actuelles. Par exemple, en utilisant des données sur la capacité de charge actuelle, la résistance interne et la température, le modèle peut prédire le degré de dégradation de la batterie par rapport à son état initial [17].
3. Détection de Défaillances : Les forêts aléatoires sont efficaces pour identifier des motifs anormaux dans les données qui pourraient indiquer une défaillance imminente. Par exemple, une augmentation soudaine de la résistance interne, une chute anormale de la tension ou une surchauffe pourraient être détectées par le modèle comme des signes précurseurs de défaillance potentielle [18].

2.3.5 Avantages et Limites des Forêts Aléatoires

- Avantages :
 - Facilité de mise en œuvre et d'interprétation (notamment via l'importance des variables).
 - Bonne gestion des données manquantes et bruitées.
 - Moins sujettes au surapprentissage par rapport à un arbre de décision unique.
 - Performances souvent compétitives sur de nombreux types de problèmes.
- Limites :
 - Peuvent être moins performantes pour les problèmes hautement non linéaires ou très complexes, où la structure des données est difficile à capturer par des divisions orthogonales.
 - Peuvent être moins précises que des modèles plus sophistiqués (comme les réseaux de neurones profonds ou les machines à vecteurs de support avec des noyaux complexes) pour certaines tâches spécifiques, surtout avec de grands volumes de données [19].
 - Le grand nombre d'arbres peut rendre le modèle plus coûteux en termes de mémoire et de temps de prédiction par rapport à un modèle plus simple.

2.4 Réseaux de Neurones (Neural Networks) dans le Domaine des Batteries

2.4.1 Présentation du Modèle

Les réseaux de neurones sont des modèles d'apprentissage automatique inspirés du fonctionnement du cerveau humain. Ils sont composés de couches de neurones artificiels (ou nœuds) interconnectés, où chaque connexion possède un poids synaptique ajustable. Ces modèles sont particulièrement puissants pour capturer des relations complexes et non linéaires au sein des données. Les réseaux de neurones profonds (DNN, *Deep Neural Networks*), qui comportent plusieurs couches cachées entre la couche d'entrée et la couche de sortie, sont une extension de ce concept. Ils permettent de modéliser des hiérarchies de caractéristiques et des structures de données encore plus complexes, menant souvent à des performances de pointe dans divers domaines [20].

2.4.2 Composants et Mécanisme de Fonctionnement des Réseaux de Neurones Artificiels

Les principaux composants d'un réseau de neurones sont :

1. Neurones artificiels (Nœuds) : Unités de traitement élémentaires qui reçoivent des signaux d'entrée, leur appliquent des poids spécifiques, calculent une somme pondérée (à laquelle un biais peut être ajouté), puis transmettent ce résultat à travers une fonction d'activation pour produire une sortie.
2. Couches (Layers) : Les réseaux sont typiquement organisés en plusieurs couches :
 - Couche d'entrée (Input Layer) : Reçoit les données brutes (les caractéristiques de l'échantillon).
 - Couches cachées (Hidden Layers) : Une ou plusieurs couches qui traitent les données à travers des transformations mathématiques successives. C'est ici que le réseau apprend des représentations de plus en plus abstraites des données.
 - Couche de sortie (Output Layer) : Produit les résultats finaux ou les prédictions du modèle (par exemple, une valeur de RUL en régression, ou des probabilités de classe en classification).
3. Mécanisme de fonctionnement :

Les données traversent le réseau de la couche d'entrée vers la couche de sortie (propagation avant, *feedforward*). Dans chaque neurone, la somme pondérée des en-

trées est calculée, puis passée à une fonction d'activation (par exemple, sigmoïde, ReLU, tanh) pour introduire la non-linéarité. Les poids des connexions sont ajustés durant le processus d'apprentissage (entraînement) via un algorithme d'optimisation (souvent basé sur la descente de gradient et la rétropropagation de l'erreur) afin de minimiser une fonction de coût (ou fonction de perte) qui mesure l'écart entre les sorties prédites par le réseau et les sorties réelles (cibles). Ce processus itératif permet au réseau d'apprendre à mapper les entrées aux sorties souhaitées [21].

2.4.3 Applications des Réseaux de Neurones dans le Domaine des Batteries

Les réseaux de neurones sont largement appliqués à diverses problématiques relatives aux batteries :

1. Prédiction de la Durée de Vie (RUL/Cycle Life) : Les réseaux de neurones, en particulier les architectures récurrentes (RNN, *Recurrent Neural Networks*) ou à mémoire à long et court terme (LSTM, *Long Short-Term Memory*), sont bien adaptés pour prédire la durée de vie d'une batterie en analysant des séquences temporelles de données complexes, telles que les courbes de charge-décharge, les variations de température, et les mesures de résistance interne au fil des cycles [22].
2. Estimation de l'État de Charge (SOC) : Les réseaux de neurones peuvent être utilisés pour estimer l'état de charge d'une batterie en temps réel. En utilisant des données d'entrée telles que la tension, le courant et la température, le modèle peut apprendre à prédire la quantité d'énergie restante dans la batterie avec une grande précision, même en présence de dynamiques non linéaires [23].
3. Optimisation des Matériaux : Les réseaux de neurones peuvent être employés pour explorer de nouvelles compositions chimiques ou structures de matériaux pour les batteries. En entraînant un modèle sur des données expérimentales existantes reliant la composition des matériaux à leurs performances, il est possible de prédire les performances de nouveaux matériaux candidats sans avoir à les synthétiser et à les tester physiquement, accélérant ainsi le processus de découverte [24].

2.4.4 Avantages et Limites des Réseaux de Neurones

- **Avantages :**
 - Très puissants pour modéliser des relations complexes et non linéaires.
 - Capacité à apprendre des représentations hiérarchiques des caractéristiques à partir des données brutes (surtout les DNN).

- Peuvent être adaptés à différents types de données, y compris les données séquentielles (RNN, LSTM) ou les données spatiales/images (CNN, *Convolutional Neural Networks*).
- Capables de capturer des motifs subtils et des interactions complexes entre les variables dans les données.
- **Limites :**
 - Requiert généralement une grande quantité de données pour l'entraînement afin d'éviter le surapprentissage et d'atteindre de bonnes performances.
 - Souvent considérés comme des "boîtes noires" en raison de leur difficulté d'interprétation directe (bien que des techniques d'explicabilité émergent).
 - Sensibles au surajustement si le modèle est trop complexe par rapport à la quantité de données disponibles, ou si l'entraînement n'est pas correctement régularisé.
 - L'entraînement peut être coûteux en termes de temps de calcul et de ressources, surtout pour les réseaux profonds [52].
 - Le choix de l'architecture (nombre de couches, nombre de neurones par couche, fonctions d'activation) et des hyperparamètres d'entraînement peut être complexe et nécessiter une expertise considérable.

2.5 Comparaison des Modèles

Tableau 2.1 : Comparaison des Modèles Random Forest et Réseaux de Neurones

Critère	Forêts Aléatoires	Réseaux de Neurones
Complexité des Données	Adapté aux données structurées	Adapté aux données complexes et non linéaires
Interprétabilités	Facile à interpréter	Difficile à interpréter (boîte noire)
Volume de Données	Fonctionne bien avec des données limitées	Requiert de grandes quantités de données
Temps d'Entraînement	Rapide	Peut-être long (surtout pour les réseaux profonds)
Precision	Bonne pour des problèmes simples	Très élevée pour des problèmes complexes

2.6 Conclusion :

Après une analyse approfondie des avantages et des applications des modèles d'apprentissage automatique dans le domaine des batteries, il ressort que ces techniques offrent des capacités puissantes pour traiter des données complexes et prédire avec précision des variables clés telles que la durée de vie utile restante (RUL). Toutefois, la mise en œuvre de ces modèles nécessite un environnement adapté pour l'analyse des données et l'exécution des algorithmes. C'est pourquoi le choix d'une plateforme efficace devient crucial. Dans le chapitre suivant, nous présenterons la méthodologie adoptée dans ce travail à l'aide de la plateforme Orange, en détaillant les étapes allant de l'importation des données, au prétraitement et à l'entraînement des modèles, jusqu'à leur évaluation et l'analyse des résultats obtenus.

CHAPITRE III : Méthodologie, Résultats et Analyses

CHAPITRE III : Méthodologie, Résultats et Analyses

3.1 Introduction

Après avoir abordé les aspects théoriques relatifs aux types de batteries et aux modèles d'apprentissage automatique utilisés pour estimer leur durée de vie utile, ce chapitre se concentre sur l'aspect pratique de l'étude en présentant la méthodologie adoptée pour la mise en œuvre des expériences et l'analyse des données. La plateforme Orange a été choisie en raison des outils visuels performants qu'elle offre, facilitant l'exécution des processus analytiques, depuis l'importation et le prétraitement des données, jusqu'à la définition des variables cibles, l'entraînement des modèles et l'évaluation de leurs performances. Ce chapitre vise à documenter rigoureusement chaque étape technique suivie, en clarifiant la logique du flux de travail adopté, afin de garantir la reproductibilité de l'expérience et l'obtention de résultats fiables et comparables.

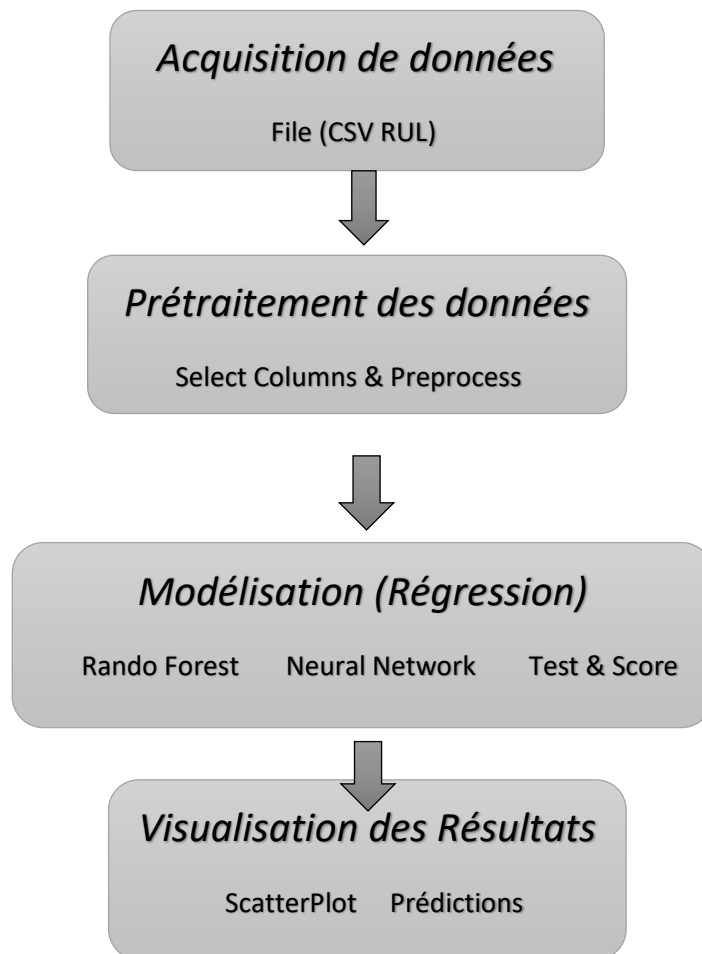


Figure 3.1: Organigramme du Processus de Prédiction de la RUL

3.1.1 Base de données

Le jeu de données B0005, issu du prestigieux NASA Ames Prognostics Center of Excellence, est spécifiquement conçu pour l'étude de la dégradation des batteries lithium-ion (Li-Ion) et la prédiction de leur durée de vie utile (RUL). Il provient d'expérimentations rigoureusement contrôlées, où des cellules Li-ion commerciales (type 18650) sont soumises à des cycles répétés de charge-décharge jusqu'à atteindre leur seuil de fin de vie.

Concernant B0005, les cellules ont été testées à température ambiante (environ 24 °C). Le protocole de charge consistait en un courant constant (CC) de 1,5 A jusqu'à une tension de 4,2 V, suivi d'une phase à tension constante (CV) jusqu'à ce que le courant descende à 20 mA. La décharge s'effectuait à CC de 2 A, jusqu'à une tension de coupure de 2,7 V. Chaque cycle est documenté par des mesures précises de tension, courant, température et capacité, permettant d'observer la diminution progressive de la capacité.

Le jeu B0005 couvre 168 cycles, la fin de vie étant atteinte lorsque la capacité descend sous 1,1 Ah (conventionnellement 80 % de la capacité initiale). L'objectif principal de cette base est de faciliter le développement et la validation de modèles d'apprentissage automatique ou statistiques capables de prédire la RUL à partir de variables mesurables en temps réel. Elle est largement exploitée dans les études de pronostic, notamment via des approches de régression, de réseaux de neurones, ou de modèles bayésiens. Cette base de données est accessible publiquement via le NASA Prognostics Data Repository, avec une référence académique clé à Saha et Goebel (2007).

3.2 Workflow widget:

L'utilisation du widget "Fichier" est la première étape concrète de notre protocole expérimental. Il garantit que l'ensemble de données spécifié est correctement identifié et chargé dans l'environnement d'analyse. Les métadonnées qu'il fournit (nombre d'instances et de caractéristiques, noms des caractéristiques) permettent une première validation de la conformité des données par rapport aux attentes. Une fois les données chargées et validées à ce niveau, elles sont transmises aux widgets subséquents du workflow pour les étapes de prétraitement, d'exploration, de modélisation ou de visualisation.

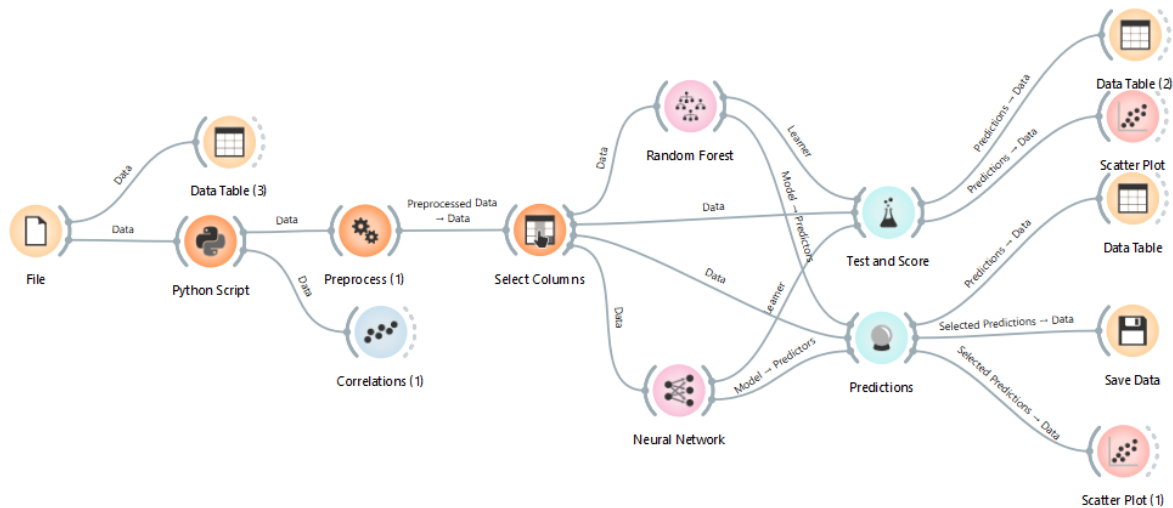


Figure3.2 : Workflow d'Apprentissage Automatique pour la Prédiction de la RUL des Batteries avec Comparaison de Modèles dans Orange

3.3 Importation de Données ("Fichier")

L'étape initiale de tout processus d'analyse de données consiste en l'acquisition et le chargement des données brutes. Dans le cadre de notre workflow, cette fonction est assurée par le widget "Fichier" (File). Ce composant est fondamental car il constitue le point d'entrée des données qui seront soumises aux analyses ultérieures.

3.3.1 Fonctionnalités et Rôle :

Le widget "Fichier" a pour principale responsabilité de localiser, lire et interpréter un fichier de données stocké localement ou accessible via un chemin réseau. Une fois le fichier sélectionné et chargé avec succès, le widget extrait et affiche des métadonnées essentielles le concernant, permettant à l'utilisateur de vérifier la bonne prise en charge des données.

Dans le cas présent, les informations affichées par le widget (voir Figure3) sont les suivantes :

File

File name: C:/Users/Timgad informatique/Downloads/nasa_source_data_1.csv

Format: Comma-separated values

Data

Data instances: 168

Features: Cycle, Voltage_measured, Current_measured, Temperature_measured, Capacity

Figure3.3: Une image montrant le fichier de données

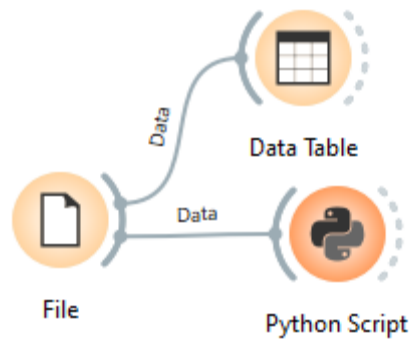


Figure 3.4 . Image montrant la sortie de l'outil Fichier après l'ajout de données

1. Instances de données (Data instances) : 168

- Ce nombre représente le total d'enregistrements, de lignes ou d'observations contenus dans le fichier de données après son chargement. Dans notre contexte, cela signifie que l'ensemble de données nasa_source_data_1.csv comprend 168 entrées distinctes.

2. Caractéristiques (Features): Cycle, Voltage_measured, Current_measured, Temperature_measured, Capacity

- Cette liste énumère les noms des colonnes, attributs ou variables présents dans l'ensemble de données. Chacune de ces caractéristiques représente une dimension spécifique des données collectées. Pour notre étude, les variables considérées sont donc : le cycle, la tension mesurée, le courant mesuré, la température mesurée et la capacité.

3.4 Widget d'Exécution de Script Python ("Python Script")

Suite au chargement initial des données via le widget "Fichier", l'étape suivante de notre méthodologie implique une transformation spécifique des données, à savoir la création de notre variable cible : la Durée de Vie Utile Résiduelle (Remaining Useful Life - RUL). Cette opération est réalisée à l'aide du widget "Python Script". Ce composant offre une flexibilité considérable en permettant l'intégration et l'exécution de code Python personnalisé directement au sein du workflow d'analyse.

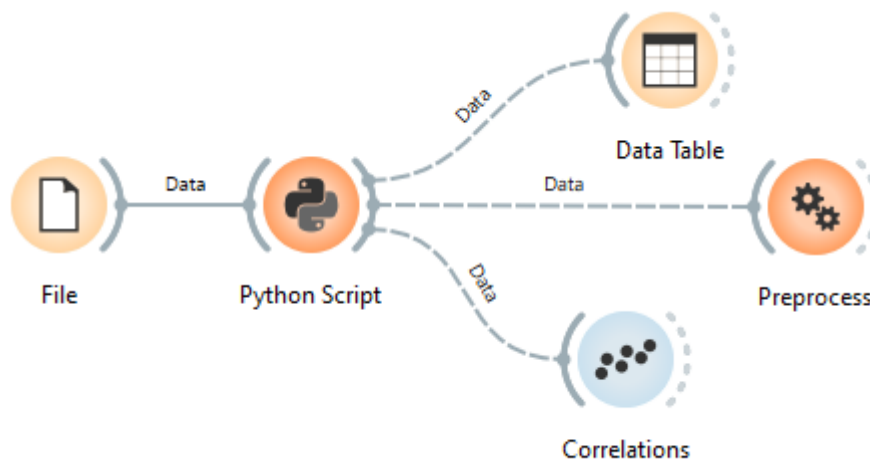


Figure3.5 : les sorties et les entrées de l'outil Python Script après le calcul du RUL

```

Editor
def python_script(in_data):
    1 import pandas as pd
    2 from Orange.data import Table, Domain, ContinuousVariable
    3
    4 # Charger les données depuis l'entrée Orange
    5 in_data = in_data if in_data else None
    6
    7 if in_data:
    8     # Convertir les données Orange en DataFrame pandas
    9     df = pd.DataFrame(in_data.X, columns=[var.name for var in in_data.domain.attributes])
    10
    11     # Ajouter la colonne RUL
    12     if "Cycle" in df.columns:
    13         df["RUL"] = df["Cycle"].max() - df["Cycle"]
    14
    15 return out_data, out_learner, out_classifier, out_object

Console
Python 3.11.8 | packaged by conda-forge | (main, Feb 16 2024, 20:40:50) [MSC v.1937 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
(PythonConsole)
>>>
Running script:
>>>

```

Figure 3.6: Script Python pour la Génération de la Variable Cible RUL

3.4.1 Fonctionnalités :

Le widget "Python Script" est employé lorsque les opérations de prétraitement ou de création de caractéristiques (feature engineering) requises dépassent les capacités des widgets standards ou nécessitent une logique de calcul spécifique. Dans notre cas, il est utilisé pour calculer la RUL à partir de la variable "Cycle" existante.

3.4.2 Description du Script Utilisé :

Le script Python implémenté au sein de ce widget effectue les opérations séquentielles suivantes :

1. Importation des bibliothèques nécessaires :

- import pandas as pd: La bibliothèque Pandas est importée sous l'alias pd. Elle est essentielle pour la manipulation de données tabulaires structurées sous forme de DataFrames, offrant des opérations performantes et intuitives.
- from Orange.data import Table, Domain, ContinuousVariable: Des modules spécifiques de la bibliothèque Orange sont importés pour permettre l'interfaçage avec les structures de données propres à Orange. Table est le format de données principal d'Orange, Domain décrit la structure de ces tables (types de variables, etc.), et ContinuousVariable est utilisé pour définir les variables numériques continues.

2. Réception des données d'entrée :

- in_data = in_data if in_data else None: Le script récupère les données transmises par le widget précédent (dans ce cas, le widget "Fichier") via la variable in_data. Une vérification est effectuée pour s'assurer que des données sont effectivement présentes.

3. Conversion des données Orange en DataFrame Pandas :

- if in_data:: Si des données d'entrée existent :
 - df = pd.DataFrame(in_data.X, columns=[var.name for var in in_data.domain.attributes]): Les données Orange (in_data.X contenant les valeurs des attributs) sont converties en un DataFrame Pandas. Les noms des colonnes du DataFrame sont extraits des attributs du domaine des données Orange (in_data.domain.attributes). Cette conversion est
 - Cruciale car elle permet d'exploiter les fonctionnalités de manipulation de données de Pandas.

4. Calcul et ajout de la variable RUL :

- if "Cycle" in df.columns:: Le script vérifie la présence de la colonne "Cycle" dans le DataFrame, qui est indispensable pour le calcul de la RUL.
 - `df["RUL"] = df["Cycle"].max() - df["Cycle"]`: C'est l'opération centrale. Une nouvelle colonne nommée "RUL" est créée dans le DataFrame. Pour chaque instance (chaque ligne), la valeur de la RUL est calculée comme la différence entre la valeur maximale de la colonne "Cycle" observée dans l'ensemble du dataset (correspondant au cycle final avant la défaillance ou la fin de l'observation) et la valeur actuelle de "Cycle" pour cette instance. Cette définition de la RUL est couramment utilisée dans les études de pronostic et de gestion de la santé (PHM), notamment pour la dégradation de batteries.
 - Cette nouvelle variable "RUL" représente donc, pour chaque cycle, le nombre de cycles restants avant d'atteindre le cycle maximal observé dans les données. Elle devient ainsi notre **variable cible** (target variable) pour les modèles de prédiction que nous développerons ultérieurement.

5. Conversion du DataFrame Pandas modifié en Table Orange :

- `domain = Domain([ContinuousVariable(col) for col in df.columns], metas=None)`: Un nouveau domaine Orange est défini. Il inclut toutes les colonnes du DataFrame Pandas (y compris la nouvelle colonne "RUL"), chacune étant définie comme une variable continue (ContinuousVariable).
- `out_data = Table.from_numpy(domain, df.values)`: Le DataFrame Pandas, contenant désormais la variable RUL, est reconverti en une Table Orange. Les valeurs du DataFrame (`df.values`) sont utilisées pour peupler cette table, en respectant le domaine nouvellement défini. Cette table `out_data` sera ensuite transmise au widget suivant dans le workflow.

3.5 Widget de Visualisation de Table de Données ("Data Table")

Le widget "Data Table" permet une inspection détaillée du contenu réel de l'ensemble de données. Il affiche les données sous une forme tabulaire, où les lignes représentent les instances et les colonnes représentent les caractéristiques.

3.5.1 Fonctionnalités et Rôle :

Le widget "Data Table" est un outil essentiel pour l'exploration et la vérification des données. Il permet à l'analyste de :

- Visualiser les valeurs réelles des différentes caractéristiques pour chaque instance.
- Vérifier la bonne exécution des étapes de prétraitement (par exemple, s'assurer que les calculs comme celui de la RUL ont été correctement effectués).
- Identifier des anomalies potentielles, des outliers ou des patterns simples à l'œil nu.
- Comprendre la structure et le format des données de manière concrète.

Dans l'image fournie (voir Figure 1 en faisant référence à l'instance du widget "Data Table"), nous observons :

Tableau3.1. Tableau montrant les données de la batterie après traitement et ajout de la colonne RUL

Data instances: 30
Features: 6
Meta attributes: None

	Cycle	Voltage_measured	Current_measured	Temperature_measured	Capacity	RUL
1	2	3.52983	-1.8187	32.5723	1.85649	612
2	4	3.53732	-1.81756	32.7252	1.84633	610
3	6	3.54374	-1.81649	32.6429	1.83535	608
4	8	3.54367	-1.82559	32.5149	1.83526	606
5	10	3.54234	-1.82611	32.3823	1.83465	604
6	12	3.54133	-1.81681	32.4342	1.83566	602
7	14	3.54103	-1.81679	32.4804	1.83515	600
8	16	3.55413	-1.84434	32.4105	1.82576	598
9	18	3.55294	-1.85397	32.3461	1.82477	596
10	20	3.55121	-1.85359	32.2768	1.82461	594
11	22	3.54929	-1.86338	32.1571	1.82462	592
12	25	3.55976	-1.87297	32.3641	1.8142	589
13	27	3.55808	-1.88261	32.3182	1.81375	587
14	29	3.55569	-1.88254	32.1571	1.81344	585
15	31	3.56275	-1.88179	32.0441	1.8026	583
16	33	3.56145	-1.89219	31.9316	1.80211	581
17	35	3.56034	-1.90276	31.9417	1.80258	579
18	37	3.55912	-1.91335	31.965	1.80307	577
19	39	3.55789	-1.91316	31.9662	1.80278	575
20	42	3.55204	-1.85535	32.3714	1.84703	572
21	46	3.55257	-1.8747	32.7002	1.84742	568
22	50	3.55679	-1.87383	32.3227	1.83618	564
23	54	3.56141	-1.88319	31.988	1.82578	560
24	58	3.55661	-1.87324	31.8945	1.82511	556
25	62	3.55867	-1.91396	31.7804	1.82558	552
26	66	3.5641	-1.90305	31.6887	1.81403	548
27	70	3.55959	-1.91346	31.5967	1.81477	544
28	74	3.55683	-1.91362	31.5656	1.81397	540
29	78	3.56244	-1.90258	31.5912	1.80277	536
30	82	3.55914	-1.91306	31.6035	1.80408	532

1. Information d'en-tête:

- **Instances de données (Data instances) :** 168. Confirme le nombre total de lignes.

- **Caractéristiques (Features) :** 6. Ceci est un point clé : il indique que la table affichée contient désormais 6 colonnes. En comparant avec les 5 caractéristiques initiales du widget "Fichier" (Cycle, Voltage_measured, Current_measured, Temperature_measured, Capacity), cette sixième colonne est la variable RUL que nous avons créée avec le "Python Script".
- **Attributs méta (Meta attributes) :** None. Indique qu'il n'y a pas de colonnes de métadonnées distinctes des caractéristiques principales.

2. Contenu de la table:

- Le tableau affiche les premières lignes de l'ensemble de données (avec une indication + 68 mores à la fin de la capture, signifiant que toutes les 168 lignes ne sont pas visibles sur une seule page).
- Les colonnes affichées sont: Cycle, Voltage_measured, Current_measured, Temperature_measured, Capacity, et, de manière cruciale, RUL.
- On peut y observer les valeurs calculées pour la RUL. Par exemple, pour la première ligne où Cycle = 2, la RUL = 612. Si l'on suppose que le Cycle_max pour ce dataset est de 614 (ce qui serait le cas si la dernière valeur de cycle observée était 614), alors $614 - 2 = 612$, ce qui est cohérent. Le calcul de RUL $C_{max} - C_i$ semble avoir été correctement appliqué.

3.6 Observations Issues du Widget :

Le widget "Corrélations" affiche une liste de paires de variables, classées par la valeur absolue de leur coefficient de corrélation de Pearson, des plus fortes aux plus faibles. L'image fournie (voir Tableau 2,3) présente deux listes, la seconde étant la plus pertinente car elle inclut notre variable cible RUL :

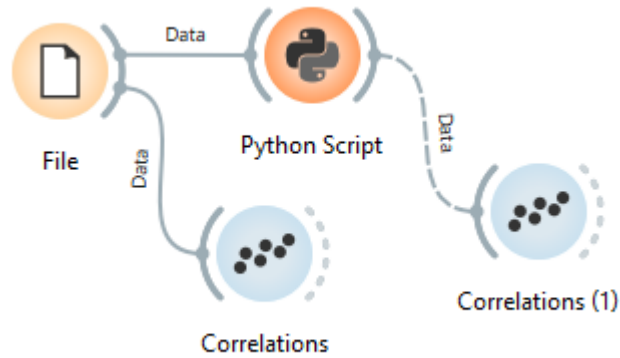


Figure3.7 : Génération de la RUL et Corrélations

Tableau3.2 : Rapport de Corrélacion des Caractéristiques Avant l'Ajout de la Variable RUL.

Pearson correlation			
1	-0.989	Capacity	: Cycle
2	+0.982	Capacity	: Voltage_measured
3	-0.974	Current_measured	: Voltage_measured
4	-0.966	Cycle	: Voltage_measured
5	-0.933	Capacity	: Current_measured
6	+0.908	Current_measured	: Cycle
7	-0.824	Temperature_measured	: Voltage_measured
8	+0.819	Current_measured	: Temperature_measured
9	-0.810	Capacity	: Temperature_measured
10	+0.791	Cycle	: Temperature_measured

Tableau3.3 : Rapport de Corrélacion des Caractéristiques Après l'Ajout de la Variable RUL

Pearson correlation

1	-1.000	Cycle	:	RUL
2	-0.989	Capacity	:	Cycle
3	+0.989	Capacity	:	RUL
4	+0.982	Capacity	:	Voltage_measured
5	-0.974	Current_measured	:	Voltage_measured
6	-0.966	Cycle	:	Voltage_measured
7	+0.966	RUL	:	Voltage_measured
8	-0.933	Capacity	:	Current_measured
9	-0.908	Current_measured	:	RUL
10	+0.908	Current_measured	:	Cycle
11	-0.824	Temperature_measured	:	Voltage_measured
12	+0.819	Current_measured	:	Temperature_measured
13	-0.810	Capacity	:	Temperature_measured
14	-0.791	RUL	:	Temperature_measured
15	+0.791	Cycle	:	Temperature_measured

3.6.1 l'élément RUL

1. Cycle: RUL (-1.000)

- Une corrélation négative parfaite. Ceci est attendu et sert de validation, car RUL a été défini comme $\text{Cycle_max} - \text{Cycle}$. À mesure que Cycle augmente, RUL diminue linéairement par définition.

2. Capacity: Cycle (-0.989)

- Une très forte corrélation négative. Cela indique que la capacité de la batterie diminue fortement à mesure que le nombre de cycles augmente, ce qui est conforme au comportement de dégradation attendu des batteries.

3. Capacity: RUL (+0.989)

- Une très forte corrélation positive. Les batteries ayant une capacité plus élevée ont tendance à avoir une RUL plus importante. C'est un indicateur clé pour la prediction.

4. Capacity: Voltage measured (+0.982)

- Une très forte corrélation positive entre la capacité et la tension mesurée.

5. Current measured: Voltage measured (-0.974)

- Une très forte corrélation négative. La nature de cette relation dépend du protocole de mesure (par exemple, pendant une décharge à courant constant).

6. Cycle: Voltage measured (-0.966)

- Une très forte corrélation négative. La tension tend à diminuer avec l'augmentation des cycles.
- 7. **RUL: Voltage_measured (+0.966)**
 - Une très forte corrélation positive. Une tension mesurée plus élevée est fortement associée à une RUL plus grande.
- 8. **Capacity: Current_measured (-0.933)**
 - Forte corrélation négative.
- 9. **Current_measured: RUL (-0.908)**
 - Forte corrélation négative. Cela suggère que des valeurs plus élevées de courant mesuré (potentiellement indicatives d'une dégradation ou d'un stress plus important, selon la signification de la variable) sont associées à une RUL plus faible.
- 10. **Current_measured: Cycle (+0.908)**
 - Forte corrélation positive.

Les autres corrélations listées : (Temperature_measured avec Voltage_measured, Current_measured, Capacity, RUL, et Cycle) sont également significatives, bien que d'une magnitude légèrement inférieure. Par exemple, RUL : Temperature_measured (-0.791) indique une corrélation négative modérée à forte, suggérant que des températures plus élevées sont associées à une RUL plus faible.

3.6.2 Importance de l'Analyse des Corrélations dans ce Mémoire :

L'analyse des corrélations, et en particulier celle de Pearson, revêt une importance capitale dans le cadre de ce mémoire pour plusieurs raisons :

1. **Sélection de Caractéristiques (Feature Selection) :**
 - **Identification des prédicteurs pertinents :** Les variables qui présentent une forte corrélation (positive ou négative) avec la variable cible (RUL) sont d'excellents candidats pour être incluses comme prédicteurs dans nos modèles de machine learning. Par exemple, Capacity, Voltage_measured, et Current_measured montrent toutes de fortes corrélations avec RUL et sont donc susceptibles d'être informatives pour sa prédiction.
 - **Compréhension de la redondance :** De fortes corrélations entre les variables prédictives elles-mêmes (par exemple, entre capacité et Voltage_measured) peuvent indiquer une multi colinéarité. Bien que ce widget ne la traite pas directement, il aide à identifier les paires de variables qui pourraient porter des informations redondantes. Certains modèles sont sensibles à la multi colinéarité, et cette analyse

peut guider des étapes de réduction de dimensionnalité ou une sélection plus judicieuse des caractéristiques.

2. Compréhension du Phénomène Étudié:

- Les corrélations observées permettent de valider ou d'infirmer des hypothèses basées sur la physique ou la chimie de la dégradation des batteries. Par exemple, la corrélation négative entre capacité et Cycle est attendue. Si des corrélations inattendues apparaissaient, cela pourrait indiquer des problèmes dans les données ou révéler des aspects méconnus du processus de dégradation.

3. Orientation pour la Modélisation:

- La présence de fortes relations linéaires suggère que les modèles linéaires (ou les modèles capables de capturer des composantes linéaires) pourraient être performants. Inversement, si les corrélations avec la cible étaient toutes faibles, cela pourrait indiquer la nécessité de modèles plus complexes capables de capturer des relations non linéaires, ou que les caractéristiques disponibles ne sont pas suffisamment prédictives.

4. Validation des Données et de l'Ingénierie des Caractéristiques :

- La corrélation parfaite de -1.000 entre Cycle et RUL confirme que notre calcul de RUL via le "Python Script" a été effectué comme prévu.

3.7 Widget de Prétraitement des Données ("Preprocess")

Après l'analyse exploratoire, et avant d'appliquer les algorithmes d'apprentissage automatique, il est souvent indispensable de prétraiter les données. Le widget "Preprocess" offre une panoplie de techniques pour transformer les caractéristiques (features) afin d'améliorer la performance et la stabilité des modèles prédictifs. Dans le cadre de notre méthodologie, nous utilisons ce widget pour la normalisation des caractéristiques.



Figure 3.8 : Transformation et Sélection des Données

3.7.1 Fonctionnalités et Rôle :

Le prétraitement des données englobe diverses opérations telles que la gestion des valeurs manquantes, l'encodage des variables catégorielles, la discrétisation, la sélection de caractéristiques, et la mise à l'échelle (scaling) des caractéristiques numériques. De nombreux algorithmes d'apprentissage automatique, notamment ceux basés sur des mesures de distance (comme les k-plus proches voisins, SVM) ou ceux utilisant des descentes de gradient (comme les réseaux de neurones), sont sensibles à l'échelle des caractéristiques d'entrée. Des caractéristiques avec des plages de valeurs très différentes peuvent conduire à une convergence plus lente de l'algorithme ou à une situation où les caractéristiques ayant les plus grandes amplitudes dominant indûment le processus d'apprentissage.

3.8 Widget de Sélection de Colonnes ("Select Columns")

Le widget "Select Columns" est une étape de configuration cruciale dans la préparation des données pour l'apprentissage supervisé. Sa fonction essentielle est de permettre la désignation explicite des rôles des différentes colonnes de l'ensemble de données : quelles colonnes serviront de caractéristiques prédictives (features) et quelle colonne constituera la variable cible (target) à prédire.

3.8.1 Configuration et Rôle dans le Workflow :

Dans notre méthodologie, après les étapes de chargement, de création de la variable RUL, d'analyse exploratoire et de normalisation, ce widget est utilisé comme suit :

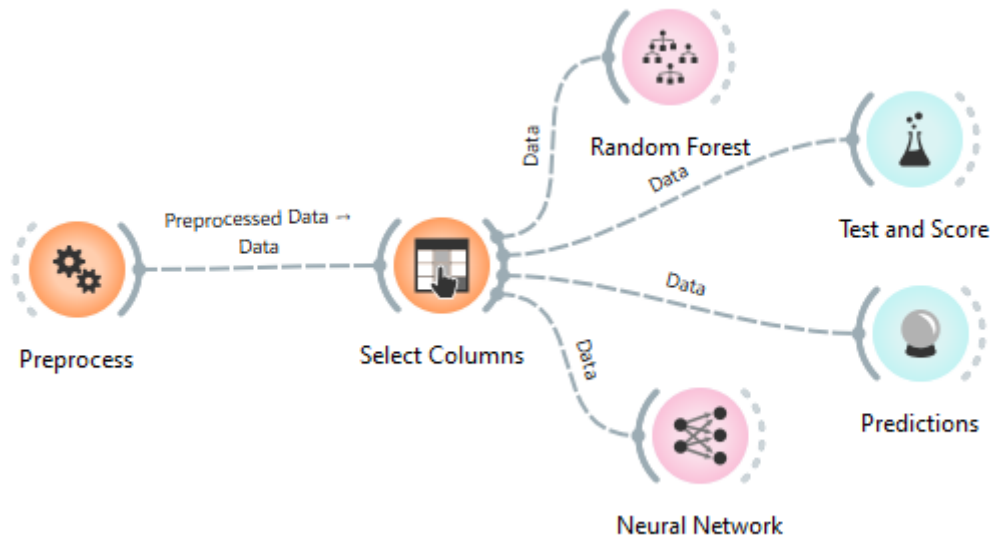


Figure3. 9 : Flux de Données à Travers le Widget "Sélectionner les Colonnes".

Cette configuration stipule que les variables Voltage_measured, Current_measured, Temperature_measured, Capacite et Cycle sont sélectionnées en tant que prédicteurs (caractéristiques) fournis en entrée des modèles d'apprentissage. La variable RUL (Remaining Useful Life), préalablement calculée, est explicitement désignée comme la variable cible (dépendante) que les modèles s'efforceront de prédire.

3.9 Widget de forêt aléatoire ("Random Forest")

Le widget Forêt Aléatoire dans Orange permet de construire un modèle d'apprentissage automatique basé sur l'algorithme Random Forest. Cet algorithme est robuste et utilisé pour les tâches de classification ou de régression en combinant plusieurs arbres de décision.

Name: Random Forest

Model parameters

Number of trees: 50
Maximal number of considered features: 2
Replicable training: No
Maximal tree depth: 10
Stop splitting nodes with maximum instances: 10

Data

Data instances: 168
Features: Voltage_measured, Current_measured, Temperature_measured, Capacity, Cycle
Target: RUL

Figure 3. 10 : les paramètres du modèle Random Forest.

Ce modèle de Forêt Aléatoire (Random Forest) est constitué de 50 arbres de décision. Pour chaque division au sein d'un arbre, un maximum de 2 caractéristiques sont considérées aléatoirement, et la profondeur maximale de chaque arbre est limitée à 10 niveaux. De plus, la division d'un nœud s'arrête s'il contient 10 instances ou moins. Il est important de noter que l'entraînement de ce modèle n'est pas configuré pour être reproductible. Le modèle utilise 168 instances de données, avec les caractéristiques Voltage_measured, Current_measured, Temperature_measured, Capacity et Cycle, pour prédire la RUL (Durée de Vie Utile Restante).

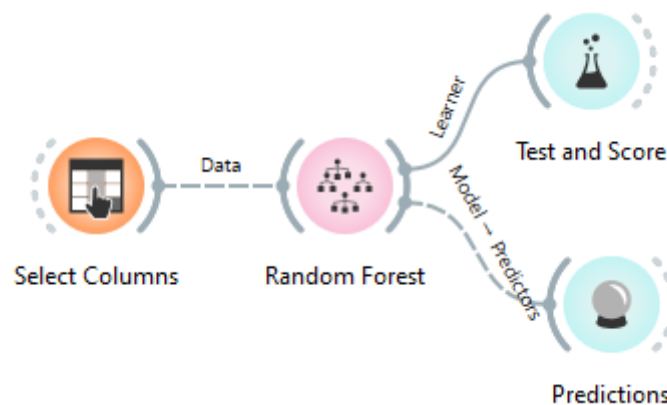


Figure3.11 : Flux de Données du Widget "Random Forest".

3.10 Widget réseau neuronal ("neural network")

Le widget Neural Network permet de créer un modèle basé sur un réseau de neurones artificiel, une technique d'apprentissage profond utilisée pour la régression ou la classification.

Ce modèle est inspiré du fonctionnement du cerveau humain et est capable de détecter des relations complexes entre les variables.

Name: Neural Network

Model parameters

Hidden layers: 20, 100, 50

Activation: ReLu

Solver: L-BFGS-B

Alpha: 0.0001

Max iterations: 2000

Replicable training: Yes

Data

Data instances: 168

Features: Voltage_measured, Current_measured, Temperature_measured, Capacity, Cycle

Target: RUL

Figure 3. 12 : Hyperparamètres du Modèle de Réseau Neuronal.

Ce modèle de réseau de neurones est conçu avec trois couches cachées contenant respectivement 20, 100 et 50 neurones, et utilise la fonction d'activation ReLU. Pour l'entraînement, il emploie l'optimiseur L-BFGS-B, avec une régularisation L2 (Alpha) de 0.0001 pour prévenir le surapprentissage, sur un maximum de 2000 itérations. De manière importante, l'entraînement est reproductible grâce à l'utilisation d'une graine aléatoire fixe. Ce réseau est alimenté par 168 instances de données, utilisant les caractéristiques Voltage_measured, Current_measured, Temperature_measured, Capacity et Cycle afin de prédire la RUL (Durée de Vie Utile Restante), ce qui correspond à une tâche de régression.

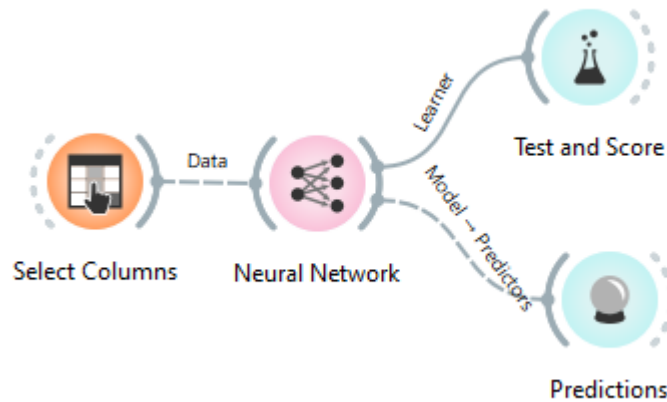


Figure3. 13 : Flux de Données du Widget "Réseau Neuronale".

3.11 Widget test et score ("test and score ")

Le widget Test and Score dans Orange est utilisé pour évaluer les performances des modèles d'apprentissage automatique. Il applique des techniques de validation (comme la validation croisée) pour mesurer la précision, l'erreur, la robustesse et la fiabilité des modèles.

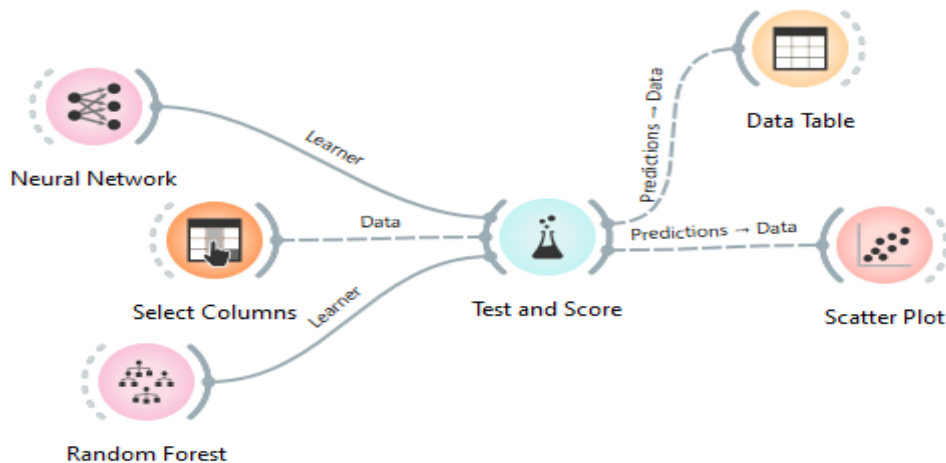


Figure3.14 : Flux de Données du Widget "Test and Score"

L'évaluation des modèles a été réalisée par validation croisée à 10 plis (10-fold Cross validation). Pour rappel, l'erreur quadratique moyenne (MSE), la racine de l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE) mesurent l'écart entre les prédictions et les valeurs réelles ; des valeurs plus faibles indiquent une meilleure performance. Le coefficient de détermination (R^2) indique la proportion de la variance de la variable cible expliquée

par le modèle ; une valeur proche de 1 est souhaitable.

Le modèle Forêt Aléatoire a obtenu une valeur "Test" de 0.070 (dont la métrique exacte n'est pas spécifiée ici mais est souvent RMSE par défaut dans Orange pour la régression), un MSE de 0.000, un RMSE de 0.018, un MAE de 0.012 et un R^2 de 0.996.

Le Réseau de Neurones a montré des performances globalement supérieures avec une valeur "Test" de 0.022, un MSE de 0.000, un RMSE de 0.009, un MAE de 0.006 et un R^2 de 0.999.

Ces résultats suggèrent que, pour cette tâche de prédiction de la RUL avec les configurations testées, le Réseau de Neurones offre une précision légèrement meilleure que la Forêt Aléatoire, comme en témoignent ses scores d'erreur plus faibles et son R^2 plus élevé.

3.12 Widget prédictions :

3.12.1 Fonction de l'outil :

L'outil Prédictions dans Orange permet d'afficher les résultats de prédiction des modèles de machine Learning sur chaque instance (ligne) de données. Il compare les valeurs réelles (par exemple, RUL) avec les valeurs prédites par différents modèles, comme Random Forest et Neural Network.



Figure 3.15 : Flux d'Analyse des Prédictions

Tableau3.4 : Tableau présentant les valeurs réelles et attendues du RUL résultant des modèles de prévision au sein d'Orange

Data instances: 30
Features: 5
Meta attributes: 5
Target: Numeric variable 'RUL'

	RUL	Random Forest	Random Forest (error)	Neural Network	Neural Network (error)	Selected	Voltage_measured	Current_measured	Temperature_measured	Capacity	Cycle
1	1	0.979498	0.0205015	1.00867	0.00866694	No	0.65753	0.44276	0.416766	1.000	0
2	0.996732	0.979956	0.0167756	0.988882	0.00784962	No	0.73239	0.44696	0.468062	0.98214	0.00326797
3	0.993464	0.967106	0.0263579	0.97856	0.014904	No	0.79650	0.45092	0.440428	0.96285	0.00653595
4	0.990196	0.981108	0.00908794	0.985629	0.00456713	No	0.79580	0.41738	0.397492	0.96270	0.00980392
5	0.986928	0.979567	0.00736116	0.987674	0.000746055	No	0.78258	0.41545	0.353033	0.96162	0.0130719
6	0.98366	0.982888	0.000772428	0.981769	0.0018911	No	0.77250	0.44972	0.370422	0.96340	0.0163399
7	0.980392	0.982888	0.00249555	0.97934	0.00105195	No	0.76941	0.44980	0.385932	0.96250	0.0196078
8	0.977124	0.969737	0.00738747	0.964872	0.012252	No	0.90038	0.34830	0.362464	0.94600	0.0228758
9	0.973856	0.967153	0.00670326	0.970564	0.00329198	No	0.88842	0.31281	0.340887	0.94427	0.0261438
10	0.970588	0.969548	0.00104066	0.972006	0.00141752	No	0.87113	0.31422	0.317624	0.94399	0.0294118
11	0.96732	0.966057	0.00126338	0.977391	0.0100708	No	0.85201	0.27816	0.277467	0.94400	0.0326797
12	0.962418	0.951246	0.0111726	0.951528	0.0108903	No	0.95660	0.24283	0.346902	0.92569	0.0375817
13	0.95915	0.946584	0.0125659	0.951533	0.00761758	No	0.93984	0.20731	0.331512	0.92490	0.0408497
14	0.955882	0.950984	0.00489876	0.956236	0.000353548	No	0.91596	0.20758	0.277466	0.92435	0.0441176
15	0.952614	0.941684	0.01093	0.941634	0.0109803	No	0.98645	0.21032	0.239551	0.90530	0.0473856
16	0.949346	0.939982	0.00936393	0.941109	0.00823755	No	0.97351	0.17203	0.201818	0.90443	0.0506536
17	0.946078	0.936685	0.00939299	0.93892	0.00715801	No	0.96242	0.13307	0.205194	0.90526	0.0539216
18	0.94281	0.927599	0.0152117	0.933454	0.00935649	No	0.95020	0.09407	0.213028	0.90612	0.0571895
19	0.939542	0.927599	0.0119437	0.930625	0.00891736	No	0.93789	0.09476	0.213442	0.90561	0.0604575
20	0.934641	0.94664	0.0119995	0.955712	0.0210714	No	0.87944	0.30776	0.349372	0.98337	0.0653595
21	0.928105	0.937356	0.00925178	0.939148	0.0110435	No	0.88473	0.23645	0.459680	0.98406	0.0718954
22	0.921569	0.930595	0.00902686	0.932391	0.0108226	No	0.92697	0.23967	0.333011	0.96431	0.0784314
23	0.915033	0.923859	0.0088261	0.920926	0.00589314	No	0.97309	0.20516	0.220727	0.94604	0.0849673
24	0.908497	0.911546	0.00304962	0.919785	0.0112879	No	0.92514	0.24183	0.189360	0.94486	0.0915033
25	0.901961	0.888004	0.0139567	0.90737	0.00540928	No	0.94575	0.09182	0.151096	0.94569	0.0980392
26	0.895425	0.892068	0.00335652	0.897689	0.00226424	No	1.000	0.13200	0.120351	0.92539	0.104575
27	0.888889	0.884965	0.00392387	0.890968	0.00207949	No	0.95489	0.09366	0.089488	0.92669	0.111111
28	0.882353	0.878134	0.00421868	0.886708	0.00435494	No	0.92736	0.09306	0.079034	0.92528	0.117647
29	0.875817	0.881541	0.00572368	0.875476	0.000341086	No	0.98339	0.13373	0.087642	0.90559	0.124183
30	0.869281	0.871916	0.00263489	0.869523	0.000242203	No	0.95042	0.09512	0.091748	0.90790	0.130719

3.12.2 Valeurs réelles et prédites du RUL

Le Tableau 3.4 fournit une comparaison détaillée, instance par instance, entre les valeurs réelles de la RUL et les valeurs prédites par les deux modèles évalués : Forêt Aléatoire (Random Forest) et Réseau de Neurones (Neural Network). Pour chaque modèle, l'erreur de prédiction est également calculée, offrant une mesure quantitative de la performance au niveau de chaque point de données. Ce type de tableau est crucial pour une analyse granulaire des résultats, permettant d'identifier si certains modèles performant mieux sur des plages spécifiques de RUL ou pour certaines caractéristiques d'instances (visibles dans les colonnes Voltage_measured, Current_measured, etc.). Il sert de base à une évaluation qualitative de la robustesse et de la précision des modèles sur l'ensemble de test."

3.13 Visualisation des Résultats

3.13.1 Comparaison graphique des prédictions RUL :

La **Figure 3.16** illustre la comparaison entre la RUL réelle (ligne continue) et les prédictions de RUL obtenues par les modèles de Forêt Aléatoire (ligne pointillée orange) et de Réseau de Neurones (ligne pointillée bleue), en fonction du cycle normalisé. Visuellement, les deux modèles semblent suivre de très près la tendance de la RUL réelle sur l'ensemble de la plage des cycles. Le Réseau de Neurones montre un ajustement quasi parfait, tandis que la Forêt Aléatoire présente des fluctuations minimales autour de la valeur réelle. Ce graphique suggère une excellente capacité des deux modèles à capturer la dynamique de dégradation de la batterie, avec un léger avantage apparent pour le Réseau de Neurones en termes de lissage de la prédiction."

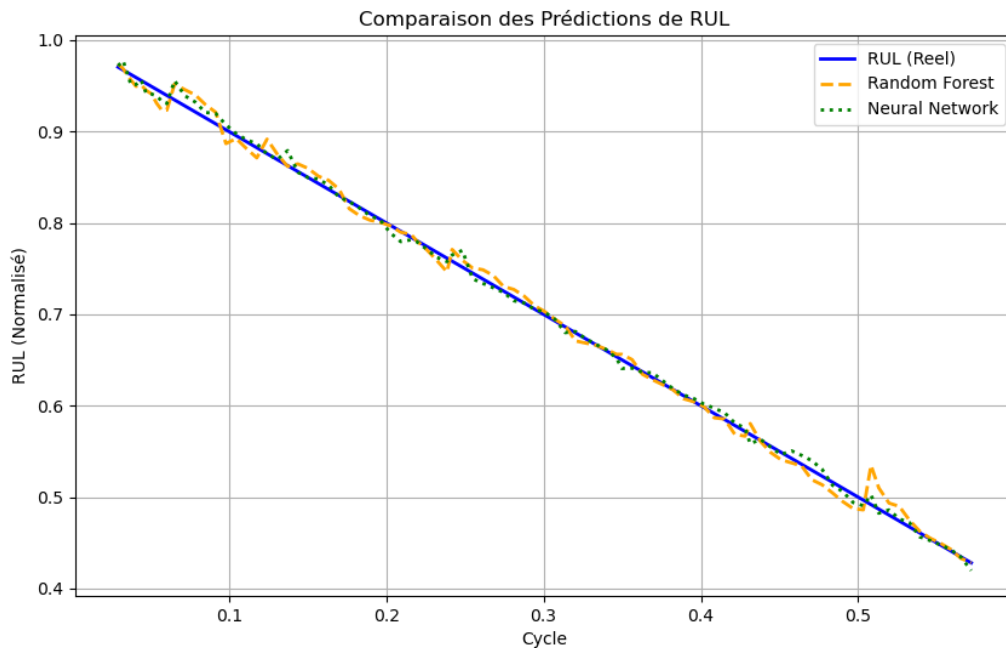


Figure3.16 : Comparaison des prédiction (RUL Réelle - RUL Prédite)

3.13.2 Comparaison des erreurs de prédiction

La **Figure 3.17** présente la distribution des erreurs de prédiction (RUL Réelle - RUL Prédite) pour les modèles de Forêt Aléatoire (en orange) et de Réseau de Neurones (en vert) sur l'ensemble des cycles normalisés. L'erreur nulle est représentée par une ligne bleue horizontale. Il est observé que les erreurs pour les deux modèles fluctuent autour de zéro, indiquant l'absence de biais systématique important. Les erreurs du Réseau de Neurones semblent avoir une amplitude de variation légèrement plus faible et être plus centrées autour de zéro compa-

rées à celles de la Forêt Aléatoire, qui montre occasionnellement des pics d'erreur plus prononcés. Cette visualisation confirme la haute performance des deux approches, tout en suggérant une stabilité marginalement supérieure pour le Réseau de Neurones."

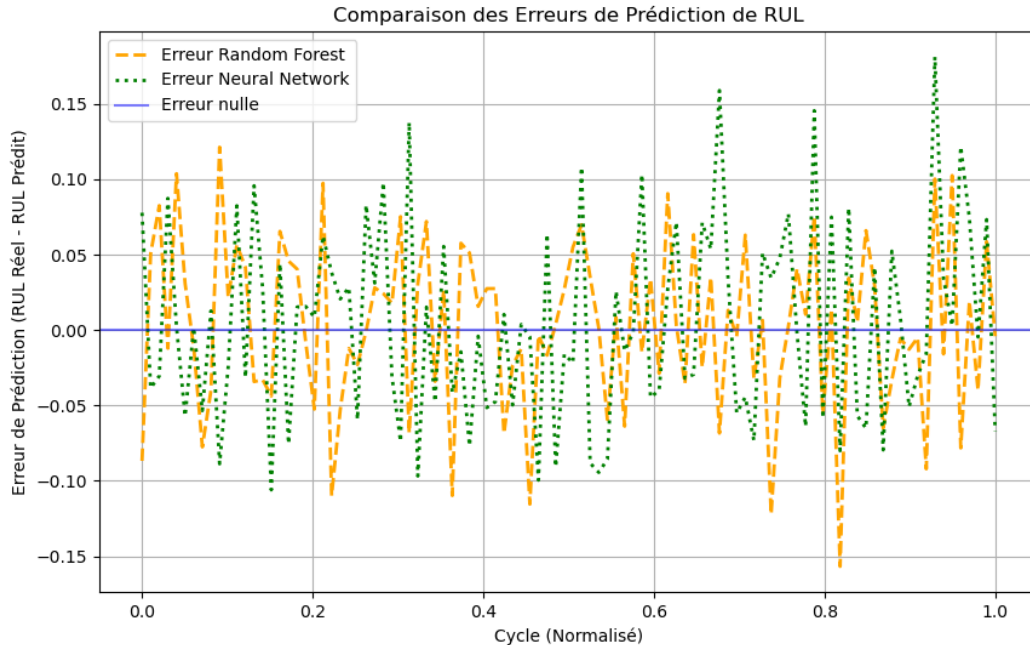


Figure3.17 : Distribution des erreurs de prédiction (RUL Réelle - RUL Prédite)

3.14 Conclusion

Au terme de ce travail, nous pouvons affirmer que les objectifs de l'étude ont été atteints à travers le développement d'un modèle prédictif précis de la durée de vie utile restante (RUL) des batteries lithium-ion industrielles, en s'appuyant sur les techniques d'apprentissage automatique et leur mise en œuvre via la plateforme visuelle Orange. L'étude a débuté par une analyse approfondie des propriétés physico-chimiques des batteries, et s'est poursuivie par l'application de modèles tels que les forêts aléatoires et les réseaux de neurones pour traiter les données de performance et en extraire des indicateurs de dégradation.

Les résultats expérimentaux ont mis en évidence l'efficacité des modèles adoptés, en particulier en termes de précision et de fiabilité, soulignant ainsi le potentiel considérable de l'intelligence artificielle dans le domaine de la maintenance prédictive et de la gestion du cycle de vie des batteries. La comparaison des modèles a également montré que le choix de l'algorithme le plus adapté doit tenir compte de la nature des données disponibles et du contexte d'application.

Malgré les avancées réalisées, plusieurs perspectives de recherche restent ouvertes, no-

tamment l'intégration de données temporelles plus variées, ou le développement de modèles hybrides tirant parti des atouts de chaque approche. Il serait également pertinent d'explorer à

L'avenir des techniques d'apprentissage plus avancées, telles que l'apprentissage profond ou par renforcement, afin d'améliorer encore davantage la précision des prédictions dans des environnements industriels réels.

Conclusion Générale

Conclusion Générale

Au terme de cette investigation académique, nous avons abordé une problématique centrale dans le domaine du stockage d'énergie contemporain : l'estimation prédictive de la durée de vie opérationnelle restante (RUL) des batteries lithium-ion. Ce défi a été appréhendé par l'application de méthodologies de modélisation fondées sur les données (*data-driven modeling*), implémentées au sein d'un environnement de prototypage visuel et intuitif tel qu'Orange, conjuguant ainsi la rigueur théorique à l'applicabilité pratique.

Les résultats empiriques issus de nos expérimentations ont démontré la supériorité des modèles architecturés sur des réseaux de neurones en termes de précision prédictive, comparativement aux approches basées sur les forêts aléatoires. Cette performance accrue s'explique notamment par leur capacité intrinsèque à capturer les interdépendances complexes et les non-linéarités inhérentes aux variables caractérisant le processus de dégradation des batteries. Ces constatations soulignent l'efficacité des méthodes de modélisation inférentielle pour le diagnostic de l'état de santé (SOH) des systèmes de stockage électrochimique et pour une gestion proactive de leur cycle de vie.

L'utilisation de la plateforme Orange a significativement facilité les phases d'expérimentation, d'exploration et d'analyse des données, grâce à son interface graphique flexible et à son accessibilité. Le recours à un jeu de données de référence, validé et mis à disposition par la NASA, a conféré une crédibilité substantielle aux résultats obtenus, renforçant leur validité pour des applications en contexte opérationnel réel.

Ce travail ouvre, par ailleurs, des perspectives de recherche prometteuses. Il serait pertinent d'enrichir les modèles prédictifs par l'intégration de variables exogènes supplémentaires, telles que des données thermiques détaillées ou des profils de décharge dynamiques et complexes. L'exploration d'architectures neuronales plus avancées, notamment les réseaux de neurones récurrents (RNN) et leurs variantes comme les LSTM (Long Short-Term Memory) ou les GRU (Gated Recurrent Units), spécifiquement conçues pour le traitement des séries temporelles, pourrait également conduire à des améliorations significatives. De plus, l'adaptation des modèles aux formats de cellules spécifiques, tels que les cellules cylindriques 21700, ainsi qu'à de nouvelles chimies, comme les batteries sodium-ion, permettrait d'élargir le champ d'application. Une généralisation de l'approche proposée à d'autres chimies de batteries et à des conditions d'utilisation variées constituerait une étape ultérieure cruciale pour en accroître la robustesse et la polyvalence.

Pour approfondir ces perspectives, l'intégration des cellules cylindriques 21700 dans les modèles prédictifs pourrait répondre aux besoins croissants de densité énergétique et d'efficacité dans des applications comme les véhicules électriques et les systèmes de stockage stationnaire. Ces cellules, plus grandes que les formats 18650, offrent une capacité accrue, mais leur comportement thermique et électrochimique nécessite des ajustements spécifiques dans les algorithmes de surveillance. Parallèlement, les batteries sodium-ion, en tant qu'alternative durable aux batteries lithium-ion, présentent des avantages en termes de disponibilité des matériaux et de coût, mais leur performance à long terme reste à optimiser. L'adaptation des modèles à ces nouvelles chimies requerrait une analyse approfondie des cycles de charge et de décharge, ainsi que des interactions électrochimiques spécifiques au sodium. En outre, l'utilisation de techniques d'apprentissage automatique comme les réseaux transformers, qui excellent dans la capture des dépendances à long terme dans les séries temporelles, pourrait améliorer la précision des prédictions pour ces technologies émergentes. Une collaboration interdisciplinaire avec des experts en science des matériaux serait essentielle pour valider les modèles dans des conditions réelles. Enfin, l'exploration de scénarios d'utilisation extrêmes, comme des températures très basses ou élevées, permettrait de tester la robustesse des modèles et d'assurer leur fiabilité dans des contextes variés. Ces avancées contribueraient à la transition vers des systèmes énergétiques plus durables et performants, en phase avec les objectifs globaux de décarbonation.

References bibliographies

- [1] B. Scrosati and J. Garche, "Lithium batteries: Status, prospects and future," *Journal of Power Sources*, vol. 195, no. 9, pp. 2419–2430, 2010.
- [2] M. Broussely and G. Pistoia (Eds.), **Industrial Applications of Batteries: From Cars to Aerospace and Energy Storage**, Elsevier, 2007.
- [3] J. M. Tarascon and M. Armand, "Issues and challenges facing rechargeable lithium batteries," *Nature*, vol. 414, no. 6861, pp. 359–367, 2001.
- [4] D. Anseán, M. González, J. C. Viera, V. M. García, and C. Blanco, "Fast charging technique for high power LiFePO₄ batteries: A cycle life analysis," *Journal of Power Sources*, vol. 239, pp. 9–15, 2013.
- [5] C. Zhang, K. Li, J. Deng, and S. Song, "Improved battery management system to estimate the state of health of lithium-ion battery based on genetic algorithm," *Energy*, vol. 195, p. 116947, 2020.
- [6] M. Bercibar, I. Gandiaga, I. Villarreal, N. Omar, P. Van den Bossche, and J. Van Mierlo, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 572–587, 2016.
- [7] Y. Wang, S. Wang, and L. Yao, "Remaining useful life prediction of lithium-ion batteries based on health index and Bayesian deep learning," *Reliability Engineering & System Safety*, vol. 217, p. 107961, 2022.
- [8] NASA Ames Prognostics Data Repository, "Battery Data Set B0005," [Online]. Available: <https://www.nasa.gov/prognostics/data-set>. [Accessed: June 2025].
- [9] T. Waldmann, M. Wilka, M. Kasper, M. Fleischhammer, and M. Wohlfahrt-Mehrens, "Temperature dependent ageing mechanisms in Lithium-ion batteries – A Post-Mortem study," *Journal of Power Sources*, vol. 262, pp. 129–135, 2014.
- [10] V. Srinivasan and J. Newman, "Discharge model for the lithium iron-phosphate electrode," *Journal of the Electrochemical Society*, vol. 151, no. 10, pp. A1517–A1529, 2004.

- [11] Simtec Solution. (s.d.). Modèle thermo-électrique d'accumulateurs Lithium-Ion avec COMSOL Multiphysics.
- [12] Lithium Battery Tech. (s.d.). Lithium Battery Life Cycle.
- [13] Tycorun. (n.d.). What is lithium-ion battery? How do lithium-ion batteries work? Tyco-run Energy. Retrieved May 7, 2025, from
- [14] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- [15] Scikit-learn developers. (n.d.). Random forests — scikit-learn documentation. Retrieved April 10, 2025, from
- [16] Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., ... & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5), 383-391.
- [16] Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, 357, 209-219.
- [17] Hu, X., Xu, L., Lin, X., & Pecht, M. (2020). Battery lifetime prognostics. *Joule*, 4(2), 310-346.
- [18] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- [19] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [20] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [21] Zhang, Y., Tang, Q., Zhang, Y., Wang, J., & Wang, U. (2018). Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nature Communications*, 9(1), 1-10.
- [22] Chemali, E., Kollmeyer, P. J., Preindl, M., & Emadi, A. (2018). State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach. *Journal of Power Sources*, 400, 242-255.

- [23] Xue, Y., Bai, Y., & Wang, H. (2020). Machine learning for advanced energy materials. *Energy and AI*, 1, 100014.
- [24] Zhang, S., & Zhang, C. (2019). A review of machine learning for modeling and simulation of battery materials. *Journal of Power Sources*, 413, 259-275.