

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A.Mira de Bejaia
Faculté des sciences exactes
Département d'Informatique



Mémoire de fin de cycle

*En vue d'obtention du diplôme de Master recherche en Informatique.
Spécialité : Systèmes d'Information Avancés.*

Thème

**Modèle proactif de gestion de la maintenance industrielle
à base d'une démarche d'analyse des données.
Cas d'étude : Groupe Agroalimentaire Cevital**

Réalisé par :

Mlle. AMRANI Lina et Mlle. YAICI Sara.

Soutenue le 03/04/2024 devant le jury composé de :

Président	Dr. FARAH Zoubeyr	U. A/Mira Béjaïa
Examineur	Dr. BENNAI Yani-Athmane	U. A/Mira Béjaïa
Encadrants	Pr. AMROUNE Kamal	U. A/Mira Béjaïa
	Dr. EL SAKAAN Nadim	U. A/Mira Béjaïa

Année universitaire 2023/2024

Remerciements

Nous souhaitons exprimer notre profonde reconnaissance et nos chaleureux remerciements à nos directeurs de mémoire,

Dr. El SAKAAN Nadim et Pr. AMROUNE Kamel, pour leur expertise, leurs précieuses orientations, leurs conseils éclairés et leur soutien constant tout au long de ce travail.

Nos remerciements vont aussi à notre maître de stage, M. RECHAM Hassan, pour ses orientations précieuses, ainsi qu'à tout le personnel de Cevital.

Nous remercions également les membres du jury d'avoir accepté d'examiner et d'évaluer notre travail.

Nous exprimons notre gratitude à nos camarades de classe qui ont partagé leurs idées et expériences avec nous. Leurs discussions stimulantes ont enrichi notre réflexion et nous ont permis d'explorer de nouvelles perspectives.

Enfin, nous sommes profondément reconnaissants à nos familles et amis pour leur soutien indéfectible.

Dédicaces

À mes chers parents, havres de paix et piliers inébranlables,

Bercés par votre amour inconditionnel et nourris par vos sacrifices sans bornes, c'est avec une immense reconnaissance que je vous dédie ce mémoire. Votre soutien indéfectible a été mon trésor le plus précieux, me propulsant vers la réussite tout au long de ce parcours. Votre foi en mes capacités est restée inébranlable même lorsque j'en doutais moi-même. Je vous en suis infiniment reconnaissante.

À mon frère bien-aimé,

Confident et ami précieux, dont le soutien constant, les encouragements sincères et la présence bienveillante ont été une source inestimable de motivation.

À mes amis qui m'ont accompagné dans les moments de doute et de joie.

À tous ceux qui ont contribué à ma trajectoire académique et professionnelle.

Je dédie ce travail, porteur de leurs précieux conseils.

Lina

Dédicaces

En ce moment précieux où j'achève ce mémoire,
je souhaite dédier ces mots empreints de gratitude et d'affection :

*À mon très chère père, mon modèle dans la vie, ton soutien inébranlable et ton affection constante ont été
essentiels à ma réussite. Je te suis profondément reconnaissante.*

Merci du fond du cœur pour tout ce que tu fais pour moi.

*À ma très chère mère, ma source de tendresse, ton affection et ton support indéfectibles ont été essentiels à mon
parcours. Mes plus sincères remerciements.*

À mon petit frère, Salas, mon compagnon de joie.

À ma petite sœur, Sandra, ma boule de bonheur.

À ma meilleure amie, Farah, ma sœur de cœur.

Sara

Table des matières

Introduction générale	1
I Concepts théoriques et contexte du projet	3
1 Introduction	4
2 Les systèmes d'informations digitaux et la maintenance des pannes	4
2.1 Définition d'un système d'information	4
2.2 Les missions d'un système d'informations	5
2.3 Les composants et les technologies des systèmes d'informations	6
2.4 Les applications métiers	6
2.4.1 Planification des ressources de l'entreprise	7
2.4.2 Gestion de la maintenance assistée par ordinateur	8
2.4.3 Gestion de la production assistée par ordinateur	8
3 La data science	9
3.1 Définition de la data science	9
3.2 La démarche data sciences	9
3.3 Les techniques de la data science	11
3.3.1 Apprentissage supervisé	11
3.3.2 Apprentissage non supervisé	13
3.3.3 Apprentissage Profond (Apprentissage arofond (DL))	14
3.3.4 Analyse de séries temporelles	18
3.3.5 Traitement du langage naturel (NLP)	18
3.4 La mise en place d'une stratégie de gestion appuyée sur la donnée (data-driven-decision)	19
3.4.1 Définition de la stratégie de gestion appuyée sur la donnée	19
3.4.2 Les étapes du processus de la mise une place de cette stratégie	19
3.4.3 Les avantages de la stratégie data-driven :	20
4 Cas d'étude : Groupe agroalimentaire Cevital	21
4.1 Historiques	21
4.2 Emplaçemant géographique	21
4.3 Les filiations de Cevital	22
4.4 L'organigramme de Cevital	22
4.5 Gestion de la Maintenance et Techniques Associées dans le Groupe Cevital	23
5 Conclusion	24
II État de l'art des méthodes de maintenance proactive industrielle assistée par ordinateur	25
1 Introduction	26
2 Taxonomie des méthodes de maintenance	26

Table des matières

2.1	La maintenance réactive (RM)	27
2.2	La maintenance proactive (PRM)	28
2.2.1	La maintenance préventive (PM)	28
2.2.2	La maintenance prédictive (PDM)	30
3	Méthodes prédictives basées sur l'apprentissage automatique (ML)	32
3.1	Les travaux basés sur l'algorithme de l'Arbre de décision (DT)	32
3.2	Les travaux basés sur l'algorithme des machines à vecteurs de support (SVM)	32
3.3	Les travaux basé sur l'algorithme des K plus proches voisins (K-NN)	32
4	Méthodes prédictives basées sur l'apprentissage profond (DL)	33
4.1	Les travaux basé sur L'algorithme de l'Auto-encodeur (AE)	33
4.2	Les travaux basé sur l'algorithme des Réseaux de neurones convolutifs (CNN)	34
4.3	Les travaux basé sur l'algorithme des réseaux de neurones récurrents (RNN)	34
4.4	Les Méthodes hybrides	35
5	Analyse comparative	36
6	Conclusion	40
III Solution proposée		41
1	Introduction	42
2	Analyse de données Cevital	42
2.1	Présentation des données de Cevital :	42
2.2	Visualisation des tendances entre les donnée	44
2.3	Défis rencontrés	48
3	Approche proposée	49
3.1	Hypothèse de la solution	49
3.2	Les Jeux de données utilisées	49
3.3	La solution proposée	52
3.3.1	Infrastructure de Développement et Ressources Techniques	52
3.3.2	Méthodologie	54
4	Conclusion	64
IV Validation et expérimentations		66
1	Introduction	67
2	Evaluation des performances	67
2.1	Les métriques d'évaluation	67
2.2	Résultats d'évaluation pour le dataset 1	68
2.3	Résultats d'évaluation pour le dataset 2	75
2.4	Étude comparative	80
3	Conclusion	82
Conclusion et perspectives		83
Bibliographie		
.....		

Table des figures

I.1	Les composants d'un système d'information	5
I.2	Qu'est-ce qu'un Progiciel de Gestion Intégré (ERP)? [1]	7
I.3	Les étapes d'un projet Data science	10
I.4	Structure d'un perceptron	14
I.5	Structure d'un ANN [2]	15
I.6	Structure d'un RNN [3]	16
I.7	Cellule d'un LSTM	17
I.8	Etapas historiques du groupe Cevital	21
I.9	Organigramme de Cevital	23
II.1	Taxonomie des méthodes de maintenance	27
II.2	Les types de maintenance industriels [4]	28
II.3	Courbe statistique en baignoire d'un équipement [4]	29
III.1	Aperçu des rapports GMAO de Cevital	43
III.2	Le nombre d'équipements pour chaque entité	45
III.3	Le nombre d'interventions sur les équipements sur quatre années	46
III.4	La somme des interventions effectuées par entité	46
III.5	le volume d'interventions moyennes par équipement pour chaque entité	47
III.6	Le nombre d'interventions	48
III.7	Le jeu de données de Télémétrie	51
III.8	Le jeu de données Filtre	52
III.9	Méthodologie de la solution proposée	55
III.10	Visualisation de la relation entre la variable cible (RUL) et une caractéristique temporelle	56
III.11	Visualisation de la relation entre la target (RUL) en fonction d'une contrainte mécanique	57
III.12	Visualisation de la relation entre la (RUL) en fonction d'une caractéristique de grandeur physique des datasets	58
III.13	Visualisation de la relation entre la (RUL) en fonction d'une caractéristique de dégradation	58
III.14	Visualisation des matrices de corrélation des datasets	60
IV.1	Graphe visualisant la target (RUL) du dataset 1	69
IV.2	Les courbes des pertes d'entraînement et de test au fil des époques pour le dataset 1	70
IV.3	Visualisation de la précision des prédictions pour le dataset 1	71

Table des figures

IV.4	L'histogramme visualisant la distribution des erreurs de prédiction pour le dataset 1	72
IV.5	Tableau des valeurs réelles et prédites de la Durée de vie restante utile (RUL) pour l'ensemble de test du dataset 1	72
IV.6	Visualisation des valeurs réelles et prédites de la (RUL) du dataset 1	74
IV.7	La dispersion des valeurs réelles vs prédites de l'ensemble de test du dataset 1 .	75
IV.8	Les courbes des pertes d'entraînement et de test au fil des époques pour le dataset 2	76
IV.9	Visualisation de la précision des prédictions pour le dataset 2	77
IV.10	L'histogramme visualisant la distribution des erreurs de prédiction pour le dataset 2	78
IV.11	Graphe des résidus du dataset 2	78
IV.12	Tableau des valeurs réelles et prédites de la RUL pour l'ensemble de test du dataset 2	79
IV.13	Le nuage de points de dispersion des valeurs réelles vs prédites pour l'ensemble de test du dataset 2	79

Liste des tableaux

II.1	Analyse comparative des études synthétisées.	38
III.1	Grilles de recherche des hyperparamètres pour les deux datasets	62
III.2	Les meilleurs hyperparamètres trouvés pour les deux datasets	64
IV.1	Résultats des évaluations du dataset 1	68
IV.2	Résultats d'évaluations du dataset 2	75
IV.3	Résultats des différents modèles	80

Liste des abréviations

AE Auto-encodeur
ANN Réseaux de neurones artificiels
ARIMA Modèle Autorégressif Intégré à Moyenne Mobile
BILSTM Mémoire à Long Terme et Court Terme Bidirectionnels
C-MAPSS Simulation de Système Aéro-propulsif Modulaire Commercial
CNN Réseaux de neurones convolutifs
CRM Gestion de la relation client
DBSCAN Regroupement spatial basé sur la densité avec bruit
DL Apprentissage arofond
DNN Réseaux de neurones profonds
DT Arbre de décision
ERP Progiciel de Gestion Intégré
ETS Modèle de Tendence, Saison et Erreur
GAN Réseaux de neurones génératifs adversariaux
GB Gradient Boosting
GMAO Gestion de la Maintenance Assistée par Ordinateur
GPAO Gestion de la Production Assistée par Ordinateur
IA Intelligence Artificielle
IoT Internet des objets
K-NN K-plus proches voisins
LSTM Mémoire à Long Terme et Court Terme
MAE Erreur absolue moyenne
MSE Erreur quadratique moyenne
ML Apprentissage automatique
NLP Traitement du langage naturel
OSA-CBM Architecture Système Ouverte pour la Surveillance Basée sur l'État
PM Maintenance préventive
PDM Maintenance prédictive
PRM Maintenance Proactive
RF Forêt aléatoire
RM Maintenance Réactive
RMSE Racine carré de l'erreur Quadratique Moyenne
RNN Les réseaux de neurones récurrents
RUL Durée de vie restante utile
SF Fonction score
SI Système d'information
SOM Réseaux de neurones auto-organisés
SVM Machines à vecteurs de supports
TL Transfert par apprentissage

Introduction générale

L'évolution des systèmes d'informations a été radicalement transformée par l'essor des outils digitaux. Autrefois, les systèmes d'information étaient principalement des bases de données statiques et des processus manuels. Aujourd'hui, grâce aux avancées numériques, ces systèmes sont devenus dynamiques, interactifs et intégrés. Les technologies comme le cloud computing, l'intelligence artificielle et les big data permettent une gestion plus efficace et plus rapide des informations. Les entreprises peuvent désormais collecter, analyser et utiliser les données en temps réel pour prendre des décisions éclairées. L'intégration des outils digitaux dans les systèmes d'information améliore non seulement la productivité, mais également la précision et la fiabilité des données, facilitant ainsi la transformation numérique des organisations [5].

Cette transformation numérique s'étend également à la gestion de la maintenance, un domaine crucial pour les entreprises industrielles. Le numérique modernise et optimise les processus de maintenance en intégrant des technologies telles que l'Internet des Objets (IoT), les capteurs intelligents et les systèmes de gestion de maintenance assistée par ordinateur (GMAO). Ces outils permettent une surveillance continue des équipements, la détection précoce des signes de pannes et une intervention avant qu'elles ne se produisent, réduisant ainsi les temps d'arrêt et les coûts de réparation. Les informations recueillies facilitent la mise en place de stratégies de maintenance préventive et prédictive, transformant la maintenance en un processus plus proactif [6].

La maintenance proactive est essentielle dans les environnements industriels modernes. Contrairement à la maintenance réactive qui intervient après la défaillance d'un équipement, peut durer des heures, voire des jours, perturbant ainsi la chaîne de production, la maintenance proactive anticipe les pannes avant qu'elles ne surviennent, avec la maintenance préventive remplaçant les équipements par crainte de perturbations, et la maintenance prédictive utilisant des techniques d'intelligence artificielle pour anticiper les pannes avant leur apparition [7].

La planification de la maintenance assistée par les données représente une avancée significative dans la gestion des actifs industriels. En collectant et en analysant des données issues de diverses sources, telles que les capteurs IoT, les historiques de maintenance des (GMAO) et les conditions opérationnelles, les entreprises peuvent élaborer des plans de maintenance plus précis et efficaces. Les outils analytiques et les algorithmes de l'apprentissage automatique et profond permettent de prédire les pannes d'équipement et d'optimiser les calendriers de maintenance. Cela signifie que les interventions peuvent être programmées de manière à minimiser les interruptions de production et à maximiser l'efficacité opérationnelle. De plus, l'utilisation des données permet d'allouer les ressources de maintenance de manière plus judicieuse, réduisant ainsi les coûts et améliorant la rentabilité globale.

Dans le cadre de la transformation numérique de la maintenance industrielle, une question cruciale se pose pour les entreprises et les usines : "Comment la mise en place d'une stratégie de maintenance prédictive peut-elle réduire significativement les coûts et les temps d'arrêt dans les environnements industriels par rapport aux approches de maintenance cor-

rective et préventive traditionnelles?"

La solution mise en œuvre repose sur l'utilisation de la maintenance prédictive, une approche proactive qui anticipe les pannes des équipements industriels grâce à l'analyse avancée des données et aux techniques d'intelligence artificielle. Les modèles basés sur les (RNN) sont utilisés pour modéliser des séquences de données temporelles. Cependant, les (RNN) traditionnels présentaient des limitations, notamment la difficulté à capturer efficacement les dépendances à long terme dans les données séquentielles. Pour surmonter ces défis, l'arrivée des (LSTM) a marqué un tournant significatif. Ces derniers sont une variante de (RNN) capable de gérer les problèmes de disparition et d'explosion du gradient, tout en conservant la capacité à mémoriser des informations sur de longues périodes.

En appliquant un (LSTM) à deux jeux de données différents, comprenant des informations historiques sur les performances des équipements et les incidents de pannes antérieurs, des résultats prometteurs ont été obtenus. Ces résultats illustrent l'efficacité accrue des (LSTM) dans l'anticipation proactive des défaillances potentielles des équipements industriels par rapport aux approches antérieures basées sur les (RNN) classiques.

L'apport majeur de ce travail réside dans la démonstration de l'efficacité de la maintenance prédictive basée sur l'apprentissage profond dans un contexte industriel. En appliquant un modèle (LSTM) avec ajustement des hyper-paramètres sur deux jeux de données réels, cette solution a montré que l'analyse avancée de données peut surpasser les approches traditionnelles de maintenance corrective et préventive. En anticipant les pannes avec une précision accrue, cette solution permet de réduire les temps d'arrêt non planifiés et les coûts associés aux interruptions de la production, tout en optimisant la durée de vie des équipements. Cette contribution met en lumière l'importance des technologies d'intelligence artificielle pour améliorer la fiabilité et l'efficacité des opérations industrielles.

Ce manuscrit est structuré en quatre chapitres.

- Le chapitre 1 se concentre sur les concepts théoriques et le contexte du projet. Nous y présenterons les systèmes d'information digitaux, suivis d'une introduction à la science des données et aux techniques d'intelligence artificielle. De plus, nous y décrirons l'entreprise qui nous a accueillis dans le cadre de ce projet.
- Le chapitre 2 aborde le concept de maintenance industrielle en se basant sur un état de l'art. Nous y présentons les différentes approches de maintenance prédictive des pannes et proposons une analyse comparative des travaux existants.
- Le chapitre 3 expose la solution proposée pour répondre à la problématique, en détaillant les différents outils nécessaires à sa mise en œuvre, ainsi que les logiciels et matériels utilisés. Nous expliquerons ensuite les étapes suivies lors de l'implémentation de la solution.
- Le chapitre 4 est consacré à l'évaluation et à la discussion des résultats obtenus. Enfin, ce manuscrit se termine par une conclusion générale.

Chapitre I

Concepts théoriques et contexte du projet

1 Introduction

Les entreprises s'appuient sur des systèmes d'informations robustes pour stimuler leurs croissances. La digitalisation de ces systèmes joue un rôle central dans l'optimisation des opérations. Parallèlement, la Maintenance préventive (PM) basée sur l'analyse de données devient essentielle pour assurer la continuité des processus industriels. C'est dans cette synergie entre le numérique et l'ingénierie proactive que la data science se révèle être un outil puissant, offrant des solutions avancées pour prédire, prévenir et résoudre les incidents de manière plus efficace.

Dans ce chapitre, nous commencerons par examiner les systèmes d'information en tant que piliers des entreprises. Ensuite, nous explorerons la démarche de la science des données, en détaillant ses différentes techniques, y compris le Apprentissage automatique (ML) et DL. Nous aborderons également la mise en place d'une stratégie de décision basée sur les données (data-driven decision). Enfin, nous parlerons de l'organisme d'accueil qui nous a accompagnés le long de ce projet, en mettant l'accent sur l'optimisation des opérations industrielles.

2 Les systèmes d'informations digitaux et la maintenance des pannes

Les systèmes d'informations digitaux sont essentiels pour le fonctionnement des entreprises modernes, notamment en ce qui concerne la gestion et la maintenance des pannes.

2.1 Définition d'un système d'information

Un Système d'information (SI) est un agencement d'agents intelligents et les techniques de technologies de l'information qui sont en mesure de recueillir, diffuser et utiliser des données, des informations et des connaissances dans une organisation dans le but d'optimiser le flux d'informations et la réalisation de ses multiples objectifs [5].

Un SI est conçu pour collecter, traiter, stocker et distribuer des informations, permettant une gestion efficace de l'information. Il est fondamentalement un système socio-technique, composé de trois sous-systèmes : Le sous-système technique englobant la technologie et les processus à savoir :

- La technologie : Les technologies de l'information comprennent le matériel, les logiciels et les équipements de télécommunication utilisés pour le traitement de l'information.
- Les processus : Un processus désigne l'ensemble des actions qu'un individu, un groupe ou l'organisation doit mettre en œuvre dans la réalisation d'une tâche

Le sous-système social englobant les personnes et la structure à savoir :

- Les personnes : La composante humaine d'un SI regroupe toutes les personnes qui sont directement impliquées dans le système.
- La structure : La composante structure d'un SI se réfère à la relation entre les individus dans la composante humaine.

Et le sous-système métier :

- ce sous-système représente l'aspect fonctionnel et opérationnel du SI. Il est responsable de fournir les outils nécessaires à l'efficacité de l'optimisation des processus clés de l'entreprise en contribuant ainsi à la création de valeur.

La performance optimale du SI exige le fonctionnement synchronisé de ces trois composantes [8].

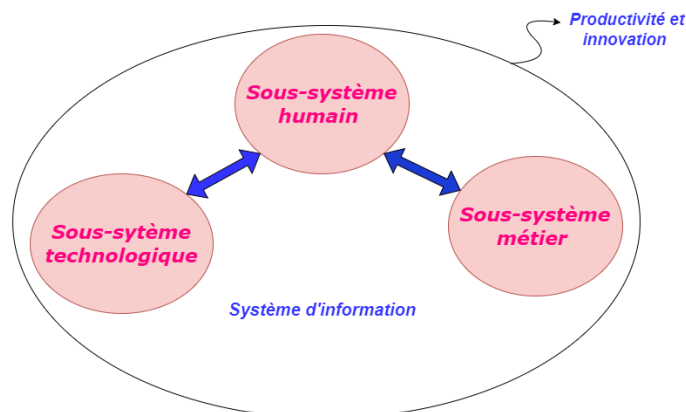


FIGURE I.1 – Les composants d'un système d'information

2.2 Les missions d'un système d'informations

La gestion de l'information a une grande importance dans les organisations, et cela est dû à leurs grande utilité qui se présentent comme suit [5] :

- Collecte de l'information : Le SI possède la capacité de recueillir des données provenant de différentes sources, qu'elles soient internes ou externes à l'organisation. La collecte de données permet d'obtenir une vision complète de l'environnement dans lequel évolue la structuration ainsi capturer les informations stratégiques.
- Stockage de l'information : Une fois que la collecte de données est faite, Le SI assure le stockage et la gestion de ces données, en garantissant leurs intégrité , disponibilité et sécurité. Cela permet aux utilisateurs autorisés d'accéder et exploiter ces informations efficacement.
- Traitement de l'information : Le traitement de l'information permet de passer des données brutes en informations significatives. le SI est capable d'extraire des connaissances et des insights à partir des données collectées en se servant des techniques spéciales telles que l'analyse de données et l'Intelligence Artificielle (IA).
- Diffusion de l'information : Une fois que les informations ont été traitées, Le SI se charge de les diffusées aux utilisateurs pertinents e l'organisation. Cela peut se faire via des tableaux de bord, des rapports automatisés, des alertes ou d'autres moyens de communication adaptés. La diffusion de l'information permet aux décideurs et aux gestionnaires d'avoir une vision claire de la situation de la structuration.
- Aide à la prise de décision : Les SI fournissent des informations stratégiques, des analyses approfondies et des prévisions basées sur des données fiables afin comprendre les enjeux, d'évaluer les différentes options et aider les dirigeants et les responsables à prendre des décisions éclairées.
- Sécurité de l'information : Le SI met en place des mesures de sécurité pour protection des données sensibles de l'organisation contre les accès non autorisés, les pertes ou les altérations.

- Évolution et adaptation : Les SI ont flexibles et évolutifs pour intégrer de nouvelles sources de données, prendre en charge de nouvelles fonctionnalités et s'adapter aux nouvelles exigences de l'organisation. Ce qui permet à la structuration de rester compétitive et de répondre aux défis et opportunités du monde.

2.3 Les composants et les technologies des systèmes d'informations

Un SI est constitué d'une multitude d'éléments inter-connectés qui collaborent ensemble pour gérer au mieux les informations au sein d'une organisation. Un SI comprends les composants suivants [9] :

- Dimension humaine : La dimension humaine d'un SI rassemble l'ensemble des individus au sein de l'organisation. Ces individus interagissent de différentes manières avec le SI, dans divers buts soit pour utiliser des applications, assurer la maintenance technique, gérer des données ou prendre des décisions basées sur les informations disponibles.
- Les plates-formes matérielles informatiques : Les plates-formes matérielles informatiques sont les équipements physiques utilisés, comme les ordinateurs clients, les serveurs qui sont des ordinateurs très puissants qui fournissent des ressources et des services aux clients et Les équipements de réseau comprenant les routeurs, les commutateurs ainsi que d'autres dispositifs qui facilitent la communication et la connectivité au sein du réseau de l'organisation.
- Plates-formes de systèmes d'exploitation : Les plates-formes de systèmes d'exploitation ce sont les logiciels qui permettent à l'ordinateur de gérer les ressources matérielles et logicielle ainsi d'exécuter les programmes affiliés et de fournir une interface utilisateur. On peut citer Comme Système d'exploitation client Mincrosoft Windows et Linux Ubuntu. Ainsi Micorosoft Server comme sysèteme d'exploitation server
- Plates-formes de logiciels d'entreprise : Les plates-formes de logiciels d'entreprise sont des logiciels conçus pour simplifier l'exécution des tâches et gérer efficacement leurs finances et ressources humaines de l'entreprise, tel que ERP et Gestion de la relation client (CRM).
- Gestion et stockage des données : La gestion des données et du stockage se concentre sur la manière dont les informations sont gérées et stockées en utilisant les bases de données relationnelles traditionnelles ainsi que les bases de données Big Data destinées au traitement de grands volumes de données complexes et tout les logiciels pour la gestion de ces insights. On cite les bases de données structurés relationnelles et les Bases de données non-structurés orientées graphes.
- Réseaux et télécommunication : Les plates-formes de réseautage et de télécommunication concernent les équipements et technologies utilisés pour établir la communication dans une organisation. Cela inclut les réseaux locaux Ethernet ainsi les technologie sans fil.

2.4 Les applications métiers

Une application métiers est une application développée pour répondre aux besoins d'une structuration. les applications métiers sont conçues dans le but d'optimiser les processus de travail, automatiser les opérations commerciales, gérer les données et faciliter la prise de décision au sein des organisations[10].

Les applications métiers se composent de divers composants offrant des fonctionnalités métier utilisables en interne, en externe ou en combinaison avec d'autres applications métier. Elles peuvent interagir et intégrer de manière transparente avec d'autres systèmes existants au sein de l'entreprise pour échanger des données et assurer la continuité des opérations.

Ces applications sont conçues avec des mesures de sécurité robustes pour protéger les données sensibles et confidentielles de l'entreprise contre les accès non autorisés et les cyber-attaques. Elles sont évolutives et extensibles, pouvant être facilement adaptées aux besoins changeants de l'entreprise.

2.4.1 Planification des ressources de l'entreprise

Un ERP (Enterprise Resource Planning) est un système logiciel qui permet de gérer l'ensemble des processus métier d'une organisation de manière intégrée. Il regroupe différents modules et fonctionnalités qui couvrent divers domaines tels que la gestion des ressources humaines, la gestion de la production, la gestion logistique, la comptabilité, la planification, la gestion des achats, la gestion des ventes, comme illustré dans la I.2.

L'ERP utilise une base de données unique qui permet de partager et de centraliser les informations, ce qui favorise la collaboration entre les différents services de l'entreprise [11].



FIGURE I.2 – Qu'est-ce qu'un ERP? [1]

— Exemple d'ERP :

Parmi les exemples couramment utilisés dans le domaine des ERP, on retrouve [11] :

- (a) **ODOO** est l'acronyme de On-Demand Open Object, également connue sous le nom d'OpenERP, est une suite logiciels d'affaire open source de gestion d'entre-

prise.

ODOO possède des outils de veille économique intégrés permettent à plusieurs utilisateurs de gérer divers aspects du logiciel, ce qui facilite la personnalisation des rapports et des flux de travail. Pour faciliter la navigation à travers ces différentes fonctionnalités, ODOO propose une interface utilisateur intuitive où les utilisateurs peuvent accéder aux différentes fonctionnalités et modules à partir d'un tableau de bord centralisé.

- (b) **Sage ERP** est une suite logicielle complète, évolutive et personnalisable qui peut aider les entreprises de toutes tailles à améliorer leurs processus financiers et à prendre de meilleures décisions.

Il offre un large éventail de fonctionnalités pour gérer l'ensemble des processus financiers. Voici quelques-unes de ses principales fonctionnalités : notamment la gestion du plan comptable et des flux de trésorerie.

2.4.2 Gestion de la maintenance assistée par ordinateur

La Gestion de la Maintenance Assistée par Ordinateur (GMAO) ou Gestion de la Maintenance Assistée par Ordinateur est un système informatisé qui permet aux entreprises de centraliser l'information liée à la maintenance dans une base de données, améliorant ainsi la gestion des interventions des différents types de maintenance (corrective, proactive), ainsi que des fonctions d'approvisionnement des équipements et de sécurisation des travaux [6].

— Les outils du GMAO :

- **IBM Maximo** : Maximo est une solution GMAO d'IBM qui aide les entreprises à gérer leurs actifs, à planifier les travaux de maintenance et à suivre les performances.
- **eMaint** : eMaint est une plateforme de gestion de la maintenance basée sur le cloud qui offre des fonctionnalités de suivi des actifs, de gestion des ordres de travail, et de reporting.
- **CMMS (Computerized Maintenance Management System) génériques** : Certains systèmes GMAO sont génériques et peuvent être adaptés à diverses industries. Ils incluent souvent des fonctionnalités de gestion des stocks, de planification des interventions, et de suivi des coûts de maintenance.

2.4.3 Gestion de la production assistée par ordinateur

Le Gestion de la Production Assistée par Ordinateur (GPAO) ou logiciel de Gestion de la Production Assistée par Ordinateur est une solution informatique conçue dans le but de superviser et optimiser les processus de production au sein d'une entreprise. Il joue un rôle dans la simplification de la gestion de toutes les activités liées à la production, aussi dans plusieurs aspects essentiels tels que la gestion des stocks et le suivi des commandes.

— Les outils du GPAO Les outils du GPAO sont les suivant :

- **DELMIA** : DELMIA ou Digital Enterprise Lean Manufacturing Interactive Application proposé par Dassault Système, il propose des solutions de GPAO qui couvrent la planification, la simulation, l'exécution et l'analyse de la production.

- Siemens Opcenter (anciennement Camstar) : Opcenter offre des solutions de gestion de la production pour améliorer l'efficacité, la qualité et la visibilité de la production.
- Sage EMS : Sage EMS ou Solutions de Gestion d'Entreprise proposant des solutions pour les activités de production avec les systèmes ERP pour une gestion globale de la production.

3 La data science

La science des données, ou Data Science, représente un domaine en plein essor qui est au cœur de la révolution des données qui façonne notre monde moderne. À l'ère numérique actuelle, nous sommes témoins d'une explosion de données provenant d'une multitude de sources, telles que les médias sociaux, les capteurs intelligents et les transactions en ligne. La science des données offre un moyen de donner un sens à cette abondance de donnée. En transformant cette richesse informationnelle en un atout majeur pour les individus et les organisations, elle offre des opportunités exceptionnelles pour exploiter tout le potentiel qu'elles renferment.

3.1 Définition de la data science

La science des données est une approche multidisciplinaire qui s'appuie sur les statistiques, l'informatique, la communication, la gestion et la sociologie pour étudier les données et leurs environnements au sein d'une organisation [12].

science des données = statistiques + informatique + calcul + communication + sociologie + gestion | données + environnement + réflexion.

où "|" signifie "conditionnel à".

Cette approche repose sur l'extraction automatisée du sens d'un ramassis de données brutes inextricables, souvent appelés «big data», en se basant sur des méthodes scientifiques, des processus et des systèmes pour transformer ces données, en suivant une approche méthodologique allant des données à la connaissance, puis à la sagesse.

Elle ne se limite pas seulement à l'utilisation des techniques d'analyse statistiques avancées pour créer des méthodes de classification, elle explore également la façon dont les données sont traitées pour pouvoir résoudre des problèmes analytiques complexe, optimiser les processus, établir des moteurs de recommandation, développer des modèles prédictifs et prendre des décisions éclairées.

De ce fait, la data science offre un large éventail de domaines d'application et joue un rôle crucial dans de nombreux secteurs, tels que les sciences sociales, l'économie, la politique, la santé, l'environnement, l'industrie, etc.

3.2 La démarche data sciences

La science des données implique l'application d'une méthodologie rigoureuse permettant aux organisations de tirer des informations exploitables ayant un impact réel à partir de leurs données. C'est un processus systématique et itératif qui représente le cycle de vie de la data science et permet de résoudre des problèmes basés sur les données dans divers

domaines.

Cette méthode comprend plusieurs étapes clés [13] :

1. **Définition des objectifs** : Cette étape est d'une importance cruciale dans la démarche de la data science. Elle établit les fondements de la collecte de données en se référant à la compréhension des besoins métier, des problématiques à résoudre et des informations recherchées. Ce qui nécessite une collaboration étroite avec toutes les parties prenantes afin de déterminer avec précision les objectifs spécifiques de l'analyse des données.
2. **Collecte des données** : Une fois les objectifs établis, il est essentiel de rassembler les données pertinentes à partir de différentes sources tout en documentant leurs origines. Cette étape implique l'établissement d'un dictionnaire de données décrivant le format, la structure, le contenu et la signification des données en veillant à ce qu'elles soient en adéquation avec les objectifs définis pour garantir une analyse fiable.
3. **Préparation et exploration des données** : Les données brutes peuvent contenir des erreurs, des valeurs manquantes ou des incohérences. Avant de les analyser, il est nécessaire de les pré-traiter pour les rendre adaptées à l'analyse, en réduisant leur dimension si elles sont trop complexes, créant de nouvelles variables, éliminant les valeurs aberrantes, normalisant les variables catégorielles, etc . Cette étape vise à garantir la fiabilité et la pertinence des données pour une analyse précise.
4. **Application de techniques pour extraire des informations à partir des données** : Une fois les données pré-traitées, différentes techniques d'analyse peuvent être appliquées pour visualiser la distribution des données et vérifier la corrélation entre les caractéristiques. Ensuite, on peut choisir les techniques de modélisation des données les plus appropriées en fonction du type de problème (régression ou classification), puis entraîner les modèles sélectionnés à l'aide des données préparées en les divisant en ensembles d'entraînement et de validation. Après cela, les performances des modèles entraînés peuvent être évaluées à l'aide de leurs précisions pour sélectionner le modèle le plus performant capable de résoudre efficacement le problème en question .
5. **Interprétation des résultats** : Les informations extraites à partir de l'analyse des données doivent être interprétées de manière à être utiles et pertinents pour les parties prenantes, impliquant leurs transformation en résultats significatives et compréhensibles pour les décideurs.
6. **Communication des résultats** : Une fois les résultats de l'analyse de données interprétés, il est important de les présenter aux parties prenantes de manière claire et compréhensible afin de permettre aux décideurs de prendre des décisions éclairées. Cela en utilisant des rapports, des visualisations graphiques, des tableaux de bord,

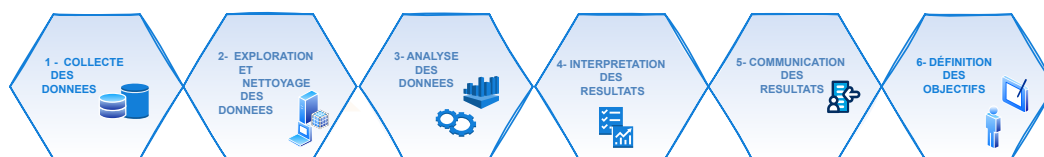


FIGURE I.3 – Les étapes d'un projet Data science

3.3 Les techniques de la data science

La data science est un domaine en constante évolution qui utilise un large éventail d'outils et de méthodes pour explorer, analyser et interpréter les données de manière efficace [14]. Parmi ses piliers fondamentaux, on peut citer l'apprentissage automatique.

ML : Le ML constitue une sous-branche de l'IA qui exploite des techniques probabilistes pour permettre à une machine d'apprendre de manière autonome à réaliser des tâches ou à faire des prédictions à partir de données. Cela consiste à développer un modèle qui repère les tendances dans les données, leur permettant de généraliser ces modèles pour des données inconnues, en se servant d'un algorithme d'optimisation pour minimiser les erreurs entre le modèle et les données.

Ces algorithmes sont catégorisés selon deux modes d'apprentissage principaux : supervisé et non supervisé [14].

3.3.1 Apprentissage supervisé

Les algorithmes supervisés utilisent des données d'entraînement contenant à la fois des entrées et des sorties correspondantes pour créer des modèles capables de prédire les sorties pour de nouvelles entrées. Il permet d'extraire des connaissances à partir d'un ensemble de données où les couples entrée-sortie sont déjà connus, en cherchant à trouver une représentation concise des associations entre les entrées et les sorties à l'aide d'une fonction de prédiction.

Quand la valeur de sortie peut prendre une infinité de valeurs dans l'ensemble continu des réels on parle de régression et quand elle prend un nombre fini K de valeurs (valeurs catégorielles) on parle de classification [14].

a. **Régression :** Les algorithmes les plus utilisés pour prédire les données incluent [14] :

- **La régression linéaire :** Cette méthode est utilisée pour modéliser la relation entre une variable dépendante cible (Y) et une variable explicative indépendante (X). Son objectif est de trouver une fonction hypothèse (une droite) qui représente au mieux cette relation, permettant ainsi de prédire la valeur Y pour de nouvelles valeurs de X tout en minimisant l'écart entre les valeurs réelles et les prédictions. On distingue : la régression linéaire univariée et la régression linéaire multivariée.
 - **Régression linéaire uni-variée**, développée au début du XXe siècle, est l'une des plus anciennes et les plus simples techniques de l'apprentissage machine. Elle est utilisée lorsque l'on souhaite prédire des valeurs (Y) en fonction d'une seule variable explicative (X).
Pour ajuster la fonction hypothèse et trouver la meilleure droite possible, on utilise une fonction de coût pour mesurer l'erreur entre les valeurs prédites et les valeurs réelles de Y , et on cherche à minimiser cette fonction de coût en ajustant les paramètres de la fonction hypothèse grâce à la descente de gradient, permettant ainsi d'obtenir une meilleure approximation de la relation entre X et Y .
 - **Régression linéaire mult-ivariée**, développée dans les années 1930, est une extension de la régression linéaire uni-variée qui utilise plusieurs variables explicatives (X) pour prédire une variable cible (Y).

- **Régression polynomiale** : Cette méthode, développée dans les années 1960, est utilisée pour capturer des relations non linéaires entre les variables d'entrée (X) et de sortie (Y), en utilisant un polynôme de degré k.
 - **Régression régularisée** : Cette méthode, développée dans les années 1990, est utilisée pour éviter le surapprentissage (overfitting) et améliorer la performance des modèles de régression linéaire. Elle vise à introduire des contraintes en ajoutant une fonction de pénalité à la fonction de coût classique du modèle afin de restreindre l'amplitude des paramètres en les rapprochant de zéro, ce qui permet de rétrécir l'espace des solutions autour de zéro, stabilisant ainsi les résultats de l'estimation.
- b. **Classification** : Il existe d'autres approches plus spécifiques de classification de données [14]. Les voici :
- **Le classifieur naïf bayésien (Naive Bayes)** : Cette méthode constitue l'un des classifieurs les plus simples utilisés pour la classification binaire et multiclasse. Le classifieur naïf bayésien, reposant sur la loi de Bayes et la théorie des probabilités, se base sur l'hypothèse de l'indépendance conditionnelle entre les caractéristiques (ou les variables explicatives) lorsqu'on connaît la classe cible (la variable dépendante).
 - **Régression logistique** : Cette méthode, développée dans les années 1950, représente un classifieur binaire performant, très répandu chez les banquiers et les assurances, offrant un pouvoir explicatif très fort. Ce modèle statistique utilise une fonction hypothèse appelée sigmoïde pour prédire la probabilité qu'une observation appartienne à l'une des deux classes possibles. En utilisant une fonction de coût basée sur la perte logarithmique (log loss), elle pénalise les erreurs de classification, améliorant ainsi la précision des prédictions, en ajustant les paramètres du modèle avec la descente de gradient.
- c. **Classification & Regression** : Divers algorithmes d'apprentissage sont polyvalents et peuvent être adaptés à la fois pour la régression et la classification. Voici quelques exemples [14] :
- **K-plus proches voisins (K-NN)** : Cet algorithme est un modèle utilisé pour la classification ou la régression. Il est basé sur la proximité des données et cherche les k plus proches voisins pour prédire la classe ou la valeur de sortie.
 - **Arbre de décision (DT)** : Ces modèles créent des arbres qui divisent les données (la série de variables continues ou qualitatives) en sous-ensembles similaires en se basant sur leurs caractéristiques afin de prédire les étiquettes de sortie pour de nouvelles instances.
 - **Forêt aléatoire (RF)** : Ce modèle construit plusieurs arbres de décision de manière indépendante à partir d'un sous-ensemble aléatoire des données d'apprentissage.

(tree bagging) et en utilisant un sous-ensemble aléatoire des caractéristiques (feature sampling).

En classification, la construction de chaque arbre se base sur deux critères : Gini, qui se concentre sur la séparation de la classe la plus représentée et entropie, qui vise à maximiser le gain d'information à chaque étape de construction de l'arbre. Les arbres créés sont ensuite combinés pour faire des prédictions finales sur de nouvelles données, en faisant la moyenne des prédictions de chaque arbre en classification et par un vote en régression, où chaque arbre "vote" pour une classe, et la classe avec le plus de votes est choisie comme prédiction finale .

- **Gradient Boosting (GB)** : Le renforcement par gradient consiste à combiner de manière itérative plusieurs arbres de décision de faible profondeur. Chaque itération vise à corriger l'erreur de la précédente en se référant au gradient de la fonction coût choisie qui indique la direction dans laquelle l'erreur du modèle peut être réduite le plus efficacement.
- **Machines à vecteurs de supports (SVM)** : Ce modèle se distingue par sa capacité à trouver une frontière de décision optimale (hyperplan) qui maximise la marge entre les classes, permettant ainsi de minimiser les erreurs de prédiction. Lorsque notre jeux de données n'est pas linéairement séparables, le SVM utilise une transformation mathématique, à l'aide d'une fonction appelée kernel trick, pour les projeter dans un nouvel espace de dimension supérieure où elles peuvent être séparées linéairement.

3.3.2 Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, les algorithmes non supervisés ne se base pas sur des données d'entraînement avec des sorties définies. Ils considèrent toutes les données comme équivalentes, se concentrant uniquement sur les motifs et les structures observables. Leur objectif principal est de partitionner un ensemble de données en un nombre défini de groupes (appelés "clusters"), en fonction de leur similitude [14].

- **K-Means** : Ce modèle est utilisé pour regrouper des données similaires en un nombre "K" de groupes prédéfini, appelé clusters. Il assigne chaque point de données au cluster dont le centre est plus proche . Ensuite, il recalcule les centroides de sorte à minimiser la somme des distances euclidiennes entre les points de chaque cluster et leur centre, et répète ce processus jusqu'à convergence [15].
- **Clustering hiérarchique** : Cet algorithme regroupe des clusters dans une structure hiérarchique en fonction de leur similarité. Il commence par considérer chaque point de données comme un cluster distinct, puis fusionne les clusters les plus proches de manière itérative jusqu'à ce qu'il ne reste plus qu'un seul cluster. L'arborescence obtenue permet de visualiser la hiérarchie des clusters et de choisir le niveau de granularité souhaité [15].
- **Regroupement spatial basé sur la densité avec bruit (DBSCAN)** : Ce modèle identifie les clusters en fonction de la densité des points de données au sein d'un voisinage. Il regroupe les points qui sont proches les uns des autres, formant des régions denses tout

en marquant les points dans les régions moins denses comme des valeurs aberrantes ou du bruit. Cela le rend capable de gérer des jeux de données avec des formes et des tailles de cluster variables [15].

3.3.3 Apprentissage Profond (DL)

L'apprentissage profond, une sous-branche de l'apprentissage machine ML, se focalise sur l'automatisation de l'apprentissage grâce à des réseaux de neurones artificiels, inspirés de la structure du cerveau humain. Ces réseaux sont capables d'apprendre de façon autonome et de détecter des schémas complexes dans les données. Ils excellent notamment dans la résolution de problèmes de reconnaissance d'images, de traitement du langage naturel, de jeux vidéo, et d'autres domaines où les données sont diverses, structurées ou non [16].

Parmi les différentes architectures de réseaux de neurones utilisées en DL [16], on trouve :

- **Réseaux de neurones artificiels (ANN) :** Ce modèle est construit à partir d'un grand nombre de neurones (perceptrons) connectés entre eux. Le Perceptron prend la somme pondérée de ses entrées, leur applique une fonction mathématique, et produit son résultat comme illustré dans la figure I.4.

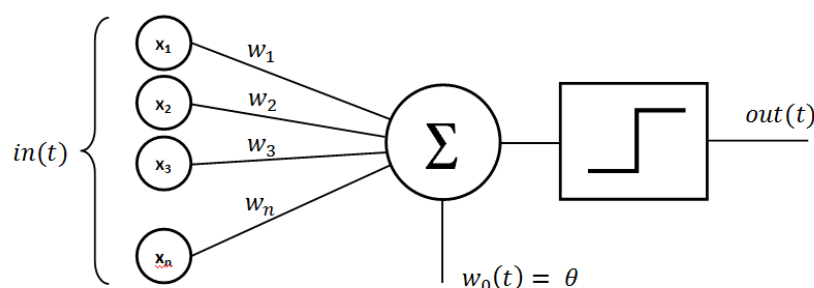


FIGURE I.4 – Structure d'un perceptron

En empilant des perceptrons les uns sur les autres, on forme des couches, et ensuite ces couches peuvent être placées séquentiellement les unes à côté des autres pour rendre le réseau plus profond tel démontré dans la figure I.5.

La couche d'entrée (Input layer) reçoit les données brutes (initiales) et les transmet aux neurones de la couche cachée (Hidden layer). Ces neurones sont connectés par des poids qui ajustent l'importance des signaux et transforment les entrées en représentations plus complexes, dont se sert la couche de sortie (Output layer) pour générer les prédictions ou classifications finales.

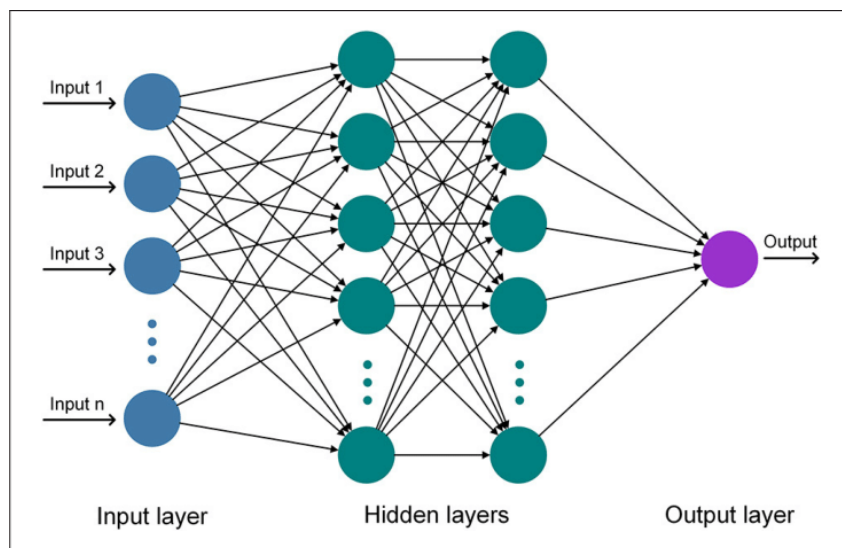


FIGURE I.5 – Structure d'un ANN [2]

- **Réseaux de neurones convolutifs (CNN)** : Ces réseaux sont particulièrement efficaces pour la reconnaissance d'images et la classification d'objets dans des images. Ils se composent de plusieurs couches. Tout d'abord, une couche de convolution qui applique des filtres sur l'image d'entrée et effectue une opération de convolution pour créer une carte d'activation. Ces filtres apprennent à détecter des caractéristiques locales de l'image tels que des bords, des textures, etc. Ensuite, pour augmenter la robustesse du modèle, la couche de pooling réduit la dimension spatiale de la carte d'activation (c'est-à-dire le nombre de paramètres) en prenant la valeur maximale (max-pooling) ou moyenne (average-pooling) dans chaque région. Enfin, la couches entièrement connectées effectue la classification ou la régression finale.
- **Les réseaux de neurones récurrents (RNN)** : Ce modèle est conçu pour le traitement de données séquentielles, telles que des séries temporelles et des données textuelles (séquence de mots) ou audio (séquence de fréquences sonores). Le RNN analyse les séquences élément par élément, en utilisant une mémoire interne pour capturer les dépendances séquentielles. À chaque pas de temps, l'entrée du neurone est combinée avec l'état caché (mémoire) précédent pour produire une nouvelle sortie, comme démontré dans la figure I.6. Cela leur permet de se souvenir des informations passées et de les utiliser pour prendre des décisions en temps réel.

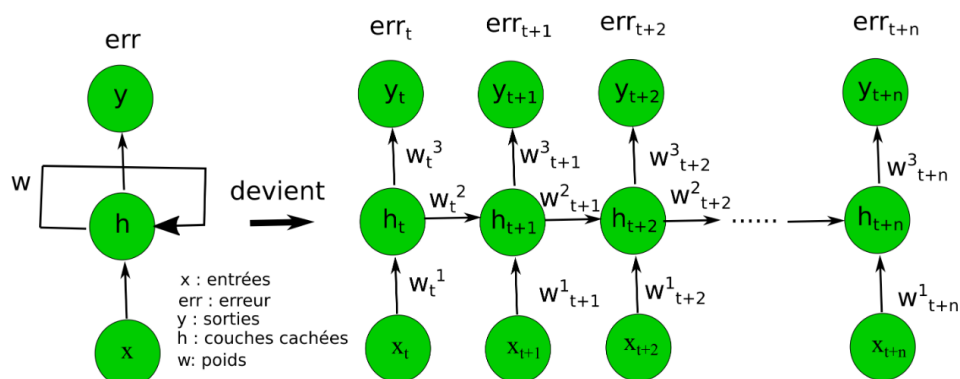


FIGURE I.6 – Structure d'un RNN [3]

- **Mémoire à Long Terme et Court Terme (LSTM)** : Les (LSTM) sont une évolution des RNN, spécialement conçue pour gérer efficacement les dépendances à long terme dans les séquences de données. Ils utilisent des cellules mémoire avancées équipées de portes d'oubli, d'entrée et de sortie pour capturer et maintenir des informations pertinentes sur de longues périodes. Cette capacité améliore la précision des modèles dans des applications telles que la prédiction temporelle et la traduction automatique, tout en atténuant les problèmes d'explosion du gradient lors de l'entraînement sur des séquences prolongées.[16].

Un réseau LSTM est composé de plusieurs couches :

- Couche d'entrée** : La couche d'entrée reçoit les données brutes, qui sont ensuite propagées à travers une ou plusieurs couches LSTM cachées.
- Couches LSTM cachées** : Chaque couche LSTM est constituée de multiples cellules ou blocs de mémoire, chacun ayant une structure particulière avec trois portes différentes qui leur permet de mieux réguler la façon d'apprendre ou d'oublier à partir des différentes sources d'entrée, comme illustré dans la figure I.7 :

La porte d'entrée, prend en compte l'entrée actuelle ainsi que l'état précédent de la cellule mémoire et génère un vecteur d'activation qui représente les informations qui doivent être mises à jour dans la cellule mémoire. Avec la porte d'oubli, la cellule peut supprimer les informations obsolètes. Elle laisse de côté certaines informations pertinentes, puis en ajoute de nouvelles par opération d'addition. Enfin, la porte de sortie filtre les informations pertinentes de la cellule via une fonction d'activation afin de produire la sortie du LSTM pour cette étape de la séquence.

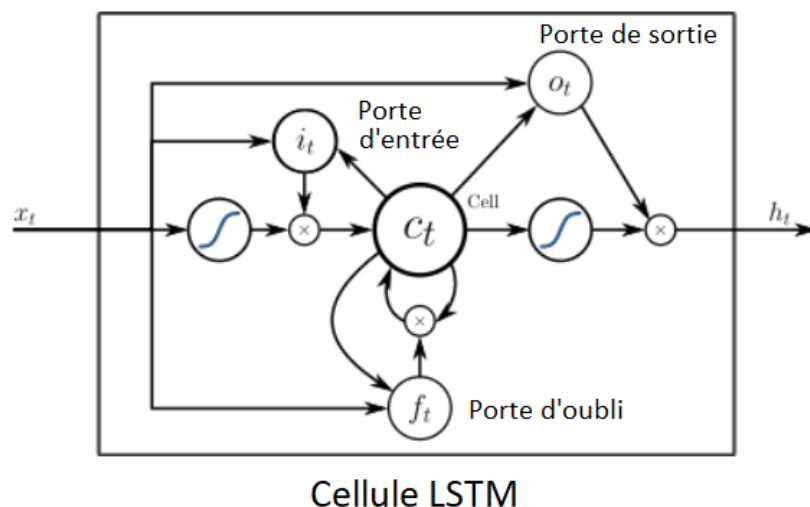


FIGURE I.7 – Cellule d'un LSTM

- c) **Couche entièrement connectées (dense)** : Les sorties des cellules LSTM sont ensuite combinées et traitées par une ou plusieurs couches denses entièrement connectées avant d'atteindre la couche de sortie finale.
- d) **Couche de sortie** : La couche de sortie s'agit de la dernière couche du réseau de neurones. Elle utilise une fonction d'activation adaptée pour produire les résultats finaux (prédiction désirée). Pour la classification, souvent on utilise la fonction softmax qui permettra d'obtenir une distribution de probabilités sur les classes. Tandis que pour la régression, des fonctions comme la fonction linéaire, la sigmoïde ou ReLU sont couramment employées.
- **Réseaux de neurones profonds (DNN)** : Aussi connus sous le nom de réseaux de neurones multicouches, ces réseaux sont composés de plusieurs couches cachées entre la couche d'entrée et la couche de sortie. Ils sont utilisés pour des tâches de classification et de prédiction sur des données complexes, structurées et non structurées.
- **Réseaux de neurones génératifs adversariaux (GAN)** : Cette architecture de réseau de neurones est utilisée pour générer de nouvelles données réalistes en apprenant à partir d'un ensemble de données existant.
GAN est composé de deux réseaux de neurones : le générateur, qui génère des données aléatoires et le discriminateur qui compare ces données avec des exemples réels puis fournit un retour au générateur sur la qualité de ses sorties permettant au générateur d'ajuster ses paramètres afin d'améliorer la qualité de sa génération de données.
- **Réseaux de neurones auto-organisés (SOM)** : Les SOM sont utilisés pour la visualisation de données et la compression de données multidimensionnelles. Ils sont souvent utilisés dans des applications où la compréhension de la structure sous-jacente des données est importante.

3.3.4 Analyse de séries temporelles

Une série temporelle représente une suite de valeurs numériques qui illustrent l'évolution d'une quantité dans le temps. L'analyse de série temporelle constitue une approche de la Data Science qui consiste à étudier et à modéliser les données temporelles afin d'en extraire des informations utiles. Les principales étapes de cette analyse sont la préparation des données, l'exploration des données, la décomposition de la série temporelle en ses composantes (tendance, saisonnalité, résidu), la modélisation et l'évaluation du modèle. Les modèles de série temporelle les plus couramment utilisés sont [17] :

- **Modèle Autorégressif Intégré à Moyenne Mobile (ARIMA) :** ARIMA est un modèle statistique utilisé pour analyser et prévoir des séries temporelles. Il combine l'auto-régression, la moyenne mobile et la différenciation des données pour modéliser les tendances et les motifs saisonniers et les comportements à court terme des données temporelles. Cependant, il suppose que la série temporelle est stationnaire et peut donc être limité dans la modélisation de tendances non linéaires ou de structures complexes.
- **Lissage exponentiel :** Le lissage exponentiel est une méthode de prévision qui attribue des poids décroissants aux observations passées. Elle est particulièrement utile pour modéliser les tendances et les effets saisonniers dans les données temporelles.

3.3.5 Traitement du langage naturel (NLP)

Le traitement du langage naturel constitue une discipline de la data science visant à analyser, comprendre et extraire des informations de manière intelligente et efficace à partir de données textuelles. En appliquant le Traitement du langage naturel (NLP), les données textuelles peuvent être représentées sous forme d'équations mathématiques ou de modèles afin de pouvoir saisir le contenu et les traiter. Il regroupe diverses techniques comme la tokenisation, la lemmatisation, la classification de texte, la traduction automatique, l'analyse de sentiments, ainsi que divers algorithmes tels que [17] :

- **Les transformers :** Les transformers sont une architecture de réseaux de neurones révolutionnaire introduite en 2017 dans le domaine du NLP. Ils utilisent une approche basée sur l'attention pour traiter les données séquentielles. Cette approche permet aux transformers de prendre en compte les relations entre toutes les parties d'une séquence, ce qui leur permet de capturer les dépendances à longue distance et d'effectuer des calculs parallèles. Les transformers sont principalement connus pour leur utilisation dans des modèles de langage tels que BERT et GPT, qui ont atteint des performances remarquables dans de nombreuses tâches de NLP, notamment la classification de texte, la traduction automatique et la génération de texte.
- **Les modèles de langage basés sur les probabilités :** Les modèles de langage basés sur les probabilités, tels que les modèles de Markov, sont une approche classique du NLP qui utilise des concepts de probabilités statistiques pour prédire des mots ou des séquences de mots. Ils ont été largement utilisés avant l'avènement des RNN et des transformers, qui sont capables de capturer des dépendances plus complexes dans les données séquentielles.

3.4 La mise en place d'une stratégie de gestion appuyée sur la donnée (data-driven-decision)

Au cours de la dernière décennie, la quantité de données produites a connu une augmentation significative, passant d'un peu plus de 1 000 milliards de gigaoctets à 59 000 milliards de gigaoctets (Press, 2020). Il est donc devenu essentiel pour les entreprises de se référer à ces données afin d'améliorer leurs performances, car la variation des données entraîne l'augmentation de la productivité de 4 à 6% [18].

3.4.1 Définition de la stratégie de gestion appuyée sur la donnée

La data-driven decision, ou décision basée sur les données, est un concept clé de la data science. Il fait référence à l'utilisation des données et de leurs corrélations pour identifier les meilleures stratégies et actions à prendre dans le processus décisionnel des entreprises. Cette approche est utilisée dans plusieurs domaines et secteurs d'activité (le marketing, la finance, la gestion des ressources humaines, la logistique, ...). Elle permet aux décideurs d'obtenir des informations objectives et véridiques qui aident à éclairer différentes décisions de l'entreprise en améliorant la précision et en réduisant les erreurs de jugement subjectives. En d'autres termes, Plutôt que de se fier uniquement à l'intuition, à l'expérience ou à des suppositions, la data-driven decision s'appuie sur les informations et les connaissances extraites des données pour prendre des décisions plus sage et mieux renseignées [19].

3.4.2 Les étapes du processus de la mise une place de cette stratégie

Pour mettre en œuvre une stratégie de data driven decision, il est essentiel de suivre un processus méthodique comme suit [19] :

1. Définition des objectifs de l'entreprise : Tout d'abord, il est important de définir les objectifs à atteindre dans le cadre de la prise de décision de sorte à ce qu'ils soient spécifiques, mesurables, réalisables, pertinents et limité dans le temps.
2. Collecte et intégration des données : Afin d'évaluer les objectifs de l'entreprise, il est nécessaire de collecter des données pertinentes. Cela peut inclure des données internes provenant de divers systèmes tels que les ventes, le marketing, les ressources humaines et les opération de maintenance industrielle, ainsi que des données externes provenant de sources telles que les médias sociaux, les données démographiques ou les fournisseurs de données tierces. Ensuite ces données doivent être nettoyer et intégrer dans un entrepôt de données centralisé.
3. Analyse des données : Les données collectés sont soumises à des techniques d'analyse afin d'en extraire des informations significatives et des modèles exploitables. Cela en se basant sur des techniques de ML et DL pour analyser les données et identifier des modèles prédictifs ou des segments de clientèle. Ce processus inclut l'exploration de données (data mining) pour découvrir des tendances et des relations intéressantes entre ces données. L'utilisation des outils statistiques et des logiciels d'analyse pour effectuer ces analyses et obtenir des informations exploitables.
4. Visualisation des données : Les résultats de l'analyse des données doivent être présenter de manière simples et intuitives pour en comprendre la signification et leur implication vis-à-vis des objectifs initiaux.
5. Prise de décision basée sur les données : A ce niveau se fait l'utilisation des informations issue des analyses générées pour prendre des décisions éclairées, en orientant les

choix et les actions à entreprendre pour atteindre les objectifs fixés. Cela peut inclure l'optimisation des processus internes, l'identification de nouvelles opportunités commerciales, la gestion de la maintenance industrielle ou l'amélioration de l'expérience utilisateur.

6. Suivi et évaluation : Cette étape repose sur la mise en place d'un mécanisme de suivi pour évaluer les résultats des décisions basées sur les données. Il est important de définir des indicateurs de performance clés appropriés en fonction des objectifs fixés afin d'analyser régulièrement les résultats et effectuer des ajustements si nécessaire pour améliorer continuellement la stratégie et les performances de l'entreprise.

3.4.3 Les avantages de la stratégie data-driven :

La data driven decision présente plusieurs avantages, on distingue parmi eux [19] :

- Prise de décision éclairée : En utilisant les données comme base pour la prise de décision, les organisations peuvent prendre des décisions plus pertinentes construites sur des preuves concrètes plutôt que sur des intuitions.
- Amélioration de la performance : la stratégie data-driven permet d'identifier les domaines où des améliorations sont requises pour optimiser les processus et réduire les coûts, permettant ainsi d'augmenter la performance globale de l'entreprise.
- Personnalisation et expérience client : Grâce à l'utilisation des données pour une meilleure compréhension des clients, de leurs préférences, de leur comportement et de leurs besoins, les entreprises peuvent personnaliser leurs produits et services et leurs interactions avec leurs clients.
- Détection des tendances et des opportunités : L'analyse des données permet d'identifier les tendances émergentes, les schémas cachés et les opportunités commerciales. Cela facilite aux entreprises d'anticiper les changements du marché et de prendre des mesures proactives pour saisir de nouvelles opportunités.
- Avantage concurrentiel : En adoptant une approche data-driven, les entreprises peuvent identifier les inefficacités, perfectionner les processus et prendre des décisions instruites plus rapidement que leurs concurrents.

4 Cas d'étude : Groupe agroalimentaire Cevital

Dans le cadre de notre projet de recherche final portant sur l'optimisation de la gestion proactive de la maintenance industrielle à travers l'application d'une méthodologie d'analyse de données, nous avons obtenu une opportunité de stage au sein du prestigieux groupe agroalimentaire Cevital. Cette collaboration nous avait permis d'approfondir notre compréhension des défis spécifiques rencontrés dans le domaine de la maintenance industrielle au sein d'un environnement agroalimentaire complexe et dynamique.

Nous avons été enthousiasmés par l'occasion qui nous avait été offerte de travailler au sein d'une entreprise de renommée internationale comme Cevital, réputée pour son engagement envers l'innovation et l'excellence opérationnelle. Grâce à cette expérience pratique, nous avons pu mettre en œuvre et valider notre modèle de gestion de la maintenance proactive, tout en bénéficiant de l'expertise et des ressources disponibles au sein de l'organisation.

4.1 Historiques

Le Groupe Cevital, fondé en 1998 par l'entrepreneur Issad Rebrab, est un conglomérat algérien opérant dans divers secteurs tels que l'industrie agroalimentaire, la grande distribution, l'industrie et les services. Il est reconnu comme le premier groupe privé algérien et occupe une position de leader en Algérie et à l'international [20].

Cevital a débuté son parcours à Béjaïa en se concentrant sur l'industrie agroalimentaire. Par la suite, le groupe a élargi ses activités en créant Mediterranean Float Glass, spécialisée dans la production et la distribution de verre pour diverses applications.

Au fil des années, Cevital s'est diversifié et consolidé en étendu sa présence à l'international, tel montré dans la figure I.8.

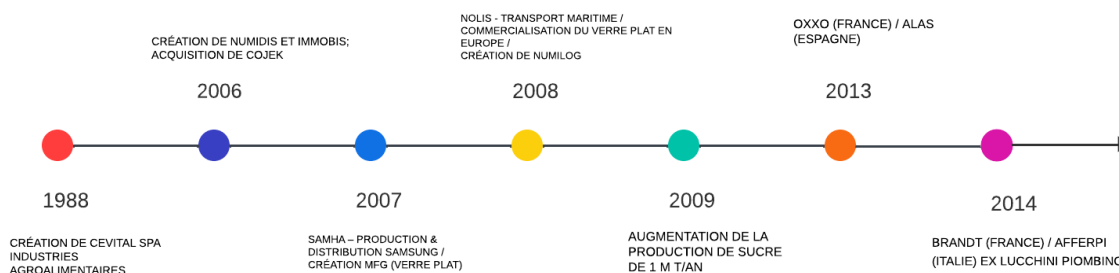


FIGURE I.8 – Etapes historiques du groupe Cevital

Cevital met l'accent sur l'innovation, le développement durable et la responsabilité sociale d'entreprise, tout en soutenant des initiatives philanthropiques.

4.2 Emplacement géographique

Cevital occupe une position de leader dans le secteur agroalimentaire en Algérie. Son complexe de production se trouve dans l'enceinte portuaire de Béjaïa, à une proximité immédiate des routes nationales RN26 et RN9. Avec une superficie de 45000 m², il s'agit du plus vaste complexe privé en Algérie.

De plus, sa situation géographique présente un avantage considérable en raison de son voisinage avec tous les facteurs socio-économiques. En effet, la gare routière est à deux minutes à pied, tandis que l'aéroport international est à seulement dix minutes en voiture.

4.3 Les filiations de Cevital

Les différentes branches familiales de Cevita sont les suivantes [20] :

— **Automotive, Immobilier et Services**

Le groupe Automotive, Immobilier et Services, créé en 2012, opère dans l'automobile, l'immobilier, la location de voitures et l'affichage publicitaire. Parmi ses filiales automobiles, on trouve HMA (Hyundai Motor Algérie), SODI Automotive (distributeur des marques Fiat, Alfa Romeo, Lancia et Fiat Professional en Algérie) et COGETP (distributeur d'engins de travaux publics) ainsi que Immobis (promotion immobilière), MTP (logistique et location d'engins) et Futur Media (médias d'affichage grand format) qui assure ses services.

— **Industrie**

Le pôle Industrie du Groupe Cevital, créé en 2014, supervise 12 filiales. Ces filiales opèrent dans les secteurs de l'électroménager (Groupe Brandt), des fenêtres et fermetures (Oxxo), du verre et de ses dérivés (MFG), de la construction (Baticompos, Cevital Entreprise, PCA, Métal Structure), ainsi que de la logistique et des services (NUMILOG, Atlas Pneu, Nolis).

— **Agro-industrie et distribution**

Cevital Agro-industrie, mis en place en 2015, est doté de plusieurs unités de production ultramodernes, comprenant une raffinerie d'huile, une raffinerie de sucre, une margarinerie, une unité de conditionnement d'eau minérale, une unité de fabrication et de conditionnement de boissons rafraîchissantes sans alcool, une unité de fabrication de chaux calcinée, une conserverie, ainsi que des silos portuaires. Par ailleurs, l'entreprise dispose d'un terminal de déchargement portuaire ayant une capacité de 2000 tonnes par heure, ce qui en fait le premier terminal de déchargement portuaire en Méditerranée.

Au sein de la commune de Béjaïa, l'entreprise Cevital déploie une activité diversifiée, mais néanmoins ancrée dans le domaine de l'industrie agroalimentaire. Cette activité englobe plusieurs volets :

- La production de margarine (600 tonnes/jour).
- Le raffinage du sucre (3000 tonnes/jour),
- Le raffinage des huiles alimentaires (1800 tonnes/jour).
- Conditionnement d'huile (1400 tonnes/jour).
- Stockage des céréales (120000 tonnes).

4.4 L'organigramme de Cevital

La figure I.9 représente un organigramme générale et hiérarchique des services du groupe agroalimentaire Cevital.

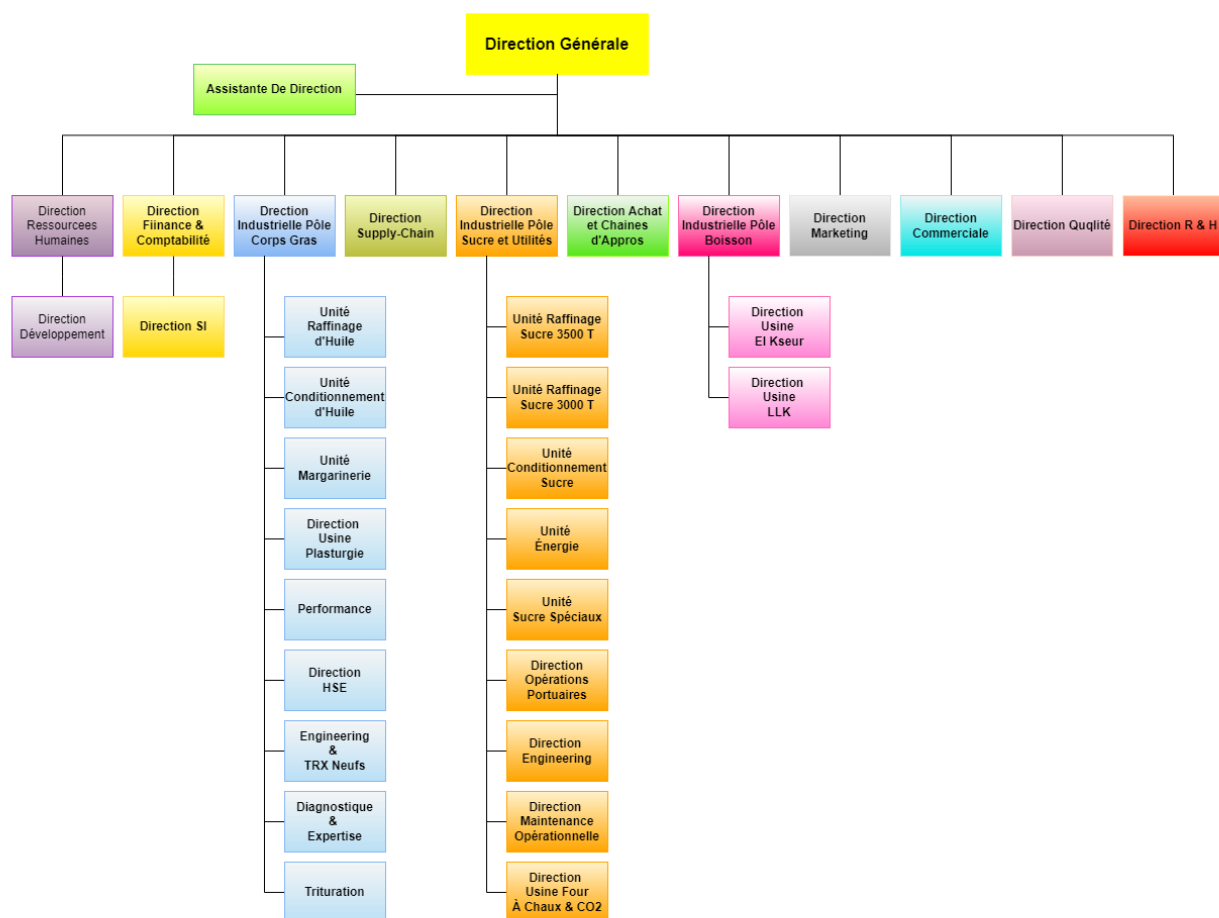


FIGURE I.9 – Organigramme de Cevital

4.5 Gestion de la Maintenance et Techniques Associées dans le Groupe Cevital

Le service de maintenance de Cevital joue un rôle crucial pour assurer le bon fonctionnement et la pérennité des équipements industriels au sein de l'organisation. Il est chargé de veiller à ce que les machines et les installations fonctionnent de manière optimale, minimisant ainsi les interruptions et les pertes de production.

Pour ce faire, le service utilise le logiciel de GMAO Coswin 8.2. Cette plateforme informatique permet de :

- Gérer les interventions correctives pour résoudre rapidement les incidents, perturbant la ligne de production, et limiter les temps d'arrêt imprévus.
- Planifier et suivre les interventions de maintenance préventive selon un calendrier basé sur les recommandations des fabricants, visant à prévenir les pannes en anticipant les problèmes potentiels avant qu'ils ne se manifestent.
- Planifier efficacement les tâches de maintenance, affecter les ressources et ordonner les travaux, garantissant ainsi une gestion optimale du temps et des compétences des techniciens.

- Collecter des données sur les performances des équipements et des activités de maintenance. Ces données permettent de générer des rapports détaillés et des analyses afin d'évaluer l'efficacité opérationnelle et d'identifier les domaines d'amélioration.
- Assurer la sécurité des données et la conformité aux normes réglementaires en vigueur. Cela garantit la confidentialité et l'intégrité des informations liées aux activités de maintenance.

5 Conclusion

En conclusion, dans ce chapitre, nous avons abordé l'importance des systèmes d'information en tant que piliers de la croissance des entreprises et leur digitalisation pour l'optimisation des opérations. Nous avons exploré la science des données, en détaillant ses techniques, y compris le ML et DL. Ainsi que la mise en place d'une stratégie de décision basée sur les données (data-driven decision). Enfin, nous avons discuté de l'organisme d'accueil qui nous a accompagnés tout au long de ce projet, en mettant en lumière son rôle dans l'optimisation des opérations industrielles. Cette exploration a mis en évidence comment la synergie entre le numérique et l'ingénierie proactive, soutenue par la data science, peut transformer les pratiques industrielles pour les rendre plus efficaces et résilientes. Dans le chapitre suivant, nous allons effectuer un état de l'art sur la maintenance industrielle et ses techniques en approfondissant sur la maintenance prédictive basée sur l'analyse de données.

Chapitre II

État de l'art des méthodes de maintenance proactive industrielle assistée par ordinateur

1 Introduction

La maintenance industrielle est essentielle pour garantir la fiabilité et la disponibilité des systèmes de production. Elle comprend une variété d'activités visant à maintenir ou à restaurer le bon fonctionnement des équipements.

Dans ce chapitre, nous commencerons par définir la maintenance industrielle, en explorant ses différents types tels que la maintenance corrective, préventive et prédictive. Ensuite, nous nous concentrerons sur un état de l'art des recherches dans le domaine de la maintenance prédictive, en mettant particulièrement l'accent sur l'application des techniques d'apprentissage automatique et profond. Ces technologies avancées permettent de prévoir les défaillances et d'optimiser les interventions de maintenance, améliorant ainsi l'efficacité et réduisant les coûts. Enfin, nous présenterons une analyse comparative détaillée des travaux synthétisés, mettant en lumière les avancées, les défis et les perspectives futures dans le domaine de la maintenance prédictive.

2 Taxonomie des méthodes de maintenance

La maintenance industrielle est un concept défini par la norme EN 13306 comme une combinaison d'actions techniques, administratives et de gestion effectuées au cours du cycle de vie d'un élément, dans le but de rétablir ses conditions de fonctionnement requises [21].

Dans le secteur industriel, la maintenance revêt une grande importance en raison de son impact considérable sur les coûts, la fiabilité et la compétitivité d'une entreprise en termes de prix, de qualité et de performance. En effet, toute interruption non planifiée d'un équipement peut entraîner l'arrêt de l'activité principale de l'entreprise, engendrant ainsi des pénalités financières importantes et une perte de réputation non-mesurable.

Selon l'étude réalisée par Park, Chulsoon, et al [22], plus de 30% des coûts de maintenance sont attribuables à une mauvaise planification. Il est donc impératif pour les entreprises de développer et de mettre en œuvre une stratégie de maintenance efficace.

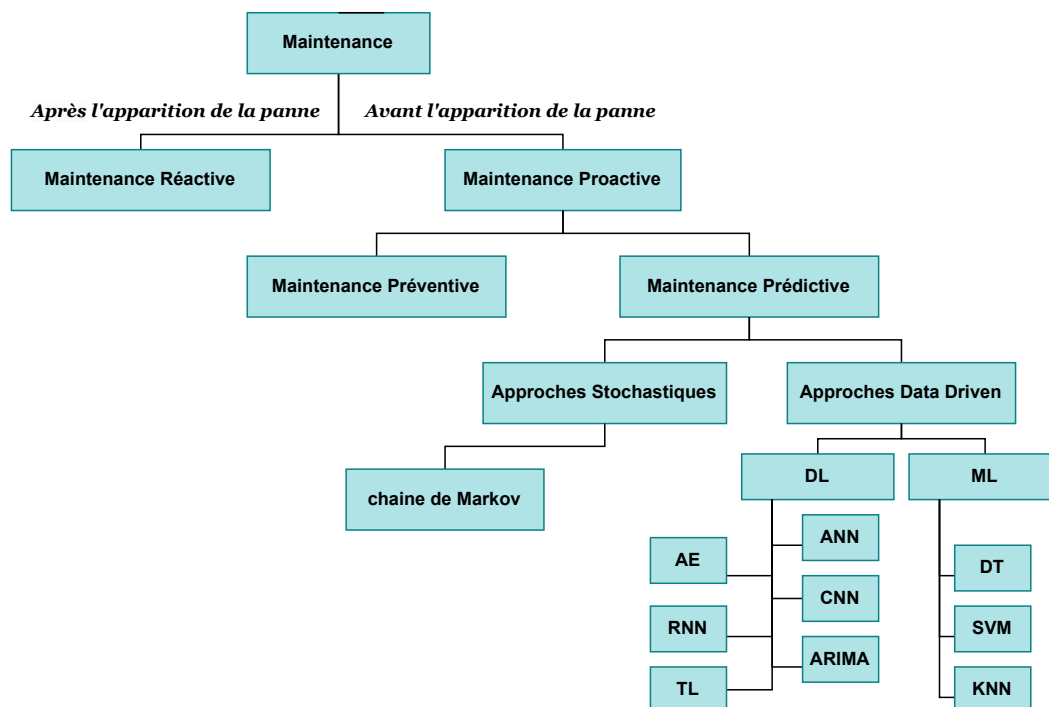


FIGURE II.1 – Taxonomie des méthodes de maintenance

Dans cette optique, la figure II.1 illustre une classification des méthodes de maintenance industrielle, élaborée à partir des recherches menées par Samberkar et al [7], Yongy et al [4] et Gackowiec et al [23].

La classification de Samberkar et al [7], distingue deux grandes catégories de méthodes de maintenance : la maintenance réactive et la maintenance proactive. Dans cette classification, la maintenance corrective intervient après l'occurrence de la panne. En revanche, la maintenance proactive est réalisée avant l'arrivée de la panne. La distinction entre maintenance réactive et proactive met en lumière l'importance de la prévoyance et de la planification dans la gestion des actifs industriels. En adoptant une approche proactive axée sur la prévention des pannes plutôt que sur leur réparation après coup, les organisations peuvent réaliser des économies significatives et améliorer la fiabilité de leurs équipements.

2.1 La maintenance réactive (RM)

La Maintenance Réactive (RM), également connue sous le nom de maintenance corrective, est effectuée uniquement en cas de dommage. Elle vise à réparer ou à rétablir l'état opérationnel après l'apparition d'une panne [4].

Bien que la RM offre une utilisation maximale de l'équipement et donc une production maximale, elle est rarement utilisée en pratique, car elle peut provoquer des temps d'arrêt non planifiés entraînant des coûts élevés de réparation ou de remplacement des composants défectueux. De plus, le recours à la RM peut causer des dommages supplémentaires à l'équipement, nécessitant des réparations coûteuses supplémentaires.

Pour gérer la RM, l'entreprise doit maintenir d'importants stocks de pièces de rechange ou

compter sur les fournisseurs d'équipement pour des livraisons immédiates. Par conséquent, elle s'est avérée inefficace.

2.2 La maintenance proactive (PRM)

La Maintenance Proactive (PRM) vise à anticiper les pannes en identifiant leurs causes. Cette approche travaille à programmer et planifier les activités de maintenance en suivant un plan d'action axé sur la prévention ou la prédiction des défaillances.

De la figure II.1 des taxonomie des méthodes de la maintenance industriel, Les travaux de Gackowiec [23] et Yongy et al[4], ont décidé d'aller plus loin dans la classification de la PRM en la découpant en deux principales méthodes : la maintenance préventive planifiée pour éviter les pannes et la maintenance prédictive pour anticiper les défaillances. Ces approches complémentaires visent à maximiser la disponibilité des équipements et à réduire les temps d'arrêt.

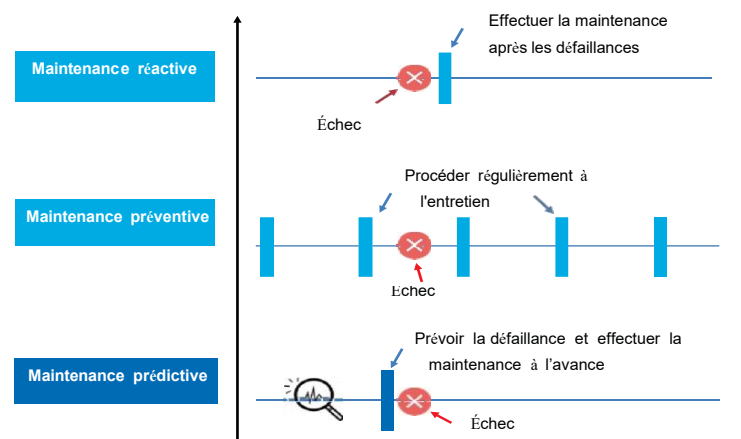


FIGURE II.2 – Les types de maintenance industriels [4]

2.2.1 La maintenance préventive (PM)

La PM également appelée maintenance planifiée, est l'une des stratégies de maintenance les plus utilisées dans le domaine de la maintenance. Elle consiste à planifier des activités de maintenance régulières sur des équipements spécifiques afin de réduire la probabilité de défaillance comme démontré dans la figure II.2 qui détaille le fonctionnement des diverses méthodes de maintenance industriel. La PM est effectuée même lorsque la machine fonctionne encore et est en opération normale, afin d'éviter les pannes imprévues et les coûts associés[4].

Selon Yongy et al [4], la plupart des programmes de gestion de la PM sont basés sur le temps, ce qui signifie que les activités de maintenance sont planifiées en fonction du temps écoulé. Ils supposent que le comportement de défaillance de l'équipement est prévisible, ce qui est représenté par des courbes en forme de baignoire illustré dans la figure

II.3. Cette courbe indique qu'un nouvel équipement présente une probabilité élevée de défaillance dues à des problèmes d'installation au cours des premières semaines d'exploitation. Après cette période de rodage, le taux de défaillance devient relativement faible pendant une période prolongée. Ensuite, la probabilité de défaillance augmente de manière significative avec le temps écoulé. Ils considèrent que la PM peut réduire les coûts de réparation et les temps d'arrêt imprévus, mais elle peut également entraîner des réparations inutiles ou des défaillances catastrophiques. Déterminer quand un équipement entrera dans la phase d'usure est basé sur le taux théorique de défaillance plutôt que sur des statistiques réelles sur l'état de l'équipement spécifique. Cela entraîne souvent des maintenances coûteuses et complètement inutiles avant qu'un problème ne survienne réellement ou après que des dommages potentiellement catastrophiques ont commencé. De plus, cela entraîne un temps d'arrêt planifié beaucoup plus long et nécessite une gestion compliquée des stocks. Si l'équipement tombe en panne avant le temps d'usure estimé, il doit être réparé en utilisant des techniques de la RM, ce qui est beaucoup plus coûteux.

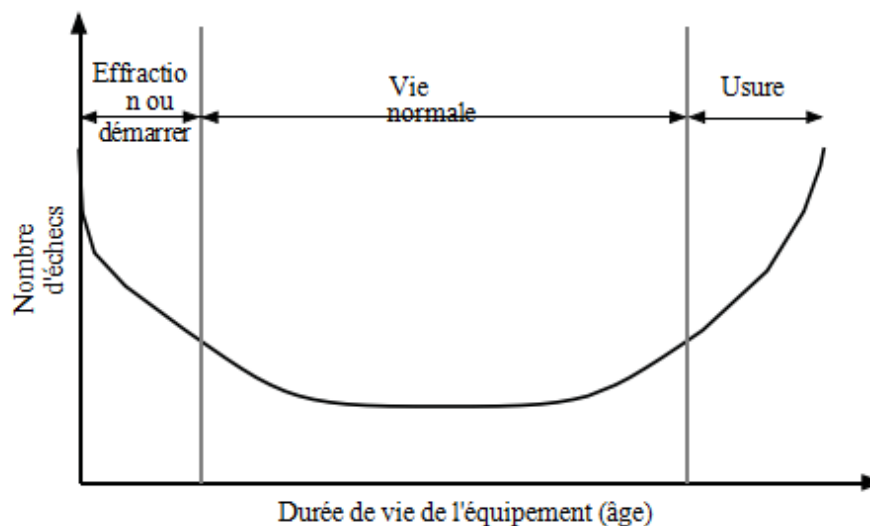


FIGURE II.3 – Courbe statistique en baignoire d'un équipement [4]

Dans l'étude menée par Drent, Kapodistria & Boxma en 2020 [24], les chercheurs se sont intéressés à la mise à jour des intervalles de maintenance, en se basant sur l'estimation de la dégradation de la machine paramétrée par un paramètre inconnu après les remplacements planifiés. Afin d'améliorer la précision des intervalles, ils ont utilisé des méthodes statistiques avancées et développent une politique bayésienne myope qui apprend presque sûrement le paramètre inconnu et converge vers la politique optimale avec une connaissance complète du paramètre.

Pour optimiser la planification et minimiser les retards des tâches critiques, des méthodes de classification multi-critères ont été développées. Par exemple, Lopes et al[25], présente une approche basée sur le risque pour hiérarchiser les tâches de la maintenance. Lorsque les actions de la PM dévient du planning prévu, il devient essentiel de prendre des mesures pour minimiser les coûts associés à cette déviation. Pour y parvenir, Sanoubar et al [26] ont identifié des politiques de maintenance optimales qui réduisent le degré de déviation attendu par rapport au planning initial et sa variance.

Néanmoins, cette approche de maintenance présente des limitations en entraînant le remplacement prématuré d'éléments du système qui n'ont pas encore atteint la fin de leur cycle de vie, ce qui entraîne des coûts inutiles [27].

Pour remédier à cela et réaliser des économies significatives, la maintenance prédictive s'est imposée comme une approche prometteuse.

2.2.2 La maintenance prédictive (PDM)

La Maintenance prédictive (PDM) est une stratégie de maintenance qui utilise des techniques avancées de surveillance et d'analyse pour évaluer l'état réel des équipements et prédire les défaillances à venir. Parmi les techniques utilisées, on retrouve l'analyse des vibrations, de la corrosion, thermique et des performances. En combinant ces méthodes et en utilisant des outils d'analyse avancés de ML et DL, les entreprises peuvent anticiper les pannes, planifier les interventions de maintenance de manière plus efficace et réduire les temps d'arrêt non planifiés comme illustré dans la figure II.2. Cela permet d'optimiser la disponibilité et la fiabilité des équipements, tout en réduisant les coûts de maintenance [21].

Étant donné que la grande majorité des processus industriels sont réalisés à l'aide de systèmes électriques et mécaniques, Coanda et al [21], proposent une vue d'ensemble des techniques utilisées pour améliorer les performances. L'étude de cas est menée sur un moteur de soufflerie et de compresseur d'air ainsi les résultats ont confirmé un désalignement de l'arbre des moteurs entraînant une perte de performances. Ce qui souligne l'importance de la PDM pour prévenir les défaillances et optimiser les performances des systèmes.

Pour détecter les anomalies qui surviennent principalement dans les équipements électriques susceptible de provoquer une surchauffe, Irfan et al [28] proposent une approche de PDM utilisant la thermographie infrarouge combinée à des techniques de ML. Cette méthode permet de surveiller les équipements pendant leurs fonctionnements afin de réduire les coûts liés aux pannes imprévues. Elle repose sur l'utilisation d'une caméra thermographie infrarouge qui capture et filtre la composante de lumière infrarouge, fournissant ainsi des images de profil thermique des équipements. L'analyse est réalisée sur un total de 150 photos, au sein desquelles environ 300 points d'intérêt sont identifiés. Ensuite, à l'aide de techniques de ML elles sont classées comme des défauts ou des paramètres de fonctionnement.

En outre, il est important de noter qu'une utilisation adéquate de l'équipement d'acquisition et de la qualité des données obtenues jouent un rôle crucial dans le succès des stratégies de la PDM.

Dans cette perspective, Giuseppe et al [29], proposent une méthode de PDM basée sur un système de surveillance sujet à des erreurs qui utilise une modélisation basée sur les chaînes de Markov pour représenter la dégradation progressive du système.

Cette méthode permet d'effectuer une simulation efficace pour étudier la corrélation entre l'efficacité de l'analyse des données, l'utilisation appropriée de l'équipement d'acquisition et la réduction des coûts de la maintenance prédictive. Les résultats de la simulation, basés sur l'analyse de 2000 éléments simulés, confirment que l'adoption de systèmes de maintenance prédictive peut améliorer les coûts de maintenance, même en tenant compte du coût de l'équipement analytique, à condition que les capteurs et les systèmes d'acquisition fournissent des données qualitatives fiables.

Staden et al [30], effectue une analyse de données de 4000 contrats de maintenance couvrant 3000 machines sur une période de 2004 à 2020 et proposent un processus décisionnel de Markov pour modéliser la PDM en fonction du temps écoulé depuis la dernière action de maintenance. En intégrant les données sur les défaillances passées et les actions de maintenance préventive, le modèle capture la dépendance temporelle des défaillances des machines, ce qui permet de recommander des actions de maintenance prédictive plus précises et adaptées en fonction de l'historique de chaque machine.

En utilisant les techniques de l'industrie 4.0 et du cloud, Wang, Jinjiang, et al [31], proposent une approche basée sur une architecture en nuage utilisant un agent mobile qui divise les tâches et de les exécutent en parallèle afin de réduire le volume de données brutes transmises et à améliorer la réactivité face aux changements dynamiques survenant dans les systèmes surveillés. Pour tester le système, six moteurs à courant alternatif ont été utilisés, cinq d'entre eux présentent différents défauts, tandis qu'un est utilisé comme norme. A l'aide des agents mobiles, des programmes logiciels sont envoyés pour être exécutés et collecter différents paramètres conformément à la méthodologie de maintenance prédictive. Les données sont ensuite traitées à l'aide d'algorithmes spécifiques, ce qui permet d'obtenir des caractéristiques de l'équipement surveillé. Les résultats ont confirmé l'efficacité du système et ont fourni des informations pertinentes sur l'état des moteurs analysés par rapport au moteur standard.

Pour améliorer la PDM en nuage et faciliter la collecte de données précises, l'analyse avancée et la prise de décision éclairée, il est essentiel de mettre en place des technologies de soutien telles que l'Internet des objets (IoT), les techniques du Big Data, les méthodes de DL, les systèmes embarqués, le web sémantique et la communication machine-machine, etc. Par exemple, Mourtzis et al [32], intègrent l'IoT ainsi que des techniques de détection dans leur système cyber-physique basé sur l'informatique en nuage, ce qui leur permet de proposer une planification adaptative de l'atelier et une maintenance basée sur l'état. Edrington et al [33], utilisent la technologie "MTConnect" pour collecter, analyser les données et notifier les événements machine.

Cachada, Ana, et al [34] proposent une structure innovante basée sur l'architecture OSA-CBM comprenant que quatre étapes, contrairement à la structure OSA-CBM, détaillé dans l'Annexe, qui comprend six étapes.

1. Acquisition de données : Cette couche fournit l'accès aux capteurs installés sur les équipements et collecte les données de fonctionnement, on utilisant les technologies de l'IoT et de l'IHM (interface homme-machine).
2. Analyse des données hors ligne : À cette étape, plusieurs technologies telles que l'apprentissage automatique, la technologie cloud et la technologie avancée d'analyse des données sont utilisées pour élaborer de nouvelles règles et procédures de surveillance. Selon le type de méthode et d'algorithmes choisis, on obtient des résultats plus ou moins efficaces.
3. Surveillance dynamique : Cette couche se compose de deux sous-ensembles : La visualisation et la détection précoce des dommages.
4. Le module de décision et de soutien : lorsqu'il est nécessaire d'effectuer des tâches ou des interventions, les informations sont transmises au module de décision et de soutien par le biais d'un outil d'organisation.

3 Méthodes prédictives basées sur l'apprentissage automatique (ML)

Avec l'augmentation continue des données, l'avancée des techniques de récolte telles que les capteurs IoT, la PDM basée sur les données devient de plus en plus attrayante. Les techniques de ML sont perçues comme une solution puissante pour extraire des connaissances utiles et prendre des décisions informées en exploitant ces données massives.

3.1 Les travaux basés sur l'algorithme de l'Arbre de décision (DT)

Benkercha et Moulahoum [35], ont utilisé l'algorithme de DT pour détecter et diagnostiquer les défauts dans les systèmes photovoltaïques connectés au réseau (GCPVS), les attributs utilisés comprennent la température ambiante, l'irradiation solaire et le rapport de puissance, tandis que les classes de défaut comprennent les défauts libres, les défauts de chaîne, les courts-circuits ou les défauts ligne-ligne, ils ont obtenue après entraînement une exactitude sur les données de test de 99,80%.

Bakir et al [36], ont utilisé l'arbre de régression (RT) pour développer un modèle de prédiction de la RUL qui calcul la durée de vie restante d'un équipement ou d'un système avant une défaillance, en se basant sur des analyses de données et de modèles prédictifs pour plusieurs composants, l'algorithme a démontré des performances assez satisfaisantes.

Sharma et al [37], ont utilisé l'algorithme de DT de classification dans le but de déterminer les machines susceptibles de tomber en panne sur un dataset de PDM comprenant 10 000 entrées avec 10 caractéristiques. La cible est binaire (0 pour pas de défaillance, 1 pour défaillance), tandis que le type de défaillance est multi-classes. ils ont acquis un résultat d'exactitude de 97,9% et 98% pour la précision, le rappel et le score F1.

3.2 Les travaux basés sur l'algorithme des machines à vecteurs de support (SVM)

Wei et al [38], ont construit un modèle d'espace d'état pour représenter la dynamique de vieillissement des batteries, ils ont utilisé la régression par vecteurs de support (SVR) pour prédire la RUL. Les performances obtenues de ce modèle (SVR) ont été comparées à celles d'un modèle basé sur (ANN) démontrant que le modèle (SVR) présentait moins d'erreur d'estimation.

Sharma et al [37], ont utilisé un SVM sur le même dataset de PDM afin de classifier les machines susceptible de représenter des défaillances, les résultats expérimentaux ont montré une exactitude de de 88,1% , 88% pour la précision, le rappel et le score F1.

3.3 Les travaux basé sur l'algorithme des K plus proches voisins (K-NN)

chen et al [39] , ont développé la méthode "CMWEKNN " qui combine plusieurs sources de preuves pour prendre des décisions dans le cadre de la surveillance de l'état des cen-

trales électriques en se basant sur la règle k-NN évidentielle (EKNN) qui est une extension du classificateur K-NN classique qui prend en compte l'incertitude associée à chaque voisin lors de la prédiction en attribuant des poids aux voisins en fonction de leur similarité avec l'échantillon de test et de leur crédibilité, cette méthode vise à fournir une détection précoce des défauts et une surveillance continue de l'état des équipements dans les centrales électriques.

Sharma et al [37], ont utilisé un K-NN sur le dataset de la PDM de classification des défaillances des machines, ils ont obtenus des résultats avec une exactitude de 95,5% et 96% pour la précision, le rappel et score F1.

4 Méthodes prédictives basées sur l'apprentissage profond (DL)

Contrairement aux approches traditionnelles du ML, le DL utilise des réseaux de neurones artificiels profonds pour apprendre des représentations hiérarchiques des données. Grâce à cette capacité à extraire des caractéristiques complexes. Le DL résout des problèmes plus complexes, traite des jeux de données massifs avec une précision accrue.

4.1 Les travaux basés sur l'algorithme de l'Auto-encodeur (AE)

Jia et al [40] ont utilisé un réseau de connexion locale (connexion avec les neurones voisins) construit par Auto-encodeur (AE) épars (nombre de neurones actifs limité) sur des données normalisées pour extraire automatiquement des caractéristiques des signaux de vibration bruts, une contrainte d'orthogonalité douce est ajoutée dans la fonction de coût pour apprendre des caractéristiques différentes (non redondantes). Une exactitude de test de 99,43% a été obtenue.

Ma et al [41], ont utilisé un AE de couplage profond (DCAE) pour trouver une caractéristique commune entre les signaux de vibration et acoustiques, ils ont obtenu une exactitude de classification de 94,3% pour l'évaluation de la santé des engrenages. Haidong et al [42], ont utilisé un AE de débruitage et d'une régression logistique pour prédire la RUL, une exactitude de prédiction de 95,20% a été obtenue avec un ELM (machine d'apprentissage extrême) et 95,03% avec softmax.

Ren et al [43], ont utilisé un modèle d'extraction de caractéristiques multidimensionnelles avec un AE pour caractériser la dégradation de la santé de la batterie lithium-ion de la NASA, en intégrant un AE et un ANN pour la prédiction de la RUL des batteries, ils ont obtenu une exactitude de 88,20% et une Racine carrée de l'erreur Quadratique Moyenne (RMSE) de 11,80.

4.2 Les travaux basé sur l'algorithme des Réseaux de neurones convolutifs (CNN)

Liu et al [44], ont proposé un modèle "CNNDM-1D" qui est un réseau de neurones convolutif peu profond conçu pour diagnostiquer les défauts à partir de signaux de vibration bruts, il se compose de deux couches convolutives, deux couches de mise en commun et une couche entièrement connectée, simplifiant ainsi l'apprentissage des caractéristiques et la classification en un seul processus, ce modèle a atteint une exactitude de 99.89% dans les expériences.

Qin et al [45], ont développé un modèle CNN unidimensionnel (1D-CNN) end-to-end pour diagnostiquer les types de pannes à partir de signaux de vibration bruts, ils ont atteint une exactitude d'environ 99.32% après ajustement des hyper-paramètres.

4.3 Les travaux basé sur l'algorithme des réseaux de neurones récurrents (RNN)

Muneer et al [46], ont appliqué sur le jeu de données Simulation de Système Aéropropulsif Modulaire Commercial (C-MAPSS) de la NASA pour prédire la RUL, un LSTM avec mécanisme d'attention (ISE). Ils se compose de deux couche LSTM (30 neurones chacune), deux couches dense avec fonction d'activation Relu , une couche de dropout, et couche de sortie avec unité linéaire afin de permettre au modèle de se concentrer sur les données pertinentes ,ils ont acquis une RMSE de 11,23 et une Erreur absolue moyenne (MAE) de 5,47.

Hesabi et al [47], ont utilisé le réseau neuronal LSTM binaire pour classer la RUL des composants en deux catégories : classe 0 lorsque la RUL est inférieure ou égale à une fenêtre de temps prédéfinie et classe 1 lorsque la RUL est supérieure à cette fenêtre, ils ont appliqué l'algorithme sur le jeu de données de C-MAPSS de la NASA, le modèle a été compilé avec l'optimiseur Adam et une fonction de perte d'entropie croisée binaire. Ils ont inclus des connexions "peephole" dans les cellules LSTM pour une meilleure stabilité. Ils ont obtenu une exactitude de 96,77% , 94,12% pour la précision ,96.96% pour le rappel et 95.52% de score F1.

Asif et al [48], ont développé un modèle de dégradation linéaire par morceaux pour estimer la RUL des systèmes basés sur des données des capteurs du jeu de données C-MAPSS , le modèle proposé se compose de quatre couches LSTM, de couches de suppression (dropout) à 0.1 pour éviter le sur-ajustement, et de la couche entièrement connectées suivies d'une couche de régression avec un taux d'apprentissage à 0.2. Ils ont obtenu sur FD0001 un ensemble des données de C-MAPSS une RMSE atteignant 7,78 et une valeur de fonction score à 100.

Zhuang et al [49], ont utilisé les données C-MAPSS pour prédire la RUL de moteurs. Elle utilise 200 moteurs du sous-ensemble "FD001", dont 100 pour l'entraînement et 100 pour les tests. Les données d'entrée incluent des signaux de capteurs et des variables opérationnelles, sans extraction de caractéristiques. Ils ont utilisé une architecture avec une couche Mémoire à Long Terme et Court Terme Bidirectionnels (BiLSTM) à 2 couches cachées (20 neurones chacune), entraîné avec RMSprop (taux d'apprentissage à 0,01, probabilité d'abandon : 0,2) et une couche Fully Connected (FNN) avec une couche cachée (5

neurones) et ils ont atteint une exactitude de 70%, 12.7 de RMSE et une valeur de 234.9 de la fonction score.

4.4 Les Méthodes hybrides

Bien que les méthodes d'apprentissage profond aient montré leur potentiel, mais leurs mise en œuvre dans les environnements industriels reste limitée en raison des exigences élevées en termes de ressources informatiques, d'interprétabilité et de gestion des données complexes causée par l'expansion continue de l'IOT). Néanmoins, l'intégration de méthodes hybrides offre une solution prometteuse pour surmonter ces défis.

Flen et al [50], ont combiné des méthodes statistiques (ETS) et l'algorithme ARIMA ainsi que des algorithmes de LSTM appliqué sur des ensembles de données provenant de trois secteurs industriels différents, à savoir la fabrication, l'énergie et les transports. Le modèle a obtenu des résultats avec une précision de 93% et un rappel de 91% pour la détection des pannes ainsi que un F1 score de 92%.

Liu et al [51] ont utilisé un AE convolutif de débruitage avec une seule dimension (DCAE-1D) pour le diagnostic des défauts des machines tournantes combiné avec un CNN unidimensionnelle (AICNN-1D) pour le diagnostic des défauts en utilisant les signaux de sortie débruités du DCAE-1D. Avec ce dernier, les exactitudes de diagnostic de l'AICNN-1D ont atteint respectivement 96,65% et 97,25%.

Song et al [52], ont intégré un AE et un BiLSTM pour améliorer la précision de la prédiction de la RUL des moteurs turbofan. AE a été utilisé comme extracteur de caractéristiques pour comprimer les données de surveillance des conditions, et le BiLSTM a été conçu pour capturer les dépendances à long terme bidirectionnelles des caractéristiques, ils atteignent une valeur fonction de score de 261.

Guo et al [53], ont développé un réseau de Transfert par apprentissage (TL) convolutif profond (DCTLN) avec deux modules, le premier utilise un CNN avec une seule dimension pour extraire automatiquement des caractéristiques des données de vibration brutes et identifier les états de santé, le second comprend un classificateur de domaine et des couches faciliter l'apprentissage de caractéristiques invariantes au domaine, ils ont acquis un résultat d'exactitude de 86,3%.

Pan et al [54], ont combiné un CNN unidimensionnel et un LSTM dans une structure unifiée, la sortie du CNN est introduite dans le LSTM pour identifier les types de défauts des paliers. Les résultats ont montré que le taux d'exactitude plus de 99%.

Habib et kamal [55], proposent une architecture utilisant des CNN-LSTM en série avec un pré-traitement des données de C-MAPSS pour prédire la RUL, en intégrant également une couche entièrement connectée. Le modèle est entraîné en utilisant la Erreur quadratique moyenne (MSE) comme fonction de perte, avec des paramètres spécifiés tels que la fonction

d'activation relu, un dropout de 0.2, un optimiseur Adam et 128 nœuds par couche LSTM. Ils ont obtenu avec ce modèle une RMSE sur FD0003 de 15.31.

Peng et al [56] ont appliqué 1-FCLCNN-LSTM se traduisant par un réseaux neuronaux convolutionnels unidimensionnels CNN avec couche de convolution complète appliqué aussi sur les même données combiné avec un LSTM dans le but de la prédiction de la RUL, ils ont obtenu une RMSE de 9,99.

Ciu et al [57], ont proposé un modèle CNN-ISE-BiLSTM pour la prédiction des pièces détachées de maintenance. La structure du modèle comprend principalement une couche d'entrée, une couche de couche convolutive pour extraire les caractéristiques visuelles, une couche ISE-Attention pour améliorer l'espace des caractéristiques et se concentrer sur des parties spécifiques des données, une couche de dépliage de séquence, une couche d'aplatissement, une couche BiLSTM et une couche de sortie pour capturer les dépendances séquentielle. Ce modèle subit un apprentissage de migration divisé en 7 couches, sur 10 ensembles de données de consommation de pièces de rechange, sur une durée 5 ans pour 6 modèles provenant de 4 unités comme ensemble de données expérimental, avec une granularité temporelle hebdomadaire ils ont obtenu un résultat avec une RMSE de 0,5036 et une MAE de 0,3826.

5 Analyse comparative

Dans le cadre de notre étude, nous avons réalisé une analyse comparative des performances des différentes méthodes basées sur les données. Nous avons évalué les performances des différents algorithmes mentionnés dans les travaux de référence en utilisant les métriques de performance suivantes [17] :

- Exactitude (Accuracy) : L'exactitude est une mesure de la justesse avec laquelle un modèle de classification prédit les étiquettes attribuées aux entrées d'un ensemble de données. Cela permet de déterminer la précision globale du modèle. Elle est calculée en divisant le nombre de prédictions correctes par le nombre total de prédictions.

Exactitude = (Nombre de prédictions correctes) / (Nombre total de prédictions)

$$Accuracy = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}}$$

- Précision (Precision) : La précision mesure la proportion de prédictions positives correctes (les vrais positifs) parmi toutes les prédictions positives effectuées. Elle est calculée en divisant le nombre de vrais positifs par la somme des vrais positifs et des faux positifs.

$$Precision = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux positifs}}$$

- Rappel (Recall) : Le rappel mesure la proportion de vrais positifs correctement prédits parmi tous les vrais positifs présents dans les données réelles. Il est utilisé pour évaluer la capacité du modèle à identifier tous les vrais positifs. Le rappel est calculé en divisant le nombre de vrais positifs par la somme des vrais positifs et des faux négatifs.

$$Recall = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}}$$

- Score F1 (F1 Score) : Le F1 score est une mesure qui combine à la fois la précision et le rappel en une seule valeur. Il fournit une mesure globale de la performance du modèle en tenant compte à la fois des faux positifs et des faux négatifs. Le F1 score est calculé en utilisant la formule :

$$F1 = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}$$

- Racine carrée de l'erreur quadratique moyenne (RMSE) : Le RMSE est une mesure de l'écart entre les valeurs prédites et les valeurs réelles. Il est couramment utilisé pour évaluer les performances d'un modèle de régression. Le RMSE est calculé en prenant la racine carrée de la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Valeurs prédites} - \text{Valeurs réelles})^2}$$

- Erreur absolue moyenne (MAE) : Le MAE est une autre mesure de l'écart entre les valeurs prédites et les valeurs réelles. Il est également couramment utilisé pour évaluer les performances d'un modèle de régression. Le MAE est calculé en prenant la moyenne des valeurs absolues des différences entre les valeurs prédites et les valeurs réelles.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\text{Valeurs prédites}_i - \text{Valeurs réelles}_i|$$

- La fonction Score Fonction score (SF) est utilisée pour évaluer la pertinence des prédictions par rapport à un temps cible. Elle est conçue pour pénaliser les prédictions tardives plus sévèrement que les prédictions précoces.

La fonction est calculée comme la somme des scores individuels s_m pour chaque prédiction m , où s_m dépend du retard d_m par rapport au temps cible.

SF est calculé comme :

$$SC = \sum_{m=1}^M s_m$$

où s_m est définie comme :

$$s_m = \begin{cases} e^{-d_m/13} - 1, & \text{if } d_m < 0 \\ e^{d_m/10} - 1, & \text{if } d_m \geq 0 \end{cases}$$

Dans cette formule :

Le d_m représente le retard de la prédiction m par rapport au temps cible.

Un retard négatif ($d_m < 0$) signifie que la prédiction est anticipée, tandis qu'un retard positif ou nul ($d_m \geq 0$) indique que la prédiction est tardive ou exacte.

Ainsi e est la base du logarithme naturel (environ 2.71828).

En effectuant une revue de la littérature, nous avons résumé les performances des différentes approches évaluées en termes de ces métriques de performance. Les résultats obtenus sont résumés dans le tableau suivant :

Travaux	Approche	Méthode	Métriques						
			Exactitude	RMSE	Précision	Rappel	Score F1	SF	MAE
Benkercha et al [35] Bakir et al [36] Sharma et al [37]	DT	RT	99,8%	–	–	–	–	–	–
–			–	–	–	–	–	–	
97.9%			–	98%	98%	98%	–	–	
Wei et al [38] Sharma et al [37]	SVM	SVR	–	–	–	–	–	–	–
88.1%			–	88%	88%	88%	–	–	
Chen et al [39] Sharma et al [37]	K-NN	EKNN	–	–	–	–	–	–	–
95,5%			–	88%	88%	88%	–	–	
Jia et al [40] Ma et al [41] Haidong et al [42]	AE	DCAE	99.43%	–	–	–	–	–	–
–		AE-Softmax	94.3%	–	–	–	–	–	–
–		AE-ELM	95.03%	–	–	–	–	–	–
–		–	95.20%	–	–	–	–	–	–
Ren et al [43]	–	–	88.20%	11.80	–	–	–	–	–
Qin et al [45] Liu et al [44]	CNN	1D-CNN	99.32%	–	–	–	–	–	–
–		CNNDM-1D	99.89%	–	–	–	–	–	–
Hesabi et al [47] Zhuang et al [49]	RNN	LSTM	96.77%	–	94.12%	96.96%	95.52%	–	–
–		FNN-BILSTM	70%	12.70	–	–	–	234.9	–
–		MCLSTM	–	13.70	–	–	–	–	–
–		LSTM	–	7.78	–	–	–	100	–
Muneer et al [46]	–	LSTM-ISE	–	11,23	–	–	–	–	5.47
Flen et al [50]	TSAE	ETS-ARIMA-LSTM	93%	–	–	91%	92%	–	–
Liu et al [51]	AE et CNN	DCAE-1D	96.65%	–	–	–	–	–	–
–		AICNN-1D	97.25%	–	–	–	–	–	–
–		–	–	–	–	–	–	–	–
Song et al [52]	AE et RNN	AE-BILSTM	–	13.63	–	–	–	261	–
Guo et al [53]	CNN et TL	DCTLN	86.3%	–	–	–	–	–	–
Pan et al [54] Habib et kamal [55] Peng et al [56] Ciu et al [57]	CNN et RNN	CNN-1D-LSTM	99%	–	–	–	–	–	–
–		CNN-LSTM	–	15.31	–	–	–	–	–
–		1-FCLCNN-LSTM	–	9.99	–	–	–	–	–
–		CNN-ISE-BILSTM	–	0.5036	–	–	–	–	0.3826

TABLE II.1 – Analyse comparative des études synthétisées.

Le tableau II.1 expose une analyse comparative des recherches synthétisées portant sur l'utilisation des différents algorithmes ML et DL dans le contexte de la PDM. L'objectif est de prédire les défaillances des machines, que ce soit en anticipant le temps restant avant la défaillance RUL ou en classifiant les pannes des machines.

Dans le domaine du ML et DL pour la PDM, les performances des différentes techniques peuvent être comparées en fonction de plusieurs métriques clés :

- Exactitude
 - Les DT ont montré une excellente exactitude de 99,80% dans la détection et diagnostique des défauts, indiquant une capacité élevée à prédire correctement les classes des défauts.
 - Les CNN ont également obtenu des résultats impressionnants avec une exactitude d'environ 99,89% dans le diagnostique des défauts à partir de signaux de vibration bruts, soulignant leur efficacité dans la classification des défauts.
 - Les AE ont affiché des performances élevées avec une exactitude de test de 99,43% dans l'extraction automatiquement des caractéristiques des signaux de vibration bruts, démontrant leur capacité à extraire des caractéristiques pertinentes des données.
- RMSE
 - Le modèle CNN-ISE-BILSTM a obtenu une RMSE de 0,5036 et une MAE de 0.3826,

ce qui indique une très faible erreur moyenne de prédiction. Cela suggère une excellente précision dans la prédiction des pièces détachées de maintenance.

- Le modèle LSTM avec couches de dropout, couches entièrement connectées et couche de régression a atteint une RMSE de 7,78, démontrant une précision élevée dans la prédiction de la RUL des systèmes basés sur des données de capteurs .
 - Le modèle LSTM combiné avec le CNN à savoir le 1-FCLCNN-LSTM a obtenu une RMSE de 9,99. Bien que légèrement plus élevée que les deux premiers modèles, cette valeur indique toujours une bonne précision dans la prédiction de la RUL.
- Précision , Rappel et f1 score :
 - Le modèle utilisant le DT pour prédire les défaillances des machines, obtenant les valeurs les plus haute en matière de précisions,rappel et f1 score, à savoir une précision de 98% témoignant de la capacité du modèle à éviter les faux positifs .Un rappel de 98% souligne la capacité du modèle à détecter correctement la quasi-totalité des machines réellement en situation de défaillance, minimisant ainsi les faux négatifs, et un score F1 de 98% combine ces deux mesures, reflétant une performance globalement élevée du modèle dans la prédiction des défaillances.
 - Le modèle LSTM basique binaire pour classer la RUL des composants en deux catégories a obtenu des valeurs plus basse que celle de DT, mais assez satisfaisantes. Avec un rappel de 96,96%, une précision de 94,12% et un score F1 de 95,52%, démontrant une bonne performance dans la classification de la RUL .
 - Fonction Score (SF) :
 - Le modèle utilisant un LSTM avec couches de dropout , couches entièrement connectées et couche de régression est conçu dans le bute de l'estimation de la RUL, c'est le plus performant avec une fonction score basse de 100. Cela indique qu'il pénalise moins les prédictions tardives et il est plus précis.
 - Les modèles (FNN BILSTM) et (AE BILSTM) pour la prédiction de la RUL ont des valeurs fonction scores plus élevés (234.9 et 261 respectivement), ce qui indique qu'ils sont légèrement moins performants.

En comparant les différents résultats, il est évident que les modèles basés sur les LSTM et leurs variantes offrent les meilleures performances en termes de RMSE et de fonction score pour l'estimation de la RUL sur les données. Plus précisément, le modèle développé par Asif et al [48] avec une RMSE de 7,78 et une fonction score de 100 est le plus performant parmi ceux étudiés. Ainsi que le modèle devloppé par Cui et al [57] utilisant un CNN-iSE-BILSTM pour l'estimation de la RUL obtenant les plus basses valeurs d'erreur à savoir une RMSE de 0.5036 et MAE de 0.3826. Aussi, Les modèles hybrides, tels que ceux combinant AE et BILSTM, montrent également des résultats prometteurs avec des fonctions score très basses, indiquant une meilleure précision dans les prédictions.

Les modèles basés sur les DT et les SVM ont montré une très bonne performance en termes de classification des défauts avec des metriques de précision, rappel et F1 score élevés, mais sont moins efficaces pour les prédictions de séries temporelles complexes telles que la RUL.

6 Conclusion

Ce chapitre a mis en lumière les progrès récents dans la maintenance prédictive (PDM) industrielle grâce aux algorithmes de l'apprentissage automatique (ML) et de l'apprentissage profond (DL). Grâce à l'exploitation efficace des données des capteurs et des informations provenant des applications métier telles que la gestion de la maintenance assistée par ordinateur (GMAO), la science des données permet d'anticiper les défaillances et les pannes, facilitant ainsi une planification proactive des activités de maintenance industrielle. Cette approche promet de réduire les temps d'arrêt imprévus, d'optimiser les ressources et de prolonger la durée de vie des équipements.

Au cours de ce chapitre, nous avons défini la maintenance industrielle et exploré ses différentes formes, notamment la maintenance corrective, préventive et prédictive. Nous avons réalisé un état de l'art des recherches dans le domaine de la maintenance prédictive, en nous concentrant sur les techniques d'apprentissage profond et d'apprentissage automatique, qui jouent un rôle crucial dans la prévision des défaillances et l'optimisation des interventions de maintenance.

Nous avons également examiné les différentes architectures utilisées pour la mise en œuvre de la maintenance prédictive, en analysant les modèles et approches adoptés dans ce domaine. Pour finir, une analyse comparative détaillée des travaux existants a été présentée, mettant en lumière les avancées, les défis et les perspectives futures de la maintenance prédictive.

Dans le chapitre suivant, nous allons voir notre solution proposée pour la mise en place de cette approche, en détaillant les étapes de son développement.

Chapitre III

Solution proposée

1 Introduction

Face à la récurrence des pannes d'équipements et à leurs impacts néfastes sur la production, l'industrie algérienne a impérativement besoin de nouvelles approches pour anticiper et prévenir ces défaillances.

Ce chapitre s'attache à exposer en détail l'approche développée dans le cadre de ce projet visant à gérer la maintenance industrielle. Notre solution proposée vise à exploiter de manière approfondie des données collectées au sein des unités de production afin d'identifier les tendances et motifs précurseurs des pannes imprévues, permettant ainsi de réaliser des prévisions précises de la durée de vie restante des différents équipements RUL.

Ce chapitre se divise en deux parties. La première présente les données de Cevital et analyse les tendances entre ces données ainsi que les défis rencontrés. La deuxième partie décrit en détail la méthodologie employée, les sources de données exploitées et les détails de l'implémentation du modèle LSTM.

2 Analyse de données Cevital

Dans le cadre de ce projet, nous avons bénéficié d'un stage au sein du groupe agroalimentaire Cevital de Bejaia.

Au cours de ce stage, nous avons eu la chance d'accéder à un ensemble de données détaillées relatives aux interventions de maintenance préventive et réactive des équipements industriels. Ces données proviennent du système de gestion de maintenance GMAO Coswin 8.0, utilisé par le service de maintenance.

2.1 Présentation des données de Cevital :

Voici ci-dessus un aperçu des rapports GMAO de Cevital, couvrant une période de quatre ans, allant de 2020 à 2023 :

- Créateur : Le créateur de l'ordre de travail, c'est-à-dire la personne qui a initié l'intervention.
- Code Équipement : L'identifiant unique de l'équipement concerné par l'intervention.
- Description Équipement : Une brève description de l'équipement.
- Code Process : Le code du processus associé à l'intervention.
- Code Intervention : Un code spécifique pour chaque type d'intervention (préventive, corrective, etc.).
- Intervention / Anomalie : Indique si l'entrée correspond à une intervention planifiée ou à une anomalie inattendue.
- N° D.I (Numéro de la Demande d'Intervention) : Le numéro de la demande d'intervention, le cas échéant.
- Type d'Intervention : La catégorie de l'intervention (par exemple, maintenance préventive, réparation, etc.).
- Machine arrêtée? : Indique si l'équipement a été arrêté pendant l'intervention.
- Début arrêt machine : Date et heure du début de l'arrêt de la machine.
- Début intervention : Date et heure du début de l'intervention.
- Fin intervention : Date et heure de la fin de l'intervention.
- Remise en service : Date et heure de la remise en service de l'équipement.
- Temps total d'intervention (h) : La durée totale de l'intervention en heures.
- Temps d'arrêt (min) : La durée de l'arrêt de la machine en minutes.
- Code Équipement Système : L'identifiant unique du système d'équipement auquel appartient l'équipement.
- Description Équipement Système : Une description plus générale du système d'équipement.
- Entité : L'entité ou le département responsable de l'intervention.
- Description Entité : Une description plus détaillée de l'entité ou du département.
- Commentaires : Toute remarque ou information supplémentaire liée à l'intervention.

2.2 Visualisation des tendances entre les donnée

Afin d'identifier des tendances et des modèles significatifs dans ces données, nous avons utilisé Power BI, un outil puissant de visualisation de données.

- Power BI : C'est une plateforme d'analyse de données et de visualisation créée par Microsoft. Elle permet aux utilisateurs de connecter diverses sources de données, de les transformer et de les visualiser de manière interactive pour obtenir des insights significatifs. Power BI offre une gamme de fonctionnalités, y compris des tableaux de bord interactifs, des rapports dynamiques, des visualisations avancées et des fonctionnalités de partage et de collaboration. En combinant des données provenant de différentes sources telles que des fichiers Excel, des bases de données, des services cloud, etc., Power BI permet aux entreprises et aux utilisateurs

individuels de prendre des décisions éclairées basées sur des analyses de données précises et personnalisées.

Voici les principales observations tirées de nos analyses :

La figure III.2 présente la distribution du parc d'équipements entre les différentes entités.

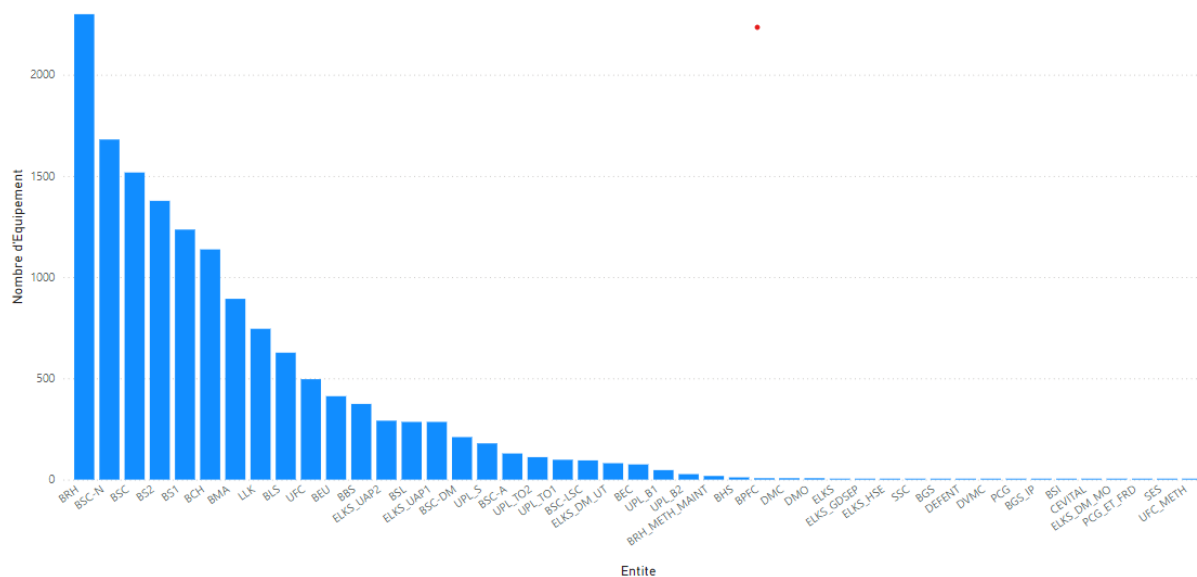


FIGURE III.2 – Le nombre d'équipements pour chaque entité

La figure III.3 montre une évolution fluctuante du nombre d'interventions au fil des quatre années présentées, avec des périodes de forte activité et d'autres plus calmes, mais une tendance générale à la hausse, passant d'environ 500 interventions en début de période à plus de 1500 en juillet 2023. On observe également des variations saisonnières, avec un nombre d'interventions plus élevé pendant les mois d'été. Cette analyse nous a permis de mieux comprendre l'évolution de la charge de travail sur la période considérée.

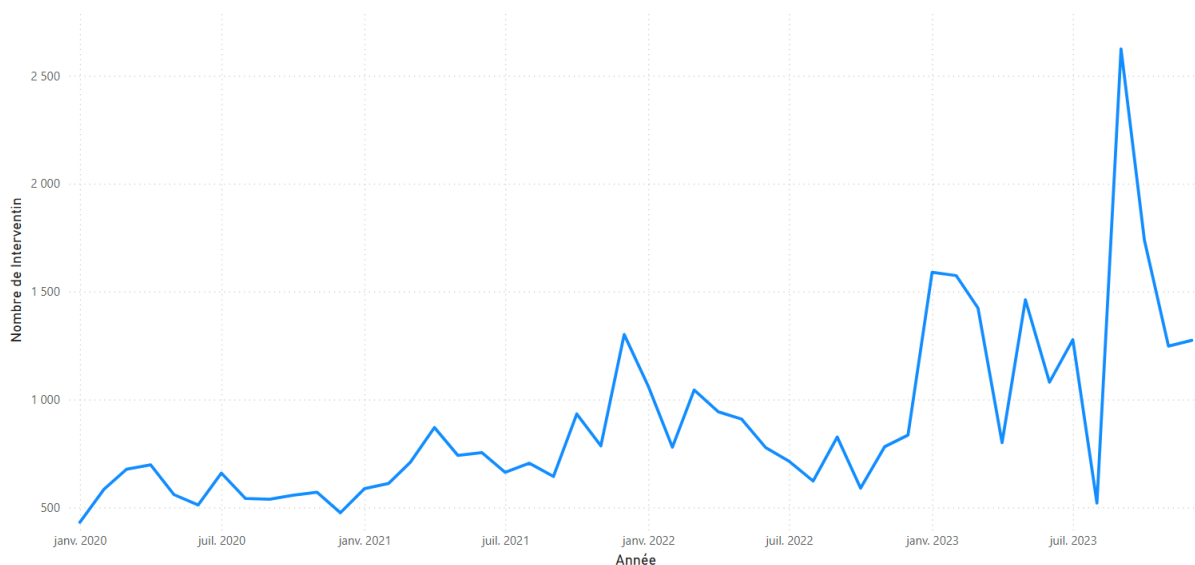


FIGURE III.3 – Le nombre d’interventions sur les équipements sur quatres années

La figure III.4 présente le nombre d’interventions effectués sur les différentes entités. On peut observer une grande disparité dans le nombre d’interventions entre les différentes entités, avec les deux entités "Bsc" et "BSC-N" se démarquant très nettement avec plus de 42000 interventions, tandis que la majorité des autres entités ont un nombre d’interventions beaucoup plus faible, descendant jusqu’à la valeur 1.

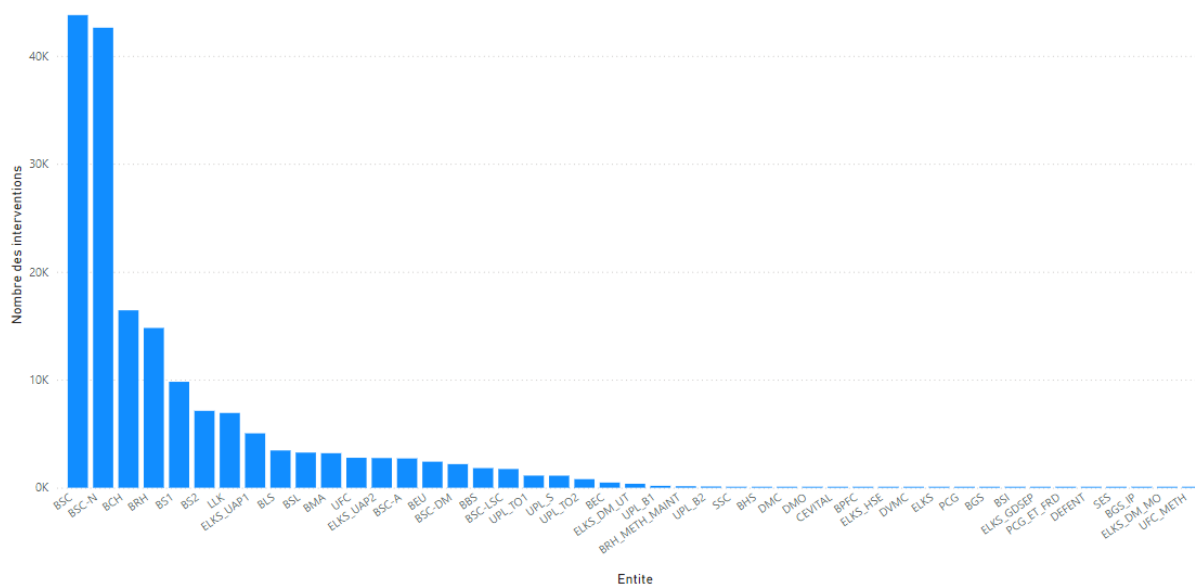


FIGURE III.4 – La somme des interventions effectuées par entité

La figure III.5 présente un tableau détaillé pour chaque entité le nombre d’équipements, le nombre total d’interventions effectuées sur ces équipements, ainsi que le volume d’interventions moyen par équipement (MoyIntervention/Equ) . Cette dernière métrique calcule le ratio entre le nombre total d’interventions et le nombre d’équipements pour chaque entité, permettant ainsi d’identifier les entités ayant à la fois un parc

d'équipements important et un volume d'interventions élevé par équipement, ce qui peut aider à cibler celles nécessitant le plus d'attention.

Entite	NbrEqu	NbrInterventionsEquipement	MoyIntervention/Equ
BSC	1517	43773	28,85
SSC	3	84	28,00
BSC-N	1680	42609	25,36
BSC-A	129	2711	21,02
BSC-LSC	95	1733	18,24
ELKS_UAP1	285	5035	17,67
CEVITAL	1	16	16,00
BCH	1137	16434	14,45
UPL_TO1	98	1123	11,46
BSL	285	3254	11,42
BSC-DM	210	2187	10,41
ELKS_UAP2	291	2748	9,44
LLK	745	6924	9,29
BS1	1235	9812	7,94
BRH_METH_MAINT	18	135	7,50
UPL_TO2	111	792	7,14
BRH	2299	14797	6,44
BEC	75	477	6,36
UPL_S	179	1116	6,23
BEU	412	2412	5,85
UFC	496	2774	5,59
BLS	627	3452	5,51
BS2	1377	7122	5,17
DMC	7	35	5,00
DMO	7	34	4,86
BBS	374	1815	4,85
ELKS_DM_UT	81	367	4,53
Total	13319	133199	10,00

FIGURE III.5 – le volume d'interventions moyennes par équipement pour chaque entité

La figure III.6 ci-dessus démontre le nombre d'arrêts par nombres d'interventions ou nous remarquons que sur un total de 133199 interventions, 68172 interventions ont conduit à l'arrêt des équipements

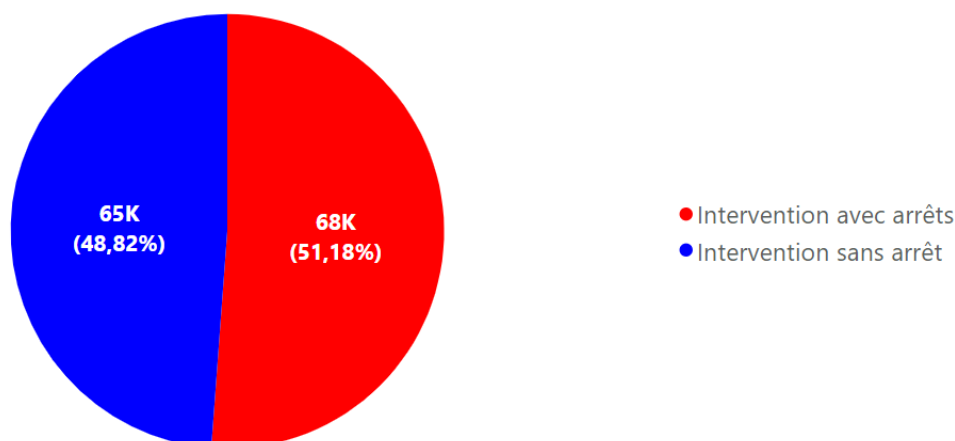


FIGURE III.6 – Le nombre d'interventions

Cela signifie que plus de la moitié des interventions ont nécessité l'arrêt des machines, soulignant l'impact significatif des pannes sur la production.

2.3 Défis rencontrés

En effectuant une analyse minutieuse de l'historique détaillé des interventions fournies par l'entreprise Cevital, nous avons constaté qu'il y a un grand nombre de données manquantes, les équipements ne disposent pas de capteurs qui enregistrent les taux d'humidité, la pression et la chaleur à intervalles définis ou ces paramètres sont pourtant essentiels pour détecter les défaillances potentielles comme vue dans les travaux synthétisés dans l'état de l'art. Ainsi, les données disponibles ne nous permettent pas de prédire la RUL des équipements.

Pour pallier ce problème, il est nécessaire de passer par l'étape de génération de données. Cependant, cette approche présente des risques, car les données générées pourraient ne pas être fiables, ce qui conduirait à des résultats relativement douteux.

Pour améliorer la fiabilité des prédictions de la maintenance, il serait judicieux de prendre des mesures supplémentaires. Cela pourrait impliquer l'installation de capteurs sur les équipements ainsi passer vers l'industrie 4.0 pour collecter des données précises sur les conditions de fonctionnement. De plus, une validation rigoureuse des données générées et une analyse approfondie de leur qualité seraient nécessaires pour garantir la fiabilité des résultats obtenus.

Face à cette réalité, nous avons décidé d'explorer d'autres sources de données. On a opté pour deux ensembles de données provenant du site web Kaggle, qui étaient mieux adaptés à notre objectif. Ces ensembles de données contenaient des informations plus structurées et des séquences temporelles plus cohérentes, incluant des mesures en temps réel de paramètres spécifiques des machines (pression, vibrations, tension, etc), ce qui permet d'entraîner un modèle efficacement et avoir des prédictions correctes.

3 Approche proposée

Après avoir étudié les travaux existants dans le domaine de la maintenance prédictive, nous avons proposé une solution pour remplacer les anciennes méthodes de maintenance réactive et préventive, qui se sont révélées inefficaces avec le temps.

3.1 Hypothèse de la solution

Pour que notre système soit réalisable, nous supposons la validité des hypothèses suivantes :

- Les données sont réelles et non générées artificiellement. Les données collectées directement des machines en fonctionnement permettent de réaliser des diagnostics précis et de prendre des décisions basées sur des situations concrètes, minimisant ainsi les risques d'erreurs liées à des simulations ou des données synthétiques.
- Le modèle fonctionne sur des jeux de données contenant données statistiques. Les moyennes et les écarts types sont des indicateurs essentiels qui permettent de résumer et de comprendre les variations des données. Notamment il peut détecter des anomalies et des tendances, facilitant ainsi la PDM et l'optimisation des performances des équipements.
- Les machines et équipements font partie de l'industrie 4.0, intégrant des capteurs IoT capables de capturer les caractéristiques influentes et apparentes sur l'équipement tout au long de sa durée de vie. Ces capteurs permettent une surveillance proactive, une PDM et une amélioration continue des processus. Les données recueillies peuvent être analysées pour optimiser les cycles de production et prolonger la durée de vie des équipements.
- Les données sont capturées à des instants précis et transmises à travers des outils tel que le cloud. La capture de données à des intervalles précis est cruciale pour garantir une surveillance efficace et réactive. Les données doivent être stockées dans des formats accessibles et structurés, permettant une analyse rapide et efficace. La précision dans la collecte des données assure la détection précoce de défaillances potentielles et l'optimisation des interventions de maintenance.
- L'ingénieur de maintenance dispose d'un outil informatique (ordinateur, tablette, etc.) afin de pouvoir visualiser les différents résultats des machines. Ces outils permettent à l'ingénieur d'accéder aux données en temps réel, d'effectuer des diagnostics et d'analyser les performances des machines. Ils facilitent également le stockage des historiques de données, permettant ainsi de détecter des tendances et de prévoir les pannes potentielles.

3.2 Les Jeux de données utilisées

Nous avons choisi d'utiliser les ensembles de données suivants :

- **Le jeu de données de Microsoft Azure "Telemetry"** : Ce premier dataset, intitulé "Predictive Useful Life based into telemetry", provient de "Predictive Maintenance

Modelling Guide Experiment", un cas d'étude mis à disposition par Microsoft dans le cadre de sa plateforme Azure AI Gallery pour développer et évaluer des modèles prédictifs capables de prévoir la durée de vie utile restante "RUL" des machines industrielles [58]. voici les colonnes de ce dataset :

- machineID - Identifiant de la machine
- datetime - Date de la télémétrie
- time_in_cycles - Lecture du cycle
- voltmean_24h - Moyenne de la télémétrie de tension des dernières 24 heures
- rotatemean_24h - Moyenne de la télémétrie de rotation des dernières 24 heures
- pressuremean_24h - Moyenne de la télémétrie de pression des dernières 24 heures
- vibrationmean_24h - Moyenne de la télémétrie de vibration des dernières 24 heures
- voltsd_24h - Écart-type de la télémétrie de tension des dernières 24 heures
- rotatesd_24h - Écart-type de la télémétrie de rotation des dernières 24 heures
- pressuresd_24h - Écart-type de la télémétrie de pression des dernières 24 heures
- vibrationsd_24h - Écart-type de la télémétrie de vibration des dernières 24 heures
- voltmean_5d - Moyenne de la télémétrie de tension des derniers 5 jours
- rotatemean_5d - Moyenne de la télémétrie de rotation des derniers 5 jours
- pressuremean_5d - Moyenne de la télémétrie de pression des derniers 5 jours
- vibrationmean_5d - Moyenne de la télémétrie de vibration des derniers 5 jours
- voltsd_5d - Écart-type de la télémétrie de tension des derniers 5 jours
- rotatesd_5d - Écart-type de la télémétrie de rotation des derniers 5 jours
- pressuresd_5d - Écart-type de la télémétrie de pression des derniers 5 jours
- vibrationsd_5d - Écart-type de la télémétrie de vibration des derniers 5 jours
- error1 - Identification d'erreur 1
- error2 - Identification d'erreur 2
- error3 - Identification d'erreur 3
- error4 - Identification d'erreur 4
- error5 - Identification d'erreur 5
- comp1 - Cycle de maintenance du composant 1
- comp2 - Cycle de maintenance du composant 2
- comp3 - Cycle de maintenance du composant 3
- comp4 - Cycle de maintenance du composant 4
- model - Modèle de la machine
- age - Âge de la machine
- RUWeek - Temps en semaines jusqu'à la défaillance
- failure - Identification du composant ayant échoué
- failed - Échoué ou non
- RUL - Durée de vie utile restante en cycles avant la défaillance
- RUL_I - Durée de vie utile restante inverse

machineID	datetime	time_in_cycles	voltmean_24h	rotatemean_24h	\
0	1 2015-01-06	2	171.025033	454.614348	
1	1 2015-01-07	3	174.139410	444.337772	
2	1 2015-01-08	4	172.135931	446.126781	
3	1 2015-01-09	5	169.350052	466.884090	
4	1 2015-01-10	6	171.006246	455.695551	

pressuremean_24h	vibrationmean_24h	voltsd_24h	rotatesd_24h	\
0	102.377665	41.506930	17.623759	43.648283
1	96.674842	41.702771	10.611608	39.450931
2	102.097039	39.281983	13.829868	38.445908
3	100.128260	38.571482	14.600352	40.676186
4	99.680231	39.789390	12.461183	33.019600

pressuresd_24h	vibrationsd_24h	voltmean_5d	rotatemean_5d	\
0	8.540959	4.298246	171.698774	452.843541
1	10.669844	5.808709	172.567505	452.344562
2	9.558219	4.133639	172.622242	449.728970
3	8.463613	4.632150	171.533767	453.357148
4	9.178725	5.427788	171.531334	453.531708

pressuremean_5d	vibrationmean_5d	voltsd_5d	rotatesd_5d	pressuresd_5d	\
0	100.529177	44.874649	1.879874	5.011519	2.141369
1	100.173224	45.360874	1.781676	5.818804	2.658025
2	100.992985	43.602453	1.758648	4.817313	2.442889
3	100.728184	40.878696	1.763873	8.948939	2.452399
4	100.191607	40.170511	1.764769	8.993078	2.293615

vibrationsd_5d	error1	error2	error3	error4	error5	comp1	comp2	\
0	5.217156	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	4.646503	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	5.012262	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	1.933709	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1.380712	0.0	0.0	0.0	1.0	0.0	0.0	0.0

comp3	comp4	model	age	DI	RULWeek	failure	failed	RUL	RUL_I	
0	0.0	0.0	model3	18	2.663840	1.0	none	0	1.0	59.0
1	0.0	0.0	model3	18	2.610489	1.0	none	0	2.0	58.0
2	0.0	0.0	model3	18	2.558206	1.0	none	0	3.0	57.0
3	0.0	0.0	model3	18	2.506971	1.0	none	0	4.0	56.0
4	0.0	0.0	model3	18	2.456761	1.0	none	0	5.0	55.0

Nombre de lignes : 28056
 Nombre de colonnes : 36

FIGURE III.7 – Le jeu de données de Télémétrie

— **Le jeux de données de filtre à gaz de "Hochschule Esslingen" :**

Ce deuxième dataset provient d'une étude réalisée sur un banc d'essai spécifique simulant les conditions de fonctionnement des filtres à gaz et les contraintes rencontrées dans des applications industrielles réelles, telles que la filtration de l'air dans des systèmes industriels [59]. Voici les colonnes de ce second dataset :

- Data_No : Numéro du cycle de test des défaillances du filtre
- Differential_pressure : Pression différentielle
- Flow_rate : Débit de gaz à travers le filtre
- Time : Durée écoulée depuis le début du cycle
- Dust_feed : Débit de poussière
- Dust : Type de poussière utilisée pour le test

Afin de faciliter la compréhension de ces différentes variables, il est important de le visualiser. La figure III.8 présente un aperçu de ces cinq premières lignes.

```
Data_No Differential_pressure Flow_rate Time Dust_feed \  
0      1                0.271267 52.281993 0.1 236.428943  
1      1                0.361690 54.986671 0.2 236.428943  
2      1                0.000000 55.424786 0.3 236.428943  
3      1                2.622251 55.524146 0.4 236.428943  
4      1                3.888165 55.852018 0.5 236.428943  
  
                                Dust  RUL  
0 ISO 12103-1, A3 Medium Test Dust 58.9  
1 ISO 12103-1, A3 Medium Test Dust 58.8  
2 ISO 12103-1, A3 Medium Test Dust 58.7  
3 ISO 12103-1, A3 Medium Test Dust 58.6  
4 ISO 12103-1, A3 Medium Test Dust 58.5  
Nombre de lignes : 39414  
Nombre de colonnes : 7
```

FIGURE III.8 – Le jeu de données Filtre

3.3 La solution proposée

Dans le chapitre précédent, nous avons constaté, au vue des études réalisées dans le cadre de la maintenance proactive, que les modèles predictifs basées sur des LSTM représentent une judicieuse solution pour le pronostic de défaillance. Il ont obtenue des meilleurs résultats notamment des taux d'erreurs faibles par rapport aux autres modèles d'apprentissage automatique.

Les LSTM, grâce à leur capacité à gérer les dépendances à long terme, modélisent efficacement les relations complexes entre les données séquentielles. Cette caractéristique est particulièrement avantageuse pour des applications telles que la maintenance prédictive PDM, où l'objectif est de prédire la durée de vie restante d'un équipement.

Notre solution proposée repose sur l'utilisation d'un RNN, plus précisément d'un réseau LSTM, pour capturer les dépendances dans les données séquentielles de mesures de surveillance. En analysant ces dépendances, notre modèle peut identifier les signes précurseurs de défaillances potentielles et estimer le temps restant avant qu'une maintenance soit nécessaire, permettant ainsi une gestion proactive et la réduction des temps d'arrêt imprévus.

3.3.1 Infrastructure de Développement et Ressources Techniques

Dans le cadre du déploiement de notre solution proposée, nous avons pris la décision de sélectionner les infrastructures de développement et les ressources techniques les plus adéquates.

■ Le langage de programmation utilisé :

Python est le langage de programmation utilisée pour le développement de notre modèle en raison de ses nombreuses fonctionnalités avantageuses et de sa large adoption dans la communauté scientifique et technologique .

Python est un langage de programmation interprété, de haut niveau et polyvalent, facile à lire et à écrire. Créé par Guido van Rossum et publié pour la première fois

en 1991, Python est devenu l'un des langages de programmation les plus populaires au monde [16].

Python est utilisé dans divers domaines tel que l'intelligence artificielle et l'analyse de données ainsi il est largement utilisé pour la prédiction, notamment dans le domaine de l'apprentissage automatique. Les bibliothèques comme Scikit-learn, TensorFlow et Keras, toutes écrites en Python, permettent de créer des modèles de prédiction pour résoudre divers problèmes tels que la classification, la régression et le clustering [16].

■ L'environnement d'exécution de la solution :

Nous avons développé notre solution d'apprentissage profond DL en Python, en exploitant l'environnement de notebook de Google Colab.

- **Google Colaboratory** : C'est une plateforme de développement et de collaboration basée sur le cloud, offerte par Google. Permet l'exécution de code Python dans un navigateur web sans installation supplémentaire. Accès gratuit à des ressources comme les GPU (Unité de traitement graphique) et les TPU (Unité de traitement de tenseurs), intégration avec Google.
- **Google Colaboratory notebook** : C'est un environnement interactif utilisant des notebooks Jupyter. Il permet aux utilisateurs d'exécuter du code, d'intégrer du texte et des visualisations, d'analyser les résultats et de partager leur travail [60].

■ Les bibliothèques utilisées

Pour la conception de notre solution, nous avons fait appel à un éventail de bibliothèques Python spécialisées, chacune apportant sa propre expertise et ses fonctionnalités distinctes.

- **NumPy** : Bibliothèque essentielle pour le calcul scientifique en Python, manipule des tableaux multi-dimensionnels et matrices, efficace pour les données volumineuses [16].
- **Pandas** : Bibliothèque dédiée la manipulation et l'analyse des données structurées en Python, utilisant des DataFrames pour nettoyer, analyser et visualiser diverses sources de données.[16]
- **Matplotlib** : Bibliothèque de traçage pour créer des visualisations statiques, animées et interactives en Python [16].
- **Seaborn** : Bibliothèque de visualisation de données basée sur Matplotlib, simplifie la création de visualisations statistiques attrayantes [61].
- **PyTorch** : Bibliothèque open-source d'apprentissage profond développée par Facebook, supporte le calcul sur GPU [62].
- **TensorFlow** : Bibliothèque open-source pour le calcul numérique et le machine learning, inclut Keras comme API haut niveau pour la création, la configuration et l'entraînement de réseaux neuronaux [16].
- **Scikit-learn** : Bibliothèque d'apprentissage automatique en Python, propose une gamme d'algorithmes et outils pour la modélisation [16].

3.3.2 Méthodologie

Pour la réalisation de notre modèle prédictif de gestion de maintenance industriel, nous avons suivi une démarche structurée reposant sur plusieurs étapes principales. La figure III.9 représente un schémas détaillé de cette méthodologie.

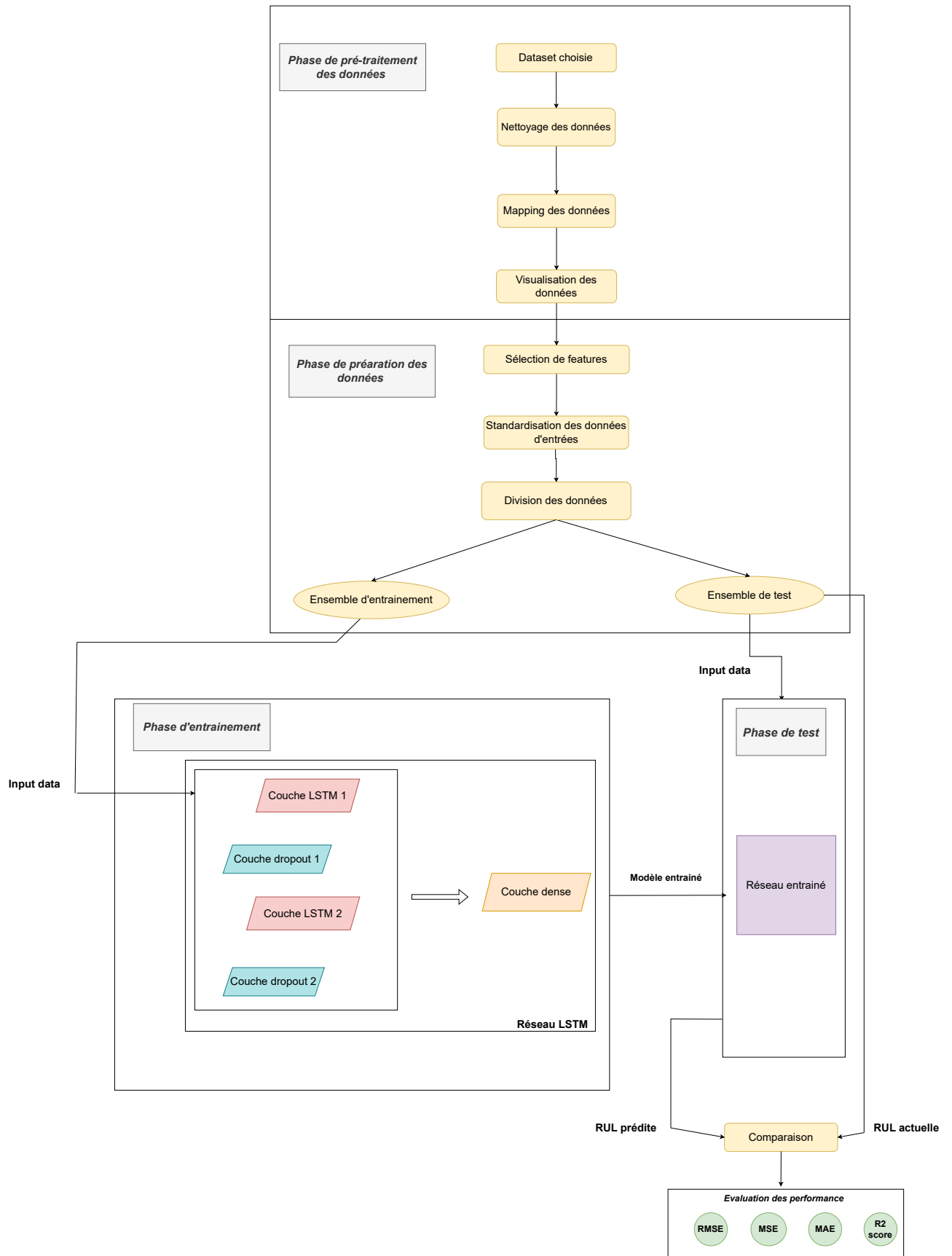


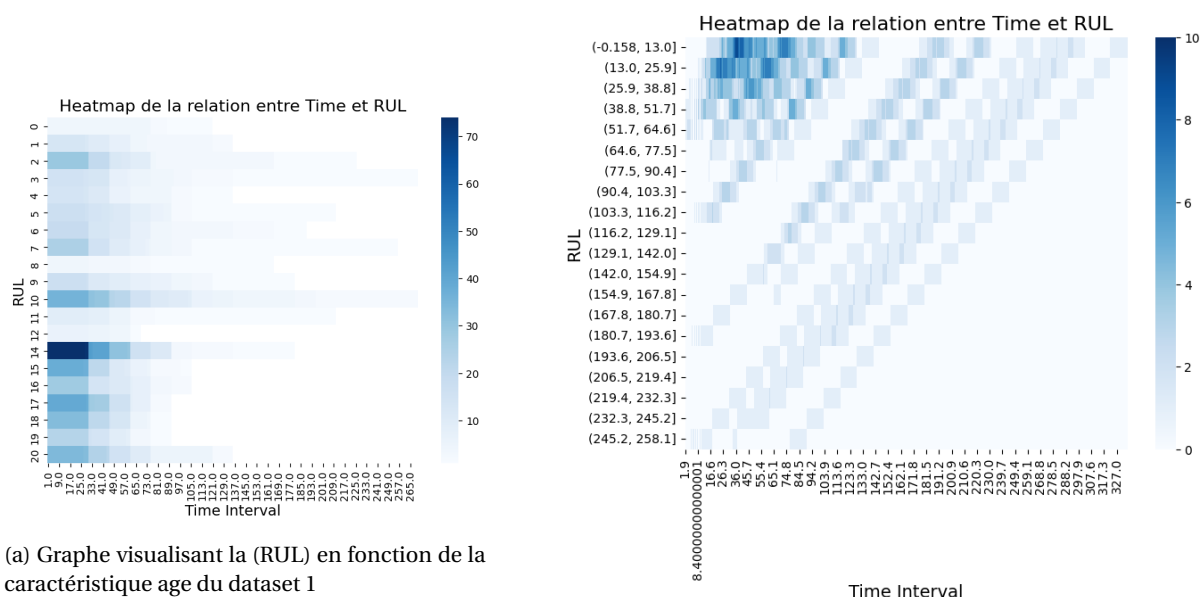
FIGURE III.9 – Méthodologie de la solution proposée

Voici une description détaillée de chaque composant de cette méthodologie :

A. Phase de pré-traitement des données :

Le pré-traitement des données est une étape cruciale pour préparer un jeu de données avant son utilisation dans un modèle d'apprentissage automatique ML.

- **Nettoyage des données :** Nous avons supprimé les lignes et colonnes contenant des valeurs nulles afin de garantir l'intégrité et la qualité des données pour les deux datasets.
- **Encodage des variables catégorielles (Mapping) :** Nous avons converti les variables catégorielles en variables numériques en utilisant des techniques comme le one-hot encoding ou le label encoding. Cela permet d'utiliser ces variables dans les modèles d'apprentissage automatique.
- **Visualisation des données :** Nous avons généré des graphiques pour visualiser les relations entre les différentes caractéristiques du jeu de données. Ces visualisations comprennent des histogrammes pour représenter la distribution des données, des scatter plots pour examiner les relations entre deux variables et des box plots pour comparer la répartition des données entre différentes catégories ou périodes.



(a) Graphe visualisant la (RUL) en fonction de la caractéristique age du dataset 1

(b) Graphe visualisant la (RUL) en fonction de la caractéristique Time du dataset 2

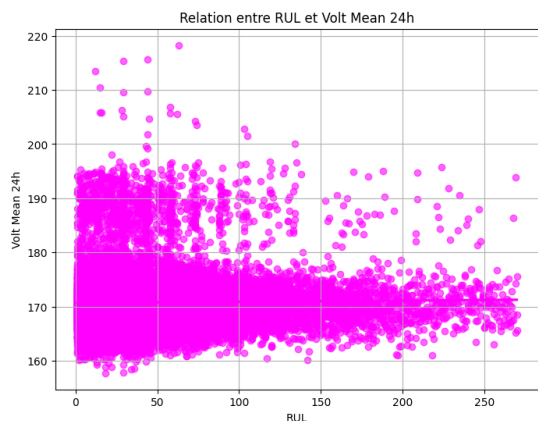
FIGURE III.10 – Visualisation de la relation entre la variable cible (RUL) et une caractéristique temporelle

La figure III.10 présente une visualisation de la variable cible RUL en fonction d'une caractéristique temporelle des datasets.

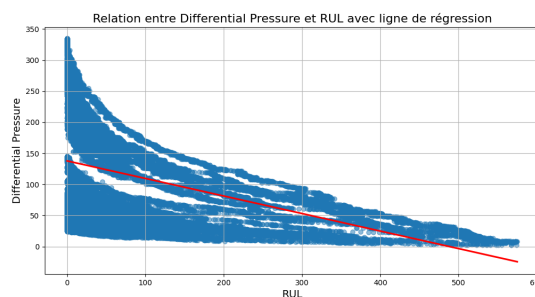
(a) présente un heatmap démontrant que la RUL d'un équipement diminue avec son âge. Un équipement neuf a une RUL de valeur supérieur, et cette dernière décroît progressivement avec le temps.

Le heatmap (b) illustre l'évolution de la RUL en fonction du temps *Time* à tra-

vers différents cycles *Data_No*. Il montre que la RUL est beaucoup plus élevée au début de la période de temps considérée et diminue au fil du temps.



(a) Graphe visualisant la relation entre la moyenne de tension en volt en fonction de la (RUL) du dataset 1



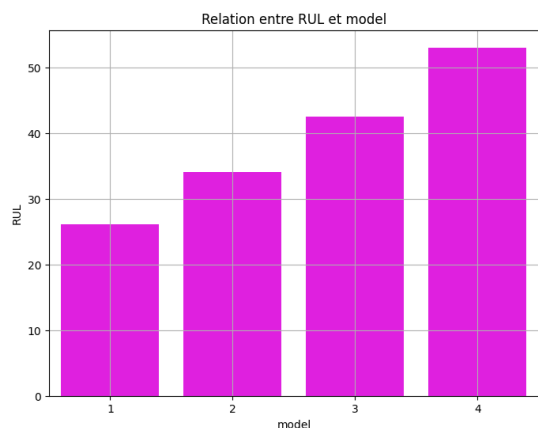
(b) Graphe visualisant la relation entre la pression différentiel et la RUL du dataset 2

FIGURE III.11 – Visualisation de la relation entre la target (RUL) en fonction d'une contrainte mécanique

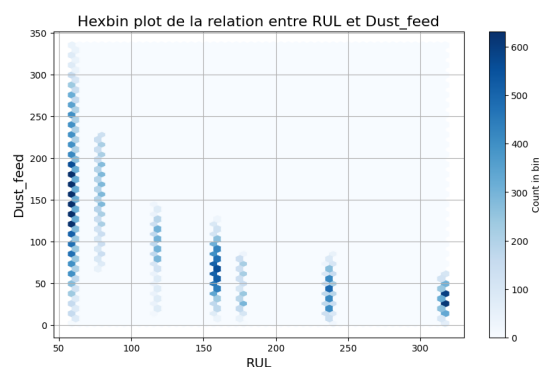
Dans (a) de la figure III.11, on observe un nuage de points illustrant la relation entre la RUL et la tension moyenne sur 24 heures *VoltMean24h* du premier dataset.

Les valeurs RUL faibles correspondent à des tensions moyennes concentrées entre 170 et 190 volts, suggérant que la tension moyenne tend à être dans cette plage en fin de vie de l'équipement. À mesure que la RUL augmente, les valeurs de tension deviennent plus dispersées, allant de 170 à environ 220 volts, indiquant que d'autres facteurs influencent la tension moyenne pour des RUL plus longues.

Le graphique (b) montre la relation entre la pression différentielle et la RUL, les points bleus indiquent les mesures réelles : une pression faible correspond à une RUL élevée (équipement en bon état), tandis qu'une pression élevée correspond à une RUL faible (équipement proche de la fin de vie).



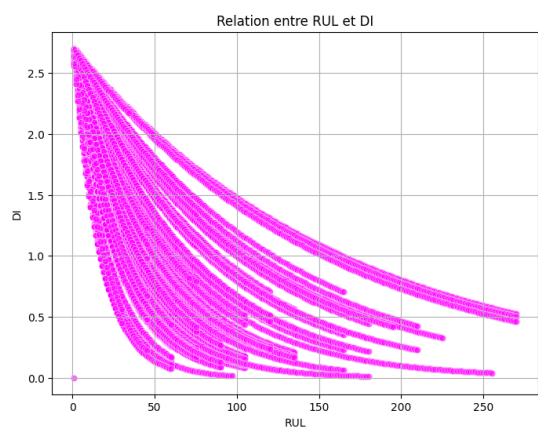
(a) Graphe visualisant la (RUL) en fonction du model du dataset 1



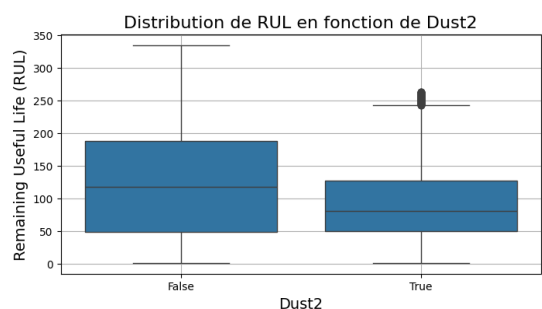
(b) Graphe visualisant la (RUL) en fonction du débit de poussière du dataset 2

FIGURE III.12 – Visualisation de la relation entre la (RUL) en fonction d’une caractéristique de grandeur physique des datasets

La figure III.12 présente une visualisation de la target RUL en fonction d’une caractéristique de grandeur physique des datasets. Nous observons dans (a) un diagramme en barres comparant la RUL d’un équipement par quatre modèles différents numérotés de 1 à 4, avec les modèle 3 et 4 on remarque une RUL plus longue que les modèles 1 et 2. Dans (b) on trouve un graphique hexbin montrant la relation entre la RUL d’un filtre et le débit de poussière (Dust_feed). On observe que de faibles valeurs de la RUL correspondent souvent à des débits de poussière plus élevés, indiquant la détérioration du filtre.



(a) Graphe visualisant la (RUL) en fonction de l’indice de dégradation (DI) du dataset 1



(b) Graphe visualisant la (RUL) en fonction de Dust 2 du dataset 2

FIGURE III.13 – Visualisation de la relation entre la (RUL) en fonction d’une caractéristique de dégradation

La figure III.13 présente une visualisation de la target RUL en fonction d’une caractéristique de dégradation dans les datasets. Nous observons dans (a) un nuage de point qui montre la relation entre la RUL et l’indice de dégradation (DI) d’un équipement, dans ce graphe lorsque le RUL est élevé le DI est faible, à mesure que le RUL diminue le DI augmente. Ce graphique illustre que le DI

peut servir d'indicateur fiable pour estimer le RUL et planifier la maintenance prédictive PDM. Dans (b) nous avons un graphe a box montrant l'impact de la poussière de diamètre moyen (Dust2) sur la RUL d'un équipement. Sans poussière (Dust2 = False), la RUL moyenne est supérieur. Avec poussière (Dust2 = True), la RUL moyenne diminue, cela indique que la présence de poussière réduit significativement la durée de vie de l'équipement.

■ Analyse de la corrélation :

La corrélation est une mesure de la liaison entre les variables, indiquant leur degré de dépendance. Ce permettant d'évaluer l'ajustement linéaire d'une variable par rapport à une autre [17].

Pour quantifier cette corrélation, on calcule un coefficient de corrélation linéaire, mesurant la force et la direction de la relation linéaire entre deux variables. Le coefficient le plus couramment utilisé est le coefficient de corrélation de Pearson, noté r . Il est calculé comme suit :

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

où $\text{Cov}(X, Y)$ représente la covariance entre les variables X et Y , et σ_X et σ_Y sont les écarts types respectifs.

La valeur du coefficient de corrélation linéaire varie entre -1 et +1 :

- +1 : forte relation positive.
- -1 : forte relation négative.
- 0 : aucune corrélation.

Nous avons utilisé des matrices de corrélation pour analyser les relations linéaires entre les caractéristiques de nos différents ensembles de données. Chaque cellule de ces matrices est colorée en fonction du coefficient de corrélation entre la paire de caractéristiques correspondante. Cette méthode nous a permis d'identifier les relations significatives entre les variables et de sélectionner les caractéristiques les plus pertinentes.

Chapitre 3 : Solution proposée

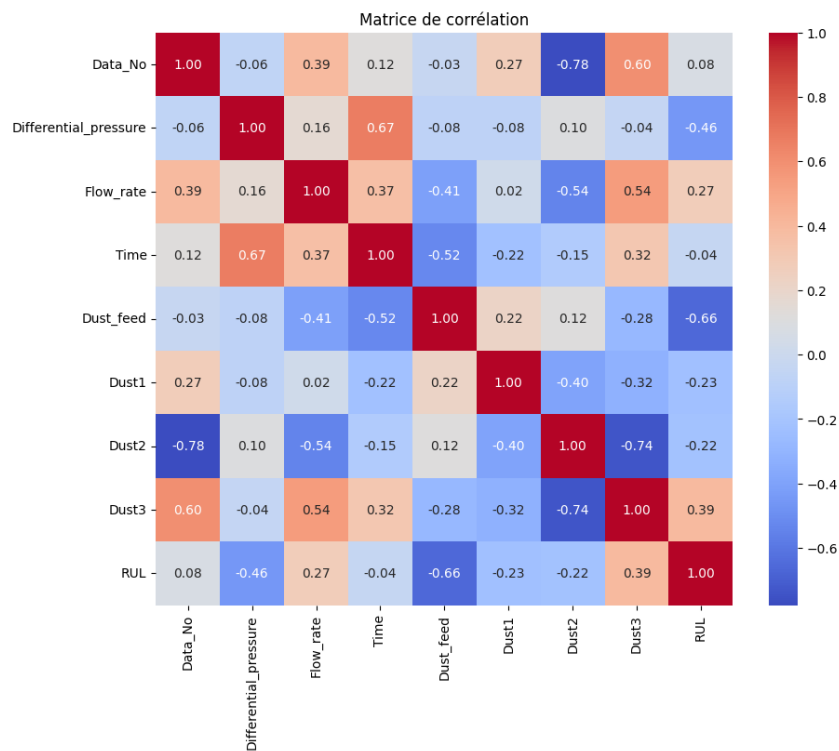
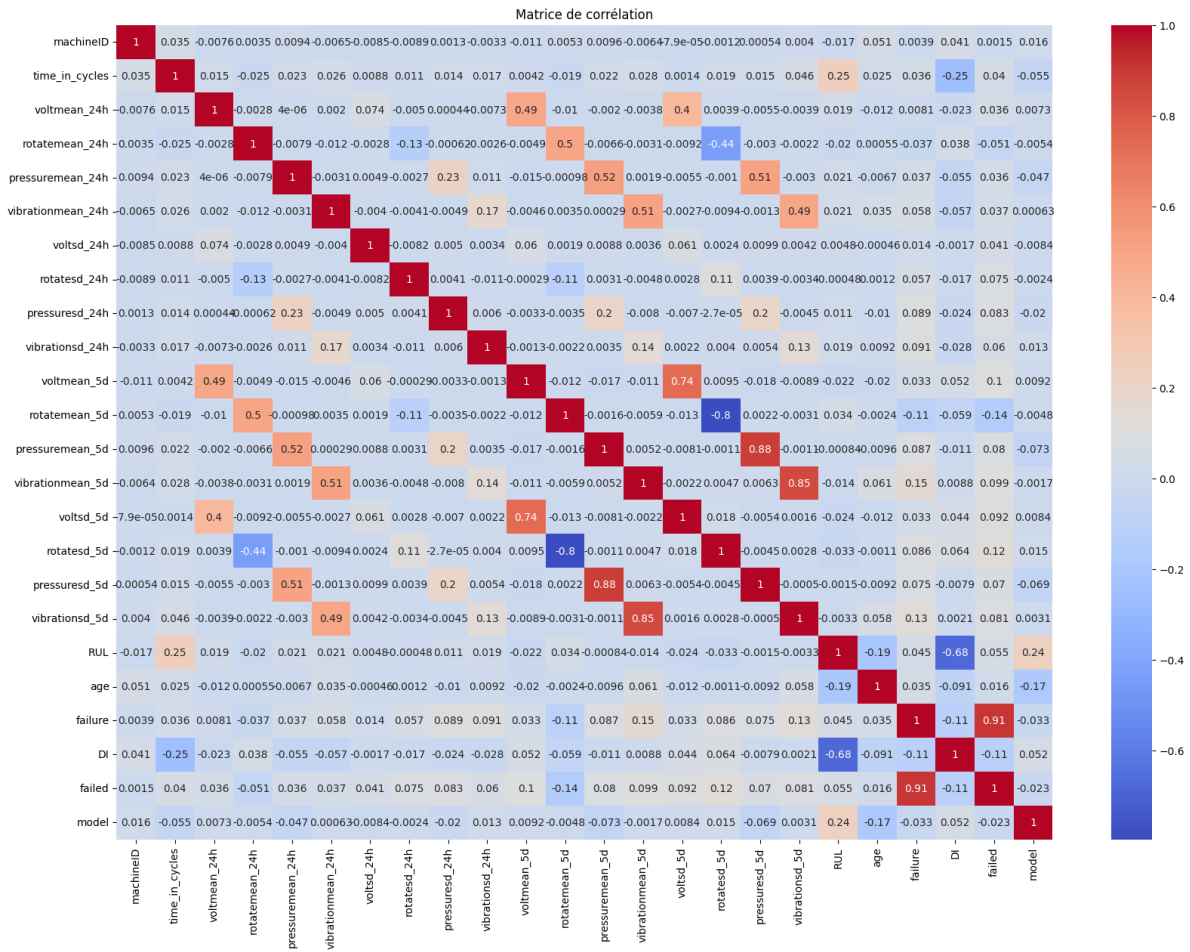


FIGURE III.14 – Visualisation des matrices de corrélation des datasets

B. Phase de préparation des Données :

L'étape de préparation des données est crucial pour les modèles d'apprentissage automatique ML. Voici ci-dessus les étapes de la préparation des données pour l'entraînement :

- **Séparation des Caractéristiques et de la Cible** : Les caractéristiques représentant toutes les colonnes à l'exception de 'RUL', sont extraites du DataFrame principal et stockées dans X. La variable cible RUL est extraite et stockée dans une variable y.
- **Division des Données** : La prochaine étape consiste à diviser les données en ensembles d'entraînement et de test. Cette division est cruciale pour évaluer la performance des modèles sur des données non vues pendant l'entraînement. Nous avons divisé les données en deux ensembles : 70% des données sont utilisées pour l'entraînement et les 30% restants sont utilisés pour le test, une graine aléatoire de 42 a été définie pour garantir la reproductibilité des résultats.
- **Standardisation des Features** : Les données d'entrée sont standardisées, ce qui signifie qu'elles sont mises à l'échelle pour avoir une moyenne de 0 et un écart-type de 1. Cela aide à améliorer les performances et la stabilité du modèle.

$$z = \frac{x - \mu}{\sigma}$$

Où : le z est la valeur standardisée

Le x est la valeur initiale

Le μ est la moyenne de l'ensemble des données

Et σ est l'écart-type de l'ensemble des données.

- **Transformation en tenseur pytorch** : Les données d'entraînement et de test sont converties en tenseurs PyTorch. une dimension supplémentaire est ajoutée à chaque tenseur d'entrée à l'index 1, représentant le nombre de séquences dans chaque batch.

C. Phase d'entraînement du modèle :

Dans la phase d'entraînement, nous avons construit notre modèle comprenant deux couches LSTM suivie chacune d'une couche de dropout, pour éviter le sur-apprentissage et accroître la robustesse du modèle. Ensuite, ce dernier est entraîné pour apprendre les tendances chronologiques du jeu de données d'entraînement. voici les étapes de la phase d'entraînement :

- **Définition de l'architecture du modèle LSTM** : Dans le modèle LSTM décrit, les tenseurs de l'ensemble d'entraînement et de test sont structurés pour être compatibles avec les exigences spécifiques de chaque couche du réseau. Initialement, ils représentent les séquences de données d'entraînement et de test, où chaque séquence est unidimensionnelle et contient le nombre de caractéristique du jeu de données. Lors du passage à travers le réseau LSTM, chaque séquence est traitée en batch.

— Première couche LSTM : Cette première couche prend en entrée les séquences

de données par pas de temps. Elle opère séquentiellement sur chaque séquence dans le batch, capturant ainsi les dépendances temporelles à long terme au sein de chaque séquence. Pendant l'entraînement, les états cachés et de cellules sont mis à jour à chaque pas de temps, et produit une sortie de dimension de la couche cachés. Ce qui permet au modèle d'apprendre des représentations séquentielles complexes.

- Deuxième couche LSTM : Cette seconde couche LSTM est connectée à la sortie de la première couche. Elle est chargée de capturer les dépendances temporelles supplémentaires et d'apprendre des caractéristiques plus abstraites des séquences temporelles. Elle opère de manière similaire à la première couche, traitant les séquences en batch avec une taille spécifiée.
- Couches dropout : À chaque couche LSTM, une couche de régularisation (dropout) est appliquée. Cette couche ajoute du bruit au réseau en désactivant aléatoirement un pourcentage (taux de dropout) de neurones pendant l'apprentissage. Cela aide à prévenir le sur-apprentissage en forçant le modèle à apprendre des représentations plus robustes des données séquentielles.
- Couches entièrement connectées (Dense) : Elle est utilisée après la dernière couche LSTM pour transformer la sortie séquentielle en une prédiction finale.

■ Configuration et sélection des hyperparametres :

La détermination des hyper-paramètres joue un rôle crucial dans l'optimisation des performances du modèle. Afin de trouver la combinaison optimale, qui minimise la perte d'évaluation et maximise la précision des prédictions, nous avons effectué une recherche en grille en utilisant la classe "GridSearchCV".

Le modèle LSTM est encapsulé dans une classe "LSTMModel" qui le rend compatible avec "GridSearchCV" qui inclut des méthodes pour définir et ajuster les paramètres, entraîner le modèle et faire des prédictions. Cette fonction teste plusieurs hyper-paramètres selon une grille de recherche prédéterminée, incluant plusieurs valeurs possibles pour les tailles des couches cachées, le taux de dropout et le taux d'apprentissage.

Paramètre	Dataset 1	Dataset 2
Première couche LSTM	32, 64, 128	64, 128
Deuxième couche LSTM	32, 64	32, 64
Taux de dropout	0.1, 0.2	0.1,0.2
Taux d'apprentissage	0.001, 0.0005	0.01, 0.1

TABLE III.1 – Grilles de recherche des hyperparamètres pour les deux datasets

La figure III.1 présente la grille de recherche des hyperparamètres pour chacun de nos jeux de données, utilisant "GridSearchCV" pour définir la taille des couches caché des couches LSTM ainsi que les taux de dropout et d'apprentissage.

Pour chaque combinaison d'hyperparamètres, le modèle est entraîné sur un

sous-ensemble des données d'entraînement via la validation croisée avec $k = 2$ et ses performances sont évaluées à l'aide de la fonction de coût MSE. Cette fonction permet de calculer l'erreur entre les prédictions du modèle et les valeurs cibles. L'optimiseur Adam est utilisé, avec un taux d'apprentissage et un paramètre de régularisation pour ajuster les poids du modèle de manière à minimiser cette perte.

Nous avons utilisé pour les couches LSTM les fonctions d'activation Sigmoidale et Tangente hyperbolique ou la fonction Sigmoidale est utilisée pour les portes d'oubli, d'entrée et de sortie, afin de réguler précisément le flux d'information, garantissant que les informations importantes sont retenues et que les informations non pertinentes sont oubliées. Et la fonction Tangente hyperbolique est utilisée pour la mise à jour de l'état de la cellule, transformant les entrées et l'état caché précédent, produisant des valeurs dans la plage de -1 à 1. Pour la couche entièrement connectée on l'a configurée comme une transformation linéaire sans fonction d'activation supplémentaire définie. Le modèle itère sur un nombre spécifié d'epochs pour chaque ensemble de données, calculant la perte d'entraînement à chaque itération grâce à la rétropropagation et optimisant les poids du modèle avec les gradients calculés. Les pertes d'entraînement et de test sont enregistrées tous les 10 epochs pour évaluer la performance du modèle sur des données d'entraînement et de test distinctes, et sont imprimées périodiquement pour suivre la progression de l'entraînement.

Après avoir testé toutes les combinaisons possibles dans la grille, "GridSearchCV" sélectionne la configuration qui minimise le MSE sur des données de validation croisée, et les meilleurs hyper-paramètres sont déterminés pour améliorer les prédictions.

Pour optimiser le processus de recherche d'hyper-paramètres dans le dataset 1, un sous-ensemble des données d'entraînement est échantillonné. Ce sous-ensemble représente 10% des données d'entraînement complètes et utilisé pour accélérer la recherche d'hyperparamètres, rendant le processus plus rapide et plus efficace. Tandis que pour le dataset 2 on a utilisé le jeu de données complet.

- **Sélections des meilleurs hyperparamètres :** Une nouvelle instance du modèle LSTM est créée avec les meilleurs hyperparamètres trouvés par GridSearchCV tel montré dans la figure III.2 qui représente les meilleurs hyperparamètres trouvés pour chacun de nos jeux de données minimisant au plus la MSE. Ce modèle est ensuite prêt pour l'entraînement sur l'ensemble des données d'entraînement.

Paramètre	Dataset 1	Dataset 2
Taille de la première couche LSTM	64	64
Taille de la Deuxième couche LSTM	64	64
Taux de dropout	0.1	0.1
Taux d'apprentissage	0.001	0.01
Nombre d'epochs	300	150
Taille du batch	32	32

TABLE III.2 – Les meilleurs hyperparamètres trouvés pour les deux datasets

Le tableau III.2 présente les hyper-paramètres optimaux utilisés pour entraîner le modèles LSTM sur Les deux datasets distincts. Chaque dataset est caractérisé par des configurations spécifiques : les réseaux LSTM comportent une première et une deuxième couche de 64 unités chacune, avec un taux de dropout de 0.1 pour régulariser l'apprentissage et un taux d'apprentissage de 0.001 pour le dataset 1 et de 0.01 pour le dataset 2. L'entraînement a été effectué sur 300 epochs pour le premier dataset et sur 150 epochs pour le second, avec une taille de batch 32 et une longueur de séquence d'une étape temporelle. Ces choix d'hyper-paramètres sont cruciaux pour obtenir des performances optimales dans la modélisation des données séquentielles, en ajustant la complexité du modèle et les paramètres d'entraînement pour chaque ensemble de données spécifique.

D. Phase de test et d'évaluation des performances du modèle :

La phase de test consiste à prendre l'ensemble de test et à appliquer le réseau entraîné avec l'ensemble d'entraînement. Cette phase permet d'évaluer la performance du modèle sur des données qu'il n'a jamais vue auparavant, afin de vérifier sa capacité de généralisation.

Pendant l'évaluation, le modèle désactive les couches dropout, les pertes moyennes sur les batchs de l'ensemble d'entraînement et de test sont calculée et enregistrée à intervalles réguliers (toutes les 10 époques) et tracées au fil des époques pour visualiser la convergence du modèle. Cela permet de voir si le modèle s'améliore et s'il y a un sur-apprentissage (lorsque la perte de test augmente alors que la perte d'entraînement diminue). Les prédictions du modèle sont comparées aux valeurs réelles pour calculer des métriques de performance telles que RMSE, MSE, MAE et R² score. Ces métriques fournissent une évaluation quantitative de la précision du modèle.

4 Conclusion

En conclusion, l'approche présentée dans ce chapitre se veut être un levier essentiel pour améliorer la fiabilité et la performance des processus industriels en Algérie. Malgré les défis posés par les contraintes des données fournies par l'entreprise Cevi-

tal, notre persévérance et notre capacité d'adaptation nous ont permis de développer un modèle LSTM efficace pour anticiper les défaillances imprévues, en s'appuyant sur l'exploitation de données plus appropriées issues de Kaggle.

Cette étude ne souligne pas seulement l'efficacité de la méthode en termes d'amélioration des processus industriels en Algérie, mais ouvrent également la voie à des applications transversales dans divers secteurs d'activité. Au-delà de cette application industrielle, les techniques développées dans ce projet pourraient trouver des applications dans d'autres domaines confrontés à des problématiques similaires de prévision de défaillances d'équipements.

Par conséquent, cette recherche pave la route vers une généralisation de pratiques de maintenance plus préventives et économiquement avantageuses, marquant un pas vers l'excellence opérationnelle et la durabilité des équipements.

Dans le chapitre suivant, nous présenterons les résultats obtenus par notre solution, effectuerons une analyse comparative.

Chapitre IV

Validation et expérimentations

1 Introduction

Ce chapitre est dédié à l'évaluation de la performance des modèles utilisés pour prédire les résultats dans le cadre de notre étude sur la maintenance prédictive. Nous introduirons tout d'abord les différentes métriques d'évaluation utilisées pour mesurer la précision de nos résultats. Ensuite, nous présenterons les résultats obtenus à partir de nos jeux de données en utilisant des visualisations appropriées. Ces visualisations permettront de mettre en lumière les performances de nos modèles sous différents angles, facilitant ainsi la compréhension des forces et des limites de chaque approche de prédiction. Enfin, nous procéderons à une étude comparative entre nos résultats et ceux obtenus dans l'état de l'art. Cette comparaison nous permettra d'évaluer la compétitivité de nos modèles par rapport aux meilleures pratiques actuelles en matière de maintenance prédictive, en identifiant les points forts et les domaines où des améliorations peuvent être apportées.

2 Evaluation des performances

L'évaluation des performances d'un modèle ML joue un rôle essentiel dans l'appréciation de sa qualité et dans la garantie de sa fiabilité. Elle permet de tester l'efficacité du modèle en termes de capacité à fournir des prédictions précises. Par conséquent, après avoir réalisé des prédictions de la RUL des équipements industriels, il est essentiel de procéder à une évaluation approfondie afin de mesurer les performances de l'approche choisie.

2.1 Les métriques d'évaluation

Les résultats du modèle ont été évalués en analysant des critères tels que le MSE (Mean Squared Error), le RMSE (Root Mean Squared Error), le MAE (Mean Absolute Error) et le R2-score [17].

- Erreur quadratique moyenne (MSE) : Le MSE est une mesure couramment utilisée pour évaluer la performance d'un modèle de régression. Il est calculé en prenant la différence entre la valeur prédite et la valeur réelle pour chaque point de données, en élevant cette différence au carré, puis en faisant la moyenne de ces carrés pour obtenir le MSE.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

où n est le nombre total de points de données.

Des valeurs plus basses de MSE indiquent une meilleure adéquation du modèle aux données réelles, c'est-à-dire que les prédictions du modèle sont plus proches des valeurs réelles.

- R2 score : Le score R^2 (ou coefficient de détermination) est utilisé pour évaluer la performance d'un modèle de régression linéaire. Il mesure la proportion de varia-

tion dans la variable de sortie qui est prédite à partir des variables d'entrée.

Il est calculé par l'expression mathématique suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

où :

- y_i représente les valeurs observées,
- \hat{y}_i représente les valeurs prédites par le modèle,
- \bar{y} est la moyenne des valeurs observées,
- n est le nombre de points de données.

cette mesure varie entre $-\infty$ et 1, où 1 indique un ajustement parfait et $-\infty$ indique un modèle constant.

2.2 Résultats d'évaluation pour le dataset 1

Les résultats des différentes métriques d'évaluations sont présentés dans le tableau ci-dessous IV.1 pour les ensembles de test du dataset 1. Cela fournit une vue d'ensemble complète des capacités prédictives du modèle et permet de juger son efficacité et sa robustesse.

Métrique	Dataset 1
MSE	13.5039
RMSE	3.6748
MAE	2.1357
R2-score	0.9913

TABLE IV.1 – Résultats des évaluations du dataset 1

- L'erreur quadratique moyenne : on observe une valeur de MSE à 13.5039 qui est relativement élevée. Une légère élévation du MSE par rapport aux autres valeurs s'explique principalement par le fait que la plupart des valeurs de la RUL sont inférieurs à 100 comme démontré dans la figure IV.1. Dans une plage plus restreinte de la RUL, même de petites erreurs peuvent significativement influencer le MSE, étant donné leurs proportion élevée par rapport à l'étendue totale des valeurs. Ainsi, une augmentation modeste du MSE dans ce contexte indique une précision moins satisfaisante des prédictions par rapport aux véritables valeurs de RUL.

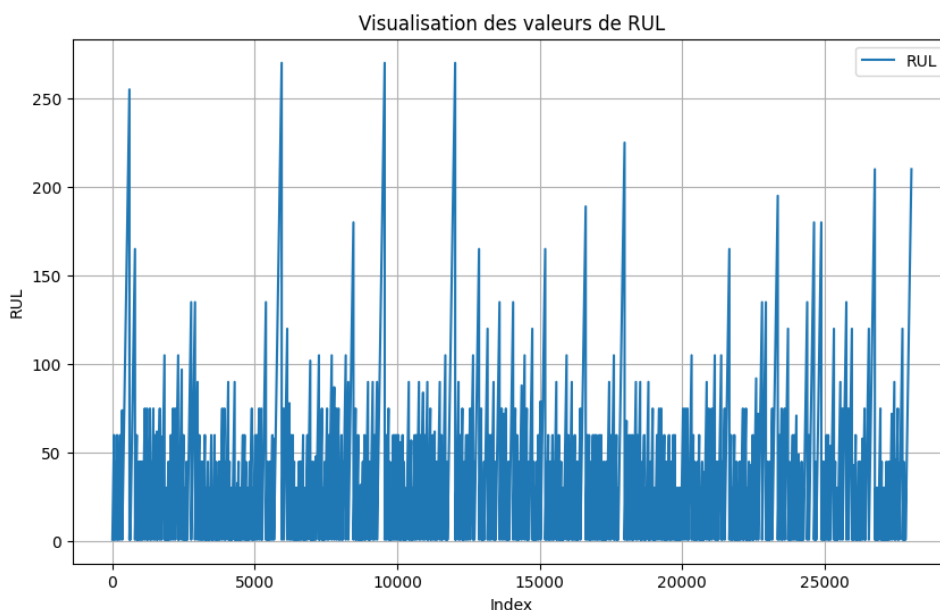


FIGURE IV.1 – Graphe visualisant la target (RUL) du dataset 1

- La racine de l’erreur quadratique moyenne : La RMSE montre une faible valeur de 3.6748, ce qui signifie que les prédictions ont en moyenne une erreur de cette faible amplitude.
- Erreur absolue moyenne : La valeur MAE est relativement minimale de 2.1357, indiquant une précision raisonnable des prédictions en termes de magnitude absolue des erreurs.
- R2-score : On constate une valeur de R2-score élevé de 0.9913 indique que le modèle explique 99.13% de la variance des données de RUL. Cela montre une capacité élevée du modèle à expliquer les variations des données.

Visualisations des résultats pour le dataset 1

Afin de mieux comprendre le comportement du modèle, nous avons d’abord tracés un graphique pour visualiser l’évolution des pertes d’entraînement et de test au fil des époques. Cela nous a permis d’identifier si le modèle souffre de sur-apprentissage (overfitting) ou de sous-apprentissage (underfitting) en observant les tendances des courbes de perte. Ensuite, nous avons affiché un graphique qui compare les valeurs réelles de la variable cible RUL avec les valeurs prédites par le modèle sur l’ensemble de test.

Cette comparaison visuelle nous a permis d’évaluer directement la précision des prédictions du modèle et de repérer d’éventuels écarts entre les valeurs réelles et prédites.

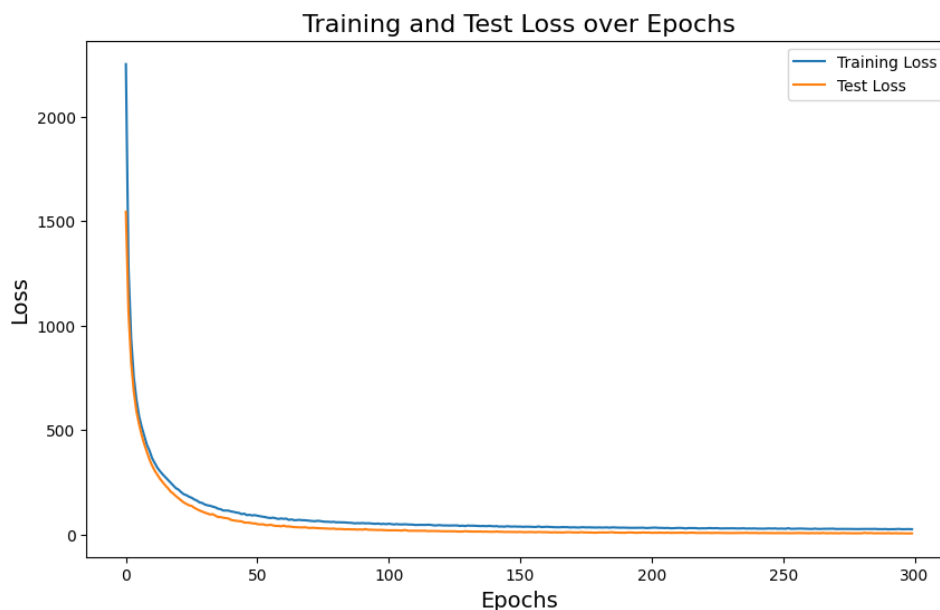
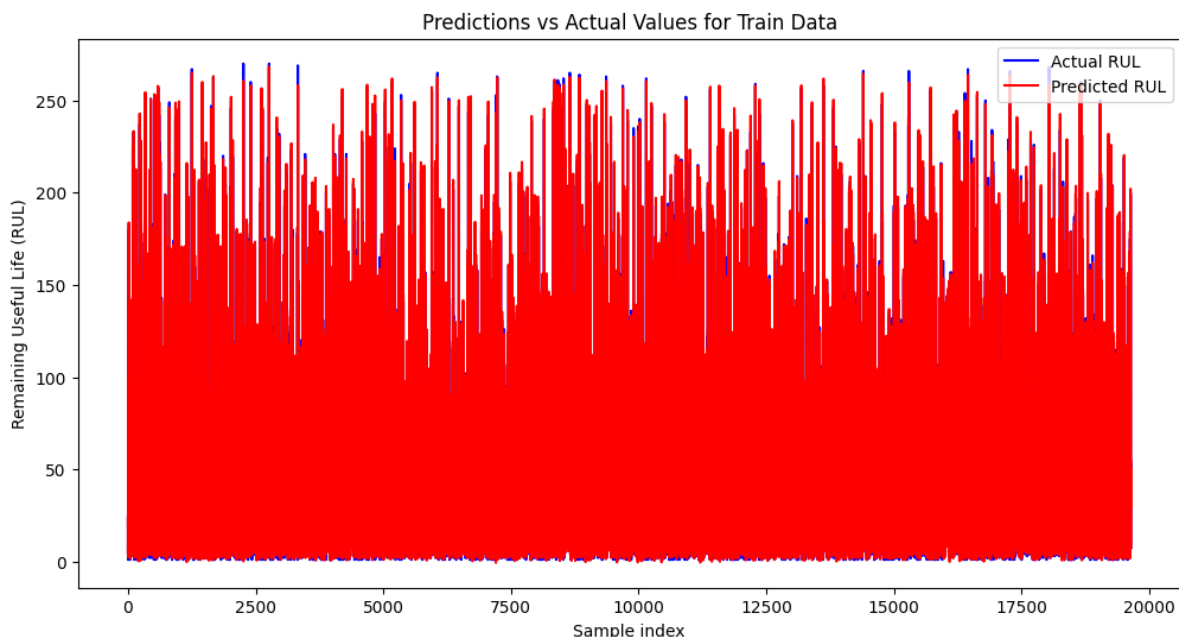


FIGURE IV.2 – Les courbes des pertes d’entraînement et de test au fil des époques pour le dataset 1

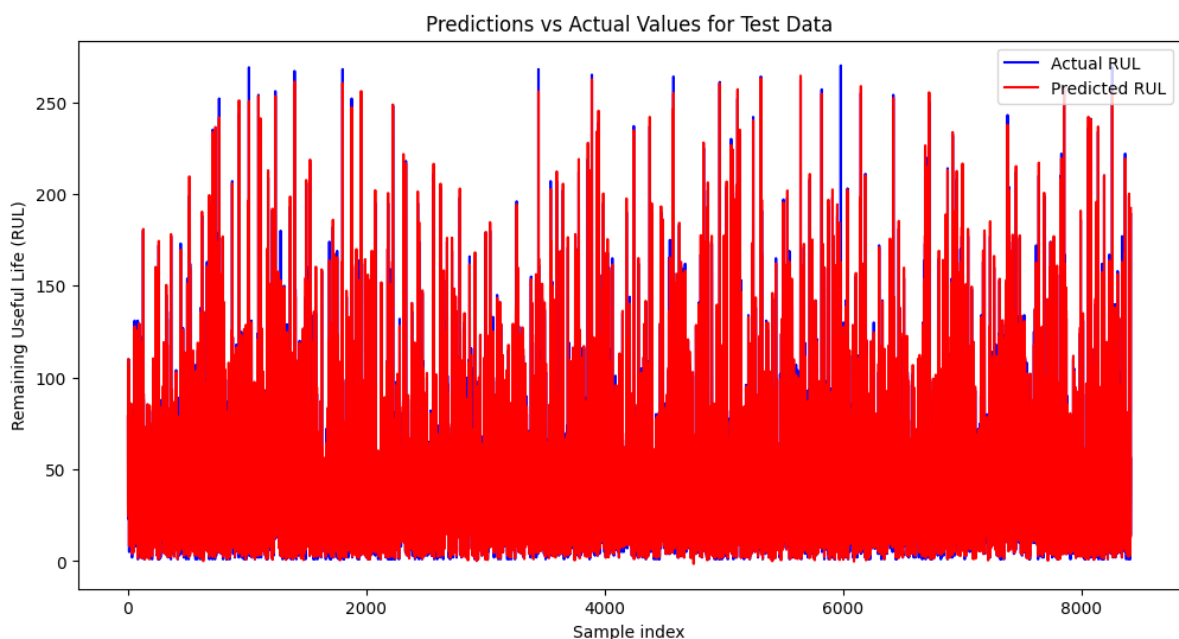
Le graphique IV.2 illustre l’évolution de la perte du modèle, sur les données d’entraînement (en bleu) et les données de test (en rouge), durant son apprentissage pour la prédiction de la RUL au fil des époques.

On observe une convergence rapide des deux courbes de perte dès les premières époques, indiquant une phase d’apprentissage initiale efficace. Après cette phase initiale, les courbes se stabilisent, suggérant que le modèle atteint un point d’équilibre dans son apprentissage.

De plus, on remarque que les courbes de perte d’entraînement et de test maintiennent une forte proximité tout au long de l’apprentissage. Cela confirme l’efficacité des techniques de régularisation employées, telles que la validation croisée et le dropout, pour prévenir le sur-apprentissage et garantir une bonne capacité de généralisation du modèle aux données non vues.



(a) Graphe visualisant la précision des prédictions pour l'ensemble d'entraînement



(b) Graphe visualisant la précision des prédictions pour l'ensemble de test

FIGURE IV.3 – Visualisation de la précision des prédictions pour le dataset 1

La figure IV.3 offrent une comparaison visuelle entre les valeurs prédites (en rouge) et les valeurs réelles (en bleu) de la RUL pour l'ensemble d'entraînement et pour l'ensemble de test.

Dans les deux cas, on observe que les prédictions suivent la tendance des valeurs réelles, avec des pics et des creux correspondants. Cependant, des écarts sont visibles, notamment pour certaines valeurs élevées où le modèle semble parfois sous-estimer la RUL. On note une densité plus élevée de prédictions dans les valeurs basses de RUL, indi-

quant que le modèle pourrait être plus précis pour les équipements proches de la fin de leur vie utile.

On peut remarqué également que la performance sur l'ensemble de test est légèrement moins précise que sur l'ensemble d'entraînement, ce qui est attendu.

Mais la différence n'est pas significative, comme démontré par l'histogramme des erreurs de prédiction IV.4. Cet histogramme montre que la distribution des erreurs de prédiction effectuées par le modèle est symétrique (positives et négatives) et centrée sur 0. Cela suggère que le modèle est capable de faire des prédictions précises de la RUL pour la plupart des cas.

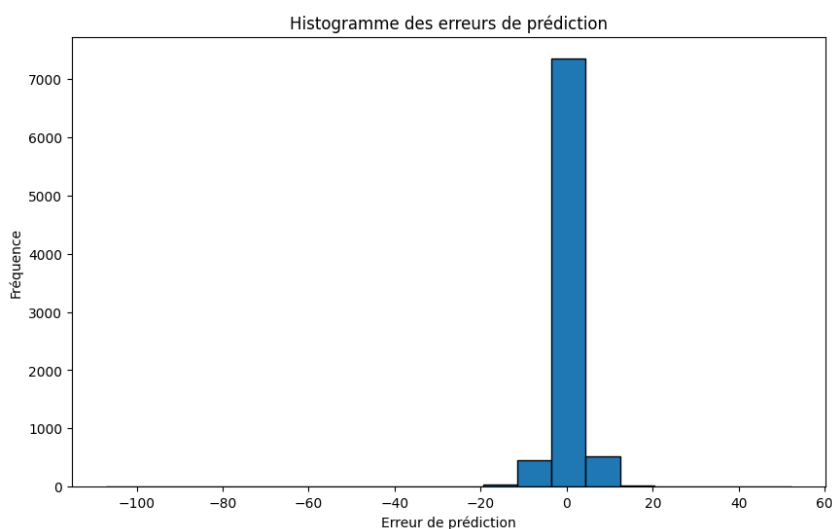


FIGURE IV.4 – L’histogramme visualisant la distribution des erreurs de prédiction pour le dataset 1

Le tableau IV.5 ci-dessous présente les valeurs réelles et prédites de la RUL pour l'ensemble de test.

	Real	Predicted
0	79.0	78.565155
1	30.0	30.316397
2	38.0	38.099907
3	23.0	24.510059
4	110.0	110.074425
...
8412	25.0	28.003302
8413	56.0	56.959015
8414	14.0	12.944644
8415	47.0	48.014771
8416	189.0	192.372192

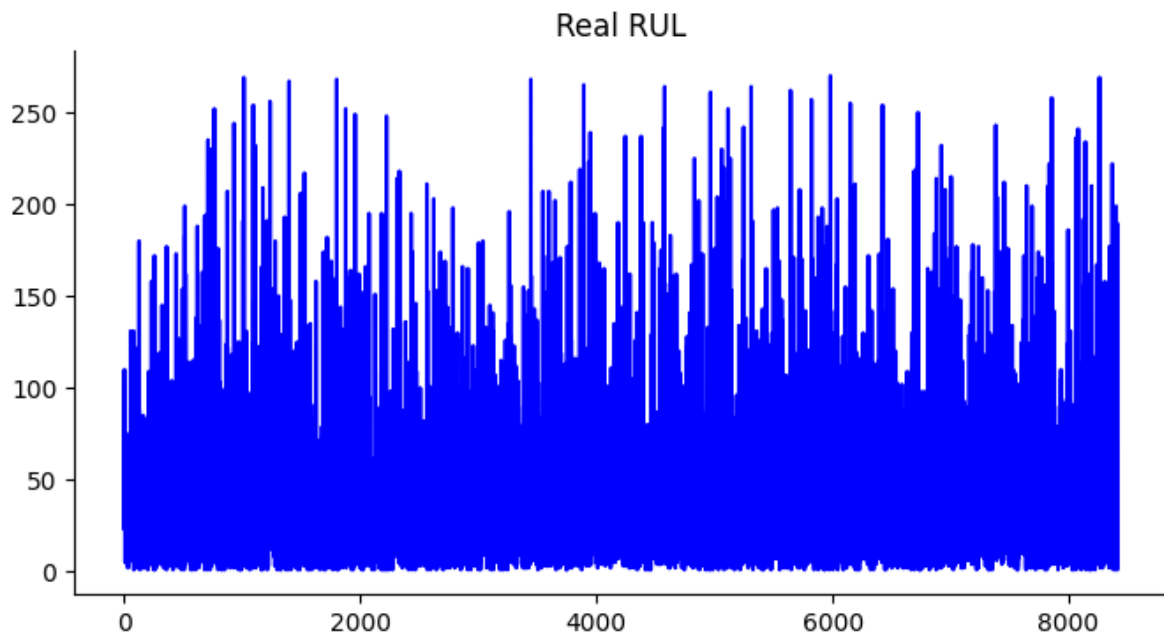
[8417 rows x 2 columns]

FIGURE IV.5 – Tableau des valeurs réelles et prédites de la RUL pour l'ensemble de test du dataset 1

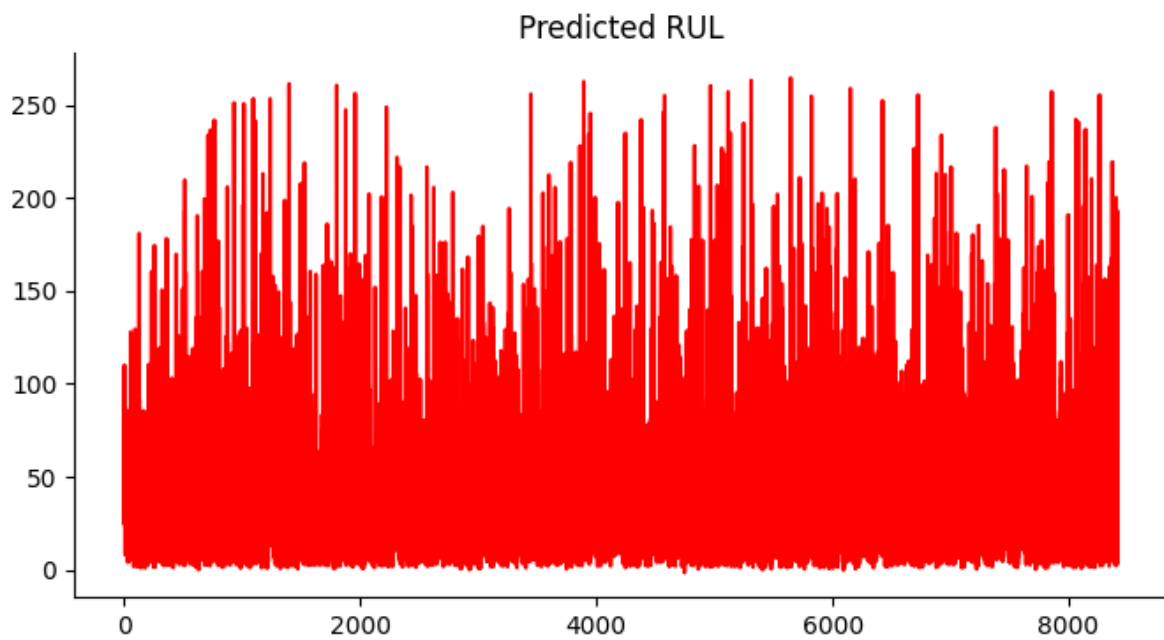
Afin de représenter graphiquement ces valeurs, les figures IV.6 illustrent une comparaison intéressante entre leur distributions.

Les deux graphiques présentent des motifs similaires, avec des valeurs allant principalement de 0 à environ 250. Cependant, on note quelques différences subtiles. Le graphique des valeurs réelles (a) semble avoir une distribution légèrement plus uniforme sur toute la plage, avec des pics plus prononcés dans les valeurs élevées. En revanche, le graphique des prédictions (b) montre une densité plus élevée dans les valeurs basses à moyennes de RUL, suggérant une tendance du modèle à prédire plus fréquemment des valeurs dans cette plage. Cette différence pourrait indiquer que le modèle est plus prudent dans ses estimations pour les équipements ayant une durée de vie restante plus longue.

Malgré ces nuances, la similarité globale des deux distributions témoigne d'une performance raisonnable du modèle dans la capture de la structure générale des données de RUL.



(a) Graphe visualisant les valeurs réelle de la (RUL) de l'ensemble de test



(b) Graphe visualisant les valeurs prédites de la (RUL) pour l'ensemble de test

FIGURE IV.6 – Visualisation des valeurs réelles et prédites de la (RUL) du dataset 1

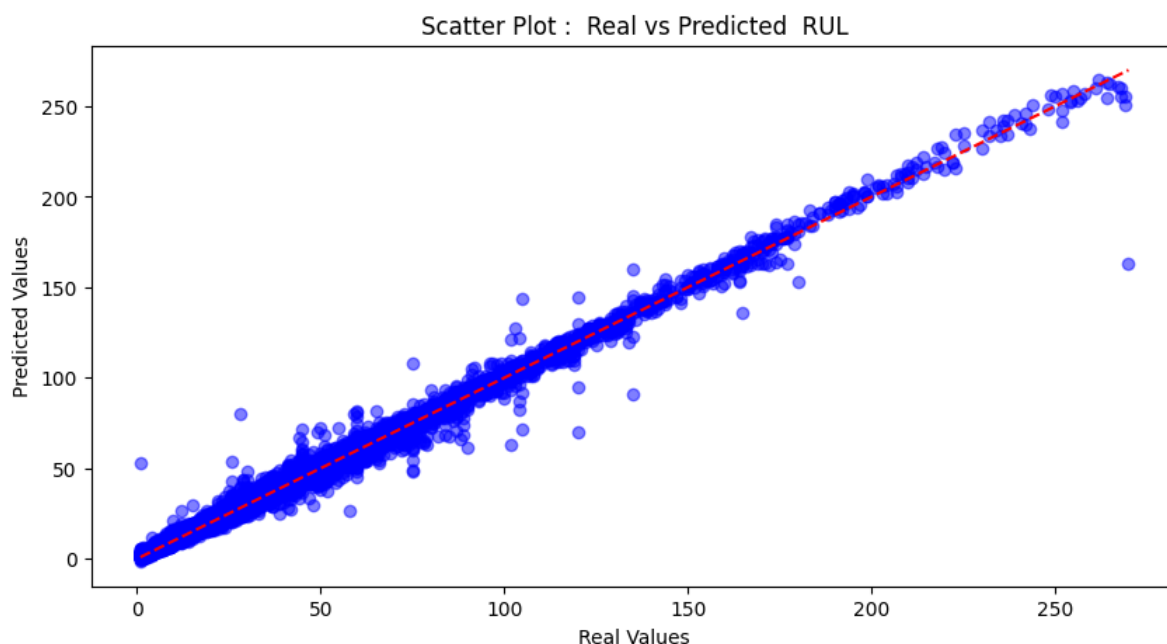


FIGURE IV.7 – La dispersion des valeurs réelles vs prédites de l’ensemble de test du dataset 1

Le diagramme à nuages de points présenté par la figure IV.7 montre la relation entre les valeurs réelles et prédites de la RUL pour chaque équipement de l’ensemble de test. Ce diagramme, marquant une corrélation positive entre les valeurs, suggère que le modèle fonctionne de manière raisonnable pour prédire la RUL des équipements. Les points de données sont dispersés autour de la ligne de régression parfaite (en rouge), mais il y a plus de points de données au-dessus de la ligne. Cela suggère que le modèle a tendance à sous-estimer la RUL plus souvent qu’il ne la surestime.

2.3 Résultats d’évaluation pour le dataset 2

Le tableau IV.2 offre une vue d’ensemble complète des résultats des capacités prédictives de notre modèle sur le deuxième dataset de filtre à gaz, permettant ainsi d’évaluer son efficacité et sa robustesse.

Métrique	Dataset 2
MSE	6.2313
RMSE	2.4963
MAE	1.7309
R2-score	0.9989

TABLE IV.2 – Résultats d’évaluations du dataset 2

Nous avons constaté que :

- L’erreur quadratique moyenne : Nous observons une très faible valeur de MSE de 6.2313 indique que les prédictions sont très précises par rapport aux valeurs réelles

de la RUL.

- La racine de l'erreur quadratique moyenne : Une valeur RMSE très basse de 2.4963, confirmant une précision élevée des prédictions.
- Erreur absolue moyenne : Une valeur de MAE minimale à 1.7309, soulignant une bonne précision des prédictions en termes d'erreur absolue moyenne.
- R2-score : Nous avons une valeur de R2-score très élevé de 0.9989 qui indique que le modèle a une forte capacité à expliquer la variation observée dans la RUL. Cela signifie que le modèle est efficace pour capturer et rendre compte des tendances dans les données de la RUL.

Visualisation des résultats pour le dataset 2

Nous avons tracer des graphique pour observer les tendances de sur-apprentissage ou de sous-apprentissage via les courbes de perte d'entraînement et de test. Ensuite, un autre graphique a comparé les valeurs réelles de la variable cible RUL avec les prédictions du modèle sur l'ensemble de test, évaluant ainsi la précision des prédictions et détectant d'éventuels écarts entre valeurs réelles et prédites pour le dataset 2.

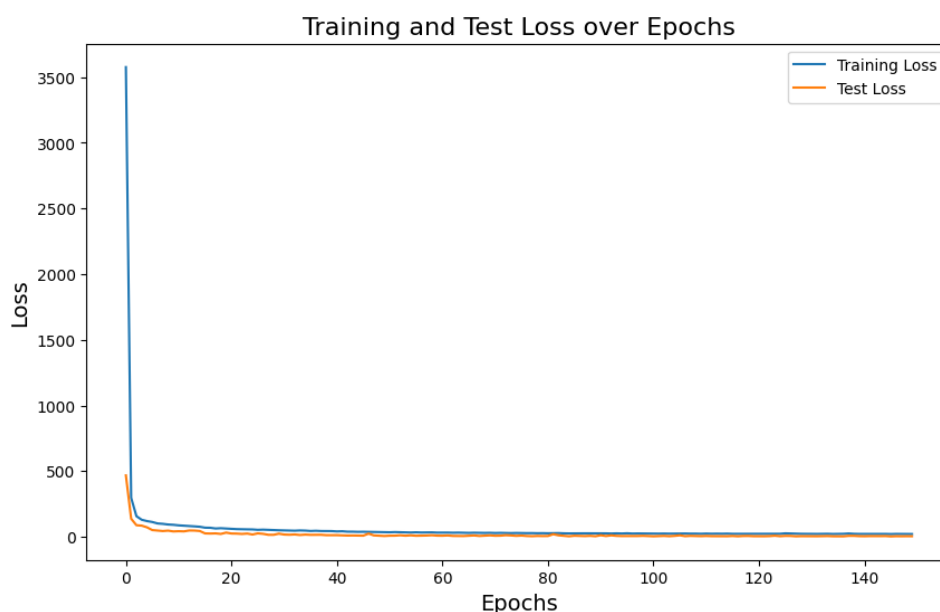
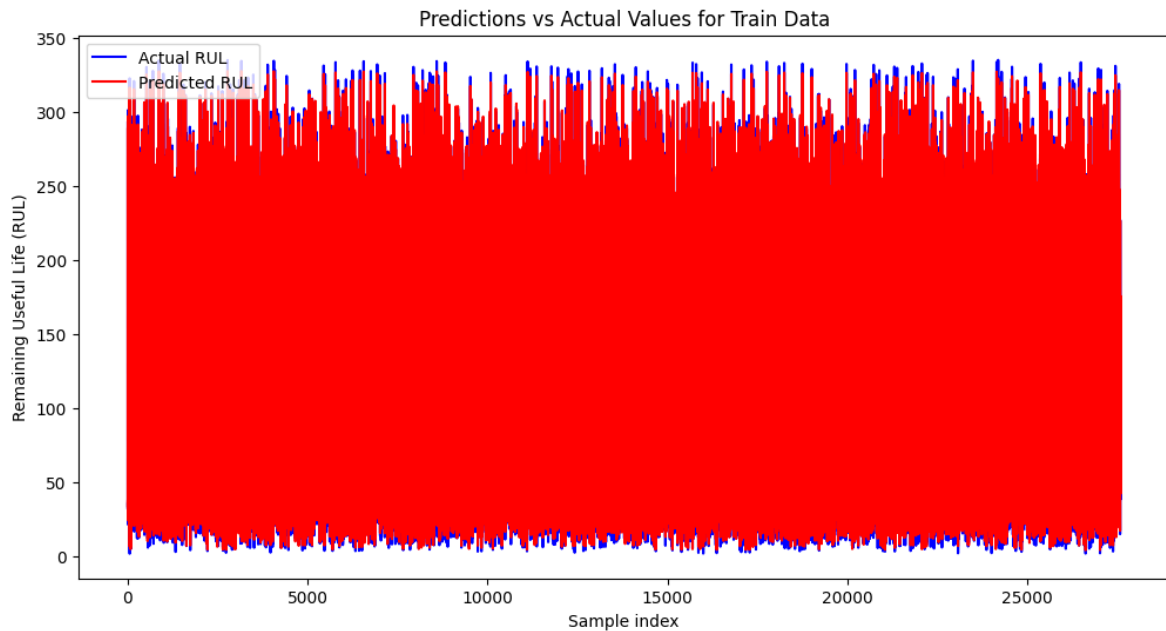


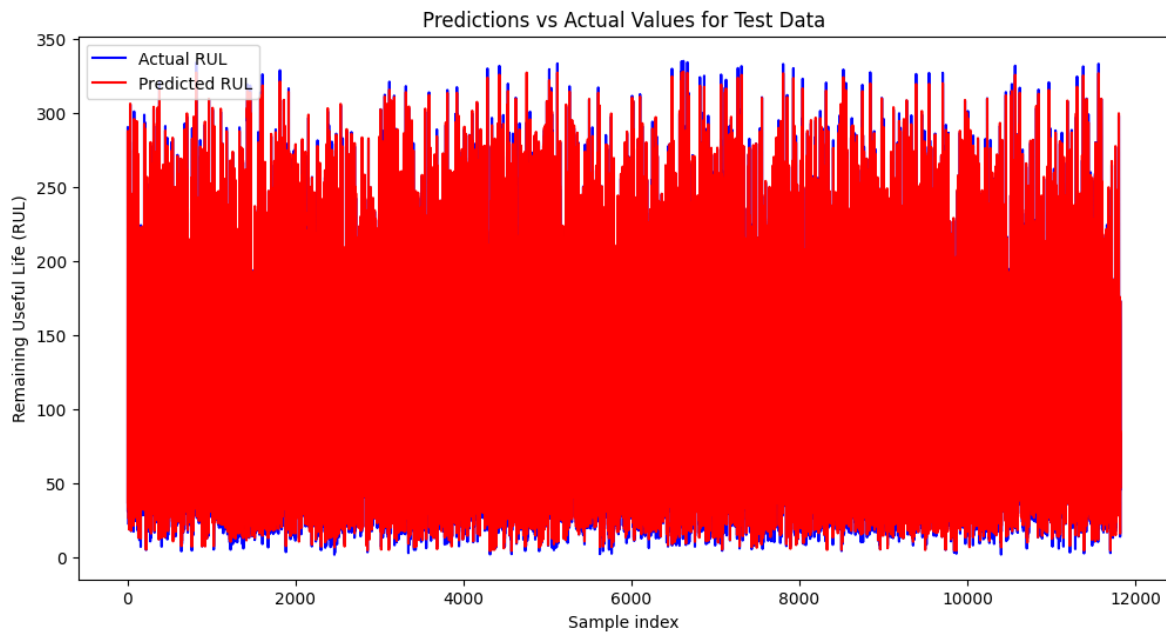
FIGURE IV.8 – Les courbes des pertes d'entraînement et de test au fil des époques pour le dataset 2

Le graphe sur la figure IV.8 illustre l'évolution de la fonction de perte MSE pour notre modèle prédisant la RUL du filtre. On observe une diminution rapide et significative des pertes d'entraînement et de test dans les premières époques, suivie d'une convergence et d'une stabilisation après environ 20 époques. Les courbes d'entraînement et de test restent très proches tout au long de l'entraînement, indiquant une bonne généralisation du modèle sans signe de sur-apprentissage. La stabilité des courbes après la convergence suggère que le modèle a atteint un état d'apprentissage optimal, capable de prédire efficacement la RUL sur des données nouvelles. Cette performance équilibrée entre l'entraînement et le test démontre la robustesse notre solution pour cette

tâche de prédiction.



(a) (a) Graphe visualisant la précision des prédictions pour l'ensemble d'entraînement



(b) (b) Graphe visualisant la précision des prédictions pour l'ensemble de test

FIGURE IV.9 – Visualisation de la précision des prédictions pour le dataset 2

Les figures IV.9 les graphiques comparant les prédictions (rouges) aux valeurs réelles (bleues) de la RUL pour les données d'entraînement et de test montrent une proximité significative entre les deux séries. Cette observation suggère que le modèle LSTM a réussi à capturer efficacement les tendances sous-jacentes des données, avec des prédictions qui suivent de près les valeurs réelles. Cela indique généralement un bon ajustement du modèle aux données disponibles.

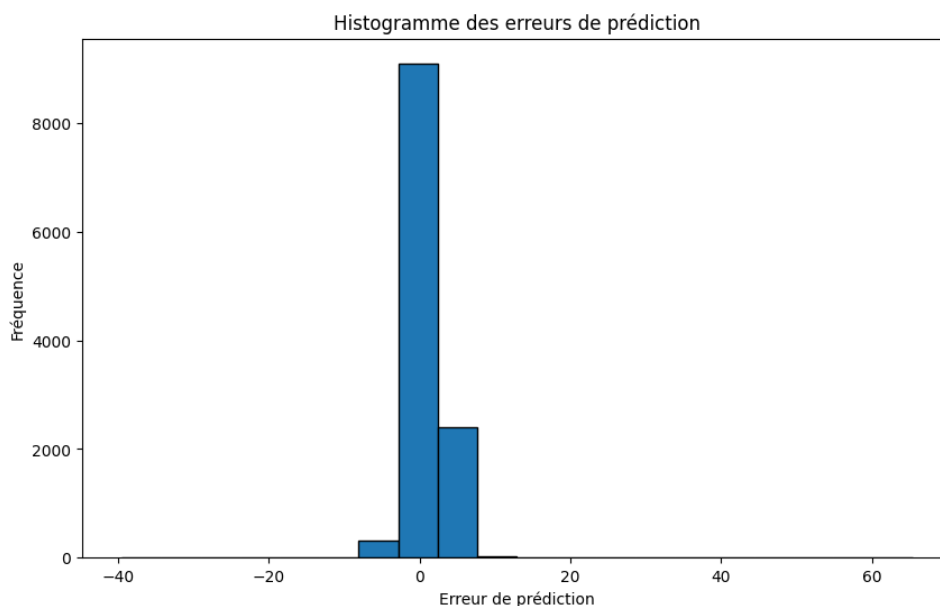


FIGURE IV.10 – L’histogramme visualisant la distribution des erreurs de prédiction pour le dataset 2

Le graphique figurant dans IV.10 représente un histogramme des résidus de test pour notre modèle LSTM prédisant la RUL. Les résidus, c’est-à-dire les différences entre les valeurs prédites et réelles. Cette histogramme des résidus montre une distribution centrée autour de zéro, indiquant des prédictions globalement précises.

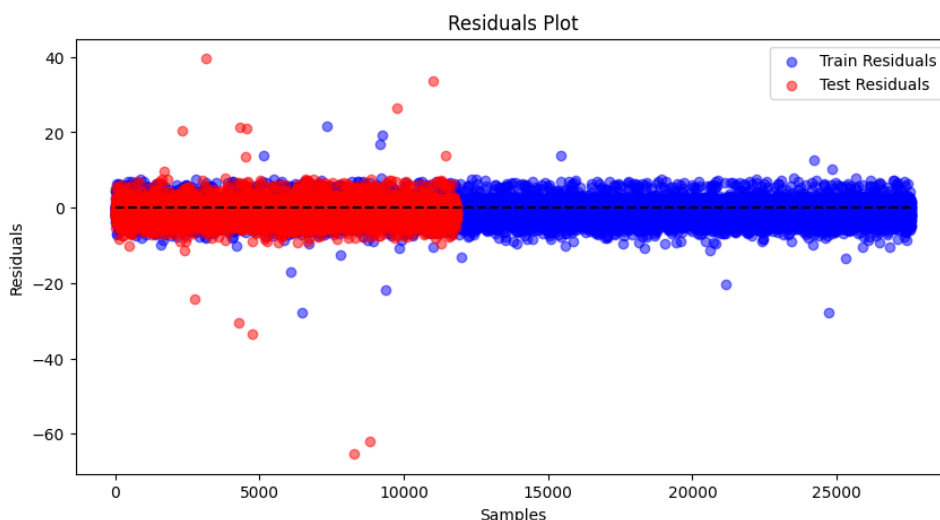


FIGURE IV.11 – Graphe des résidus du dataset 2

La figure IV.11 montre un graphe des résidus montre les différences entre les valeurs prédites et réelles de notre modèle utilisé pour prédire RUL, Avec des points bleus pour les données d’entraînement et rouges pour les données de test. Les résidus d’entraînement sont concentrés autour de zéro, indiquant une bonne performance sur les données connues. Cependant, les résidus de test présentent une légère variance et quelques valeurs aberrantes, suggérant une performance très satisfaisantes sur les

données non vues du test.

	Real	Predicted
0	31.8	37.234905
1	290.3	287.644928
2	143.6	141.499725
3	35.2	36.844704
4	60.6	60.538769
...
11820	83.8	85.415428
11821	14.1	16.185516
11822	173.2	172.970184
11823	152.3	152.953857
11824	46.0	47.416183

[11825 rows x 2 columns]

FIGURE IV.12 – Tableau des valeurs réelles et prédites de la RUL pour l’ensemble de test du dataset 2

Le tableau figurant dans IV.12 montre une comparaison entre les valeurs réelles et prédites de RUL sur un ensemble de 11825 échantillons. Le modèle démontre une précision remarquable, avec des prédictions très proches des valeurs réelles sur une large gamme de la RUL. On observe de légères variations entre les valeurs réelles et prédites, mais ces écarts restent généralement minimales. La performance du modèle semble constante tout au long du jeu de données.

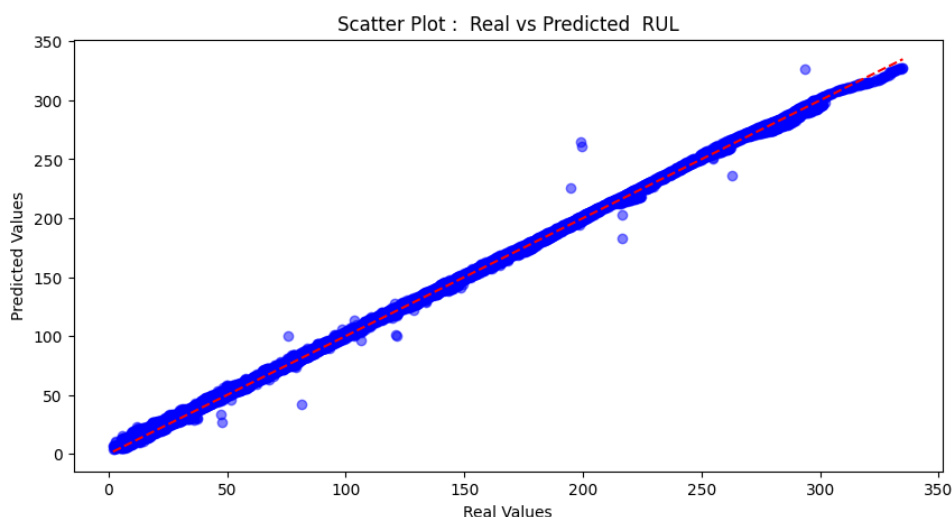


FIGURE IV.13 – Le nuage de points de dispersion des valeurs réelles vs prédites pour l’ensemble de test du dataset 2

Le diagramme de dispersion présenté dans la figure IV.13 comparant les valeurs réelles de la (RUL) aux prédictions notre modèle révèle une corrélation positive forte et gé-

nérale entre les deux variables, illustrée par une ligne diagonale bien alignée. La précision globale des prédictions est élevée, avec la plupart des points alignés près de cette diagonale, indiquant que le modèle est souvent précis. Il couvre efficacement une large gamme de valeurs RUL. Malgré quelques écarts observés, surtout dans la plage moyenne il montre une performance robuste aux extrémités des valeurs RUL

2.4 Étude comparative

La figure IV.3 ci-dessus présente un tableau comparatif entre différentes méthodes de LSTM pour la prédiction de la durée de vie restante d'un équipement RUL.

Référence	Dataset	Type de couche	Nbrs de couches (LSTM)	Nbrs de neurones (LSTM)	Taux de dropout	Taux d'apprentissage	Nbrs epoch	RMSE	Score	MAE	MSE	R2-Score
Asif et al. [48]	C-MAPSS ensemble FD0001	LSTM , Dropout , Dense , Regression	4	60	0.1	0.2	150	7.78	100	-	-	-
Muneer et al. [46]	C-MAPSS ensemble FD003	LSTM, dense 1 , dense 2, Sortie lineaire	2	30, 20	-	-	-	11.23	-	5.47	-	-
Modèle proposé	Dataset 1	LSTM, Dropout1 , dropout2, Dense	2 (LSTM)	64	0.1	0.001	300	3.6748	-	2.1357	13.5039	0.9913
Modèle proposé	Dataset 2	LSTM, Dropout1 , dropout2, Dense	2 (LSTM)	64	0.1	0.01	150	2.4963	-	1.7309	6.2313	0.9989

TABLE IV.3 – Résultats des différents modèles

Pour effectuer une étude comparative détaillée entre les différentes études qui prédisent la RUL (Remaining Useful Life) d'un équipement, nous allons examiner plusieurs aspects des modèles présentés dans ce tableau.

— Architecture du Modèle :

- Asif et al [48], utilisent un modèle plus profond avec 4 couches LSTM, ce qui peut potentiellement capturer des relations temporelles plus complexes, mais cela peut également augmenter la complexité et la difficulté d'entraînement.
- Muneer et al [46], optent pour un modèle plus simple avec seulement 2 couches LSTM, mais avec des configurations de neurones plus modestes.
- Les modèles proposés choisissent une configuration similaire à Muneer et al. en termes de nombre de couches LSTM, mais avec des neurones plus nombreux (64 par couche), ce qui pourrait permettre une meilleure capacité de modélisation des données.

— Complexité :

Le modèle de Asif et al [48] utilise une structure plus simple avec des couches LSTM et Dense, adaptée à la fois pour capturer les séquences temporelles et produire une prédiction de régression.

Le modèle de Muneer et al [46] ajoute des couches Dense supplémentaires, ce qui peut potentiellement augmenter la capacité du modèle à apprendre des relations complexes dans les données.

Le modèle proposé utilise deux couches Dropout , Cette régularisation aide à prévenir le sur-apprentissage, ce qui est crucial pour la généralisation des modèles.

— Hyperparamètres :

- Asif et al. [48] ont utilisé un taux d'apprentissage relativement élevé de 0,2, ce qui peut influencer la convergence du modèle et nécessite une gestion efficace du surapprentissage avec le dropout. Leur modèle s'entraîne sur 150 époques, un choix qui paraît équilibré, permettant au modèle de converger de manière adéquate sans engendrer un surcoût computationnel excessif.
- Le modèle proposés diffèrent sur les jeux de données principalement dans leurs taux d'apprentissage et le nombre d'époques. Le modèle pour Dataset 1 a un taux d'apprentissage plus faible (0.001) et un plus grand nombre d'époques (300), ce qui indique un apprentissage plus lent mais potentiellement plus stable. En revanche, le modèle pour Dataset 2 a un taux d'apprentissage plus élevé (0.01) et moins d'époques (150), suggérant une convergence plus rapide mais avec des risques potentiels d'oscillations.

— Performances :

- Les résultat de Asif et al [48] ainsi que Muneer et al [46] sont assez remarquable en terme de RMSE
- Les modèles proposés montrent des performances significativement meilleures en termes de RMSE par rapport aux études précédentes de Asif et al [48] et Muneer et al [46].
- Le modèle proposé pour Dataset 2 atteint même un RMSE remarquablement basse de 2.4963, indiquant une prédiction très précise de la RUL.
- Les mesures MAE, MSE et R2-Score montrent également de bonnes performances générales des modèles proposés, suggérant une bonne capacité de généralisation et de précision dans la prédiction de la RUL.

Ainsi, Les différentes études de prédiction de la RUL des équipements industriels mettent en lumière des approches variées en termes d'architecture de modèle et d'hyperparamètres. Alors qu'Asif et al [48] privilégient un LSTM profond avec plusieurs couches pour capturer des relations temporelles complexes, Muneer et al [46] optent pour une architecture plus simple avec moins de couches LSTM mais avec un mécanisme d'attention . Les modèles proposés pour Dataset 1 et Dataset 2 se positionnent avec des architectures intermédiaires, utilisant deux couches LSTM mais avec une capacité de modélisation augmentée grâce à un nombre plus élevé de neurones par couche. Les performances globales des modèles proposés montrent souvent une amélioration notable par rapport aux approches précédentes, notamment en termes de précision de prédiction mesurée par des RMSE réduits. Cette diversité d'approches souligne l'importance de choisir judicieusement l'architecture du modèle et les hyper-paramètres

en fonction des caractéristiques spécifiques des données pour optimiser la prédiction de la RUL et soutenir efficacement les stratégies de PDM dans l'industrie.

3 Conclusion

Dans ce chapitre dédié à aux résultat notre modèle prédictif de la durée de vie restante d'un équipement RUL dans la maintenance prédictive, nous avons obtenu des résultats remarquablement prometteurs. Nos analyses et expérimentations ont révélé une précision significative dans la prédiction de la durée de vie restante des équipements, ouvrant ainsi la voie à des interventions de maintenance plus prédictives et efficaces. Ces résultats témoignent de l'efficacité de notre approche et justifient son adoption pour améliorer la fiabilité opérationnelle et réduire les coûts associés à la maintenance. En conclusion, notre modèle représente une avancée substantielle dans le domaine de la maintenance prédictive, offrant des perspectives tangibles pour optimiser la gestion des actifs industriels à l'avenir.

Conclusion et perspectives

La maintenance industrielle est cruciale pour assurer la fiabilité et la durabilité des équipements, minimisant ainsi les temps d'arrêt imprévus et les pertes de production. Elle permet également d'optimiser les coûts en planifiant les interventions de manière proactive plutôt que réactive, tout en assurant la sécurité des opérations. En adoptant des stratégies avancées telles que la maintenance prédictive, les entreprises peuvent améliorer leur efficacité opérationnelle et maintenir leur compétitivité sur le marché.

Dans ce mémoire, nous avons étudié comment la science des données peut être mise au service de la maintenance industrielle proactive. Nous avons exploré les concepts fondamentaux des systèmes d'information, notamment la gestion de la maintenance assistée par ordinateur (GMAO), et comment ils peuvent être intégrés avec des techniques avancées de science des données pour anticiper et prévenir les pannes d'équipements. Nous avons étudié les techniques d'analyse de données, notamment l'apprentissage automatique et profond, en mettant l'accent sur leur utilisation pour la prédiction des pannes et les défaillances des équipements.

Dans notre étude approfondie de l'état de l'art, nous avons constaté que l'utilisation de techniques avancées d'apprentissage automatique et profond permet d'améliorer considérablement la maintenance prédictive. Nous avons investigué les différentes méthodes appliquées dans ce domaine pour mieux comprendre leur efficacité et leurs applications pratiques. Notamment, les réseaux de neurones récurrents (RNN), offre une approche robuste pour la maintenance prédictive dans l'industrie. En se concentrant spécifiquement sur la prédiction de la durée de vie restante RUL des équipements, cette méthode permet non seulement d'anticiper les défaillances potentielles mais aussi de planifier efficacement les interventions de maintenance.

L'avantage clé de prédire la RUL est la capacité à prendre des décisions proactives basées sur des données en temps réel, réduisant ainsi les interruptions imprévues de la production et les coûts associés à la réparation d'équipements défaillants. En exploitant les caractéristiques temporelles des données collectées, comme les séries temporelles des performances des équipements, les modèles RNN peuvent capturer les tendances et les motifs complexes qui indiquent la dégradation des équipements.

En intégrant ces approches avancées dans nos stratégies de maintenance, nous visons à optimiser la disponibilité opérationnelle des équipements tout en minimisant les coûts de maintenance. Cette approche proactive transforme fondamentalement la gestion des actifs industriels en permettant aux entreprises de passer d'une réparation réactive à une maintenance prédictive et préventive, alignée sur des objectifs de production efficace et durable.

À travers notre solution basée sur les réseaux de neurones LSTM (Mémoire à Long Terme et à Court Terme), reconnus pour leurs performances supérieures dans la main-

tenance prédictive, nous avons significativement réduit les pertes et amélioré les résultats en optimisant les hyperparamètres et la structure du réseau neuronal. Notre approche vise spécifiquement à prédire la durée de vie restante (RUL) des équipements industriels. Nous avons entraîné notre modèle sur deux ensembles de données industrielle dans un environnement de développement adapté. Nous avons évalué ses performances à l'aide de métriques telles que la RMSE, le MSE, le MAE et le R² score. Les résultats obtenus démontrent clairement que notre architecture proposée surpasse les méthodes précédentes, confirmant ainsi l'efficacité des techniques d'intelligence artificielle, notamment de l'apprentissage profond, en matière de maintenance prédictive.

Pour l'avenir, nous souhaitons appliquer notre solution à des cas concrets dans notre pays. Nous envisageons d'encourager l'adoption de l'industrie 4.0 par nos entreprises, notamment l'entreprise Cevital. Cela inclurait l'installation de capteurs capables de collecter des données machines à des intervalles précis. Ces données permettraient de générer des rapports précis, sur lesquels nous pourrions effectuer une analyse de données et déployer notre solution pour prédire les pannes, améliorant ainsi l'efficacité opérationnelle et réduisant les coûts de maintenance.

Annexe

Les architectures de la maintenance prédictive

Pour concevoir et mettre en œuvre un système de maintenance prédictive PDM efficace, il est important de s'appuyer sur des architectures appropriées, telles que l'Architecture Système Ouverte pour la Surveillance Basée sur l'État (OSA-CBM), l'informatique en nuage (cloud) et l'industrie 4.0 [63].

■ Architecture Système Ouverte pour la Surveillance Basée sur l'État (OSA-CBM)

L'OSA-CBM, ou "Open System Architecture for Condition-Based Monitoring" (Architecture Système Ouverte pour la Surveillance Basée sur l'État), définie dans la norme ISO 13374 en 2003, qui est détenu par l'Alliance Ouverte des Systèmes de Gestion de l'Information des Machines (MIMOSA). OSA-CBM fournit un cadre unifié et en couche ainsi elle définit des formats et des méthodes pour communiquer, présenter et afficher des informations pertinentes dans le contexte de la surveillance et de la maintenance des équipements industriels [4].

Le cadre de OSA-CBM comprends sept couches fonctionnelles qui sont les suivantes :

1. Acquisition de données : Cette couche fournit l'accès aux capteurs installés sur les équipements et collecte les données de fonctionnement.
2. Manipulation des données : Cette couche effectue des transformations sur les données collectées, telles que des filtrages ou des conversions, et applique des algorithmes d'extraction de caractéristiques pour extraire des informations pertinentes.
3. Détection de l'état : Cette couche surveille l'état de fonctionnement des équipements en comparant les caractéristiques des données aux valeurs attendues ou aux limites opérationnelles définies, et génère des indicateurs d'état ou des alarmes en cas de déviations significatives.
4. Évaluation de la santé : Cette couche évalue la santé globale des équipements en analysant les tendances de l'historique de la santé, du statut opérationnel et de l'historique de maintenance, afin de détecter les signes de détérioration ou de défaillance imminente.
5. Évaluation de la prognose : Cette couche projette l'état de santé actuel des équipements dans le futur en prenant en compte les estimations des profils d'utilisation futurs, ce qui permet de prévoir les besoins en maintenance et de planifier les interventions de manière proactive.
6. Génération de recommandations : Enfin, cette couche fournit des recommandations pour les activités de maintenance et les modifications de configuration des équipements, en tenant compte de l'historique opérationnel, des profils de mission actuels et futurs, ainsi que des contraintes de ressources.

■ L'informatique en nuage

Le système de Maintenance Prédictive PDM amélioré par le cloud tire son inspiration du potentiel de l'informatique en nuage (cloud computing) et de la fabrication en nuage (cloud manufacturing). Le cloud permet la fourniture de ressources

informatiques telles que les infrastructures, les plateformes et les applications en tant que services, tandis que la fabrication en nuage transforme les ressources et les capacités de fabrication en services de fabrication adaptatifs, sécurisés et à la demande via l'Internet des objets (IoT). En tant qu'élément important de la fabrication en nuage, la PDM est améliorée par le concept de cloud pour soutenir les entreprises ou les usines dans le déploiement et la gestion des services de PDM via Internet. Les systèmes de maintenance améliorés par le cloud fournissent un nouveau paradigme de provisionnement de maintenance, c'est-à-dire la maintenance en tant que service.

Dans cette architecture générique de la PDM améliorée par le cloud, plusieurs étapes clés sont impliquées :

1. Surveillance de l'état des machines : Les données sont collectées à distance et dynamiquement depuis le plancher de l'usine via des capteurs équipés.
2. Traitement distant des données : Les données collectées sont nettoyées, intégrées et les caractéristiques pertinentes sont extraites à distance.
3. Diagnostic et pronostication : Les données traitées sont utilisées pour identifier les défaillances actuelles et prédire les défaillances potentielles à venir.
4. Planification de la maintenance prédictive : Les résultats du diagnostic et de la pronostication servent à planifier la PDM, qui peut être exécutée à distance et dynamiquement sur le plancher de l'usine.

Cette architecture permet à l'équipement connecté de fournir des données en tant que service à la PDM améliorée par le cloud. De plus, l'équipement peut souscrire à des services de pronostication ou de maintenance. En plus de l'informatique en nuage et de la fabrication en nuage, les technologies de support incluent également IoT, les systèmes embarqués, le web sémantique et la communication machine à machine. D'un point de vue de la Maintenance Prédictive PDM, l'environnement du cloud peut efficacement soutenir divers services intelligents et résoudre plusieurs problèmes tels que la capacité mémoire des équipements, la puissance de calcul du processeur, la sécurité des données et la fusion des données provenant de sources multiples. Par conséquent, ce paradigme de PDM amélioré par le cloud présente les caractéristiques suivantes :

- Orienté service : Toutes les fonctions de la PDM peuvent être dérivées sous forme de services basés sur le cloud. Les utilisateurs n'ont plus besoin d'héberger ni de maintenir un grand nombre de serveurs informatiques ou de logiciels associés.
- Accessible et robuste : Les services de maintenance à la demande via Internet peuvent accroître l'accessibilité, tandis que les services modulaires et configurables peuvent accroître la robustesse et l'adaptabilité.
- Sensible aux ressources : Le stockage des données de surveillance et le calcul analytique peuvent être effectués localement ou à distance, il doit être sensible aux ressources pour prendre des décisions de maintenance afin de réduire la quantité de transmission de données.
- Collaboratif et distribué : L'informatique en nuage facilite le partage et l'échange d'informations entre différentes applications/machines à différents endroits de manière transparente et collaborative.

■ L'industrie 4.0

L'Industrie 4.0 représente un changement de paradigme dans les processus et les produits industriels, propulsé par des approches intelligentes de traitement de l'information, des systèmes de communication, des techniques orientées vers l'avenir, et bien plus encore. L'objectif de l'Industrie 4.0 est de rendre une machine ou une usine plus intelligente. Le terme "intelligent" ne se limite pas seulement à améliorer la gestion de la production, mais aussi à réduire les temps d'arrêt des équipements. La Maintenance Prédicative PdM alignée sur les principes de l'Industrie 4.0, dessine une feuille de route pour des systèmes de PdM intelligents.

La PdM 4.0 utilise une analyse avancée et en ligne des données collectées pour détecter plus tôt l'occurrence de possibles défaillances de machines, et elle soutient les techniciens pendant les interventions de maintenance en fournissant un support décisionnel intelligent guidé. Les stratégies de maintenance sont classées en 4 niveaux appliqués dans les industries modernes :

- Niveau 1 : Inspections visuelles : ce niveau consiste en des inspections physiques périodiques, et les stratégies de maintenance sont basées uniquement sur l'expertise des inspecteurs.
- Niveau 2 : Inspections instrumentales : ce niveau consiste en des inspections périodiques, et les stratégies de maintenance sont basées sur une combinaison de l'expertise des inspecteurs et des lectures des instruments.
- Niveau 3 : Surveillance continue en temps réel de l'état : ce niveau consiste en une surveillance continue en temps réel des actifs, et des alertes sont données en fonction de règles prédéfinies ou de niveaux critiques.
- Niveau 4 : PdM 4.0 : Ce niveau consiste en une surveillance continue en temps réel des actifs, et des alertes sont délivrées en fonction de techniques prédictives, telles que l'analyse de régression.

Bibliographie

- [1] ENODEA. Logiciel erp ou pgi : Définition. <https://www.enodea.com/definition-logiciel-erp/>, 2022. [Consulté le 28 avril 2024].
- [2] Amir Sahraei, Alejandro Chamorro, Philipp Kraft, and Lutz Breuer. Application of machine learning models to predict maximum event water fractions in stream-flow. *Frontiers in Water*, 3 :652100, 06 2021.
- [3] Zhe wu and Panagiotis Christofides. Economic machine-learning-based predictive control of nonlinear systems. *Mathematics*, 7 :494, 06 2019.
- [4] Zhou Xin Lin Pengfeng Wen Yonggang Deng Ruilong Ran, Yongyi. A survey of predictive maintenance : Systems, purposes and approaches. *IEEE Communications Surveys & Tutorials*, 2019.
- [5] Sahbi Sidhom, Malek Ghenima, and Philippe Lambert. *Systèmes d'information et Intelligence économique : enjeux et perspectives*. Open Hall, 2010.
- [6] A.B. Ondimba. *Le Management des affaires : Esquisse de réflexion sur 25 thèmes majeurs de management*. Publibook/Société des écrivains, 2012.
- [7] A.A. Sambrekar, C.R. Vishnu, and R. Sridharan. Maintenance strategies for realizing industry 4.0 : An overview. In Jiji Vanchipura, editor, *Emerging Trends in Engineering, Science and Technology for Society, Energy and Environment*, pages 341–348, 2018.
- [8] Richard T. Watson. *Information Systems*. 2007.
- [9] Alaa F. Sheta, Aladdin Ayesh, and Peter Rausch. *Business Intelligence and Performance Management*. Theory, Systems and Industrial Applications, 2013.
- [10] Kalyan Chatterjee and William Samuelson. *Game Theory and Business Applications*, volume 194. International Series in Operations Research Management Science, 2013.
- [11] Klaus-Dieter Gronwald. *Integrated Business Information Systems : A Holistic View of the Linked Business Process Chain ERP-SCM-CRM-BI-Big Data*. Textbook, 2017.
- [12] Longbing Cao. Data science : A comprehensive overview. *ACM Computing Surveys (CSUR)*, 50(3) :43, 2017.
- [13] Andrigh Van Wyk. *Machine Learning with LightGBM and Python : A practitioner's guide to developing production-ready machine learning systems*. Packt Publishing, September 2023.

- [14] E. Biernat, M. Lutz, and Y. LeCun. *Data science : fondamentaux et études de cas : Machine learning avec Python et R*. Eyrolles, 2015.
- [15] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Sebastopol, CA, 2nd edition edition, 2019.
- [16] Vinita Silaparasetty. *Deep Learning Projects Using TensorFlow 2 : Neural Network Development with Python and Keras*. Packt Publishing, Birmingham, UK, 2020.
- [17] Paul Fergus and Carl Chalmers. *Applied Deep Learning : Tools, Techniques, and Implementation*. Computational Intelligence Methods and Applications, 1st edition, 2022. ISBN : 978-3-030-72182-5.
- [18] Foster Provost and Tom Fawcett. *Data Science for Business*. O'Reilly Media, Inc., August 2013.
- [19] Avinash S. Jagtap. *Data-Driven Decision Making*, volume 314. Lulu.com, 2019.
- [20] Cevital. Cevital - premier groupe privé algérien, 2024. Consulté le 11 janvier 2024.
- [21] P. Coandă, M. Avram, and V. Constantin. A state of the art of predictive maintenance techniques. *IOP Conf. Series : Materials Science and Engineering*, 997 :012039, 2020.
- [22] Chulsoon Park et al. A predictive maintenance approach based on real-time internal parameter monitoring. *The International Journal of Advanced Manufacturing Technology*, 85(1-4) :623–632, 2016.
- [23] Paulina Gackowiec. General overview of maintenance strategies - concepts and approaches. *MAPE*, 2(1) :126–139, 2019. Date of submission to the Editor : 05/2018, Date of acceptance by the Editor : 08/2018, ORCID ID : 0000-0002-9669-6879.
- [24] Collin Drent, Stella Kapodistria, and Onno J. Boxma. Censored lifetime learning : Optimal bayesian age-replacement policies. *Operations Research Letters*, 48(6) :827–834, November 2020.
- [25] I. S. Lopes, P. Senra, B. Neto, R. Costa, M. Sousa, T. Cabo, and J. A. Oliveira. Multi-criteria classification for prioritization of preventive maintenance tasks to support maintenance scheduling. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, 2017.
- [26] Shadi Sanoubar, Lisa M. Maillart, and Oleg A. Prokopyev. Age-replacement policies under age-dependent replacement costs. pages 425–436, October 2020. Received 28 Aug 2019, Accepted 27 Jul 2020, Published online : 23 Oct 2020.
- [27] H. M. Hashemian and Wendell C. Bean. State-of-the-art predictive maintenance techniques. *IEEE Transactions on Instrumentation and Measurement*, 60(10), October 2011.
- [28] Irfan Ullah et al. Maintenance prédictive des équipements des sous-stations électriques par thermographie infrarouge à l'aide d'une approche d'apprentissage automatique. *Énergies*, 10(12) :1987, 2017.

- [29] Giuseppe Curcurù, Giacomo Galante, and Alberto Lombardo. Une politique de maintenance prédictive avec une surveillance imparfaite. *Reliability Engineering & System Safety*, 95(9) :989–997, 2010.
- [30] Heletjé E. van Staden, Laurens Deprez, and Robert N. Boute. A dynamic "predict, then optimize" preventive maintenance approach using operational intervention data. *European Journal of Operational Research*, 302(1), January 2022.
- [31] Jinjiang Wang et al. A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. *Journal of Intelligent Manufacturing*, 28(5) :1125–1137, 2017.
- [32] D. Mourtzis and E. Vlachou. A cloud-based cyber-physical system for adaptive shop-floor scheduling and condition-based maintenance. *Journal of Manufacturing Systems*, 47 :179–198, 2018.
- [33] B. Edrington, B. Zhao, A. Hansel, M. Mori, and M. Fujishima. Machine monitoring system based on mtconnect technology. *Procedia CIRP*, 22 :92–97, 2014.
- [34] Ana Cachada et al. Maintenance 4.0 : Architecture du système de maintenance intelligente et prédictive. In *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, volume 1. IEEE, 2018.
- [35] Moulahoum S. Benkercha, R. Fault detection and diagnosis based on c4.5 decision tree algorithm for grid connected pv system. *Solar Energy*, 173 :610–634, 2018.
- [36] M. Hassan A. Hamid M. Bakir, A. Zaman. Prediction of remaining useful life for mech equipment using regression. In *Journal of Physics : Conference Series*, volume 1150, page 120. IOP Publishing, 2019.
- [37] R. Dharithri B;Sharma. Sripradha. Nikita . Ashwini Kodipalli. Trupthi Rao. Rohini, B. Machine predictive maintenance classification using machine learning. In *Actes de la Conférence internationale sur l'intelligence informatique pour l'information, la sécurité et les applications de communication (CIISCA)*. IEEE, IEEE, 2023.
- [38] J. Wei and Z. Dong, G. Chen. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. *IEEE Transactions on Industrial Electronics*, 65(7) :5634–5643, July 2018.
- [39] P.-H. Hao Y.-S. Zhao M. Chen, X.-L. Wang. Evidential knn-based condition monitoring and early warning method with applications in power plant. *Neurocomputing*, 315 :18–32, 2018.
- [40] Y. Guo-L. Lin J. Xing S. Jia, F. Lei. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing*, 272 :619–628, 2018.
- [41] C. Chen-X. Ma, M. Sun. Deep coupling autoencoder for fault diagnosis with multimodal sensory data. *IEEE Transactions on Industrial Informatics*, 14(3) :1137–1145, 2018.
- [42] S. Haidong, J. Hongkai, and W. Xingqiu, L. Shuaipeng. Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. *Knowledge-Based Systems*, 140 :1–14, 2018.

- [43] L. Hong-S. Zhao S. Wang H. Zhang L. Ren, L. Zhao. Remaining useful life prediction for lithium-ion battery : A deep learning approach. *IEEE Access*, 6 :50587–50598, 2018.
- [44] W. Zeng-X. Yang C. Zhou X. Liu, Z. Mei. Remaining useful life estimation of insulated gate bipolar transistors (igbts) based on a novel volterra k-nearest neighbor optimally pruned extreme learning machine (vkopp) model using degradation data. *Sensors*, 17(11) :2524, 2017.
- [45] K. Ren-L. Qin, H. Xu. Rolling bearings fault diagnosis via 1d convolution networks. In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pages 617–621. IEEE, 2019.
- [46] A Muneer, SM Taib, S Naseer, RF Ali, and IA Aziz. Data-driven deep learning-based attention mechanism for remaining useful life prediction : Case study application to turbofan engine analysis. *Electronics*, 2021.
- [47] Mustapha . Hajji Adnene Hesabi, Hadis . Nourelfath. A deep learning predictive model for selective maintenance optimization. *Reliability Engineering and System Safety*, 219, 2022.
- [48] Haider Sajjad Ali Naqvi Syed Rameez Zaki John F. W. Kwak Kyung-Sup Islam S. M. Riazul Asif, Owais. A deep learning model for remaining useful life prediction of aircraft turbofan engine on c-mapss dataset. *IEEE Access*, 2022.
- [49] Ancha . Wang Xiao-Lin Zhuang, Liangliang . Xu. A prognostic driven predictive maintenance framework based on bayesian deep learning. *Reliability Engineering and System Safety*, 234, 2023.
- [50] Ayush Gandhi. Integrating big data analytics and machine learning for predictive maintenance within industrial iot frameworks. In *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*. Centre de recherche interdisciplinaire en commerce et en technologie, Institut universitaire d'ingénierie et de technologie de Chitkara, Université de Chitkara, Pendjab, Inde, 2023.
- [51] R. Liu and A. G. Yang, B. Hauptmann. Simultaneous bearing fault recognition and remaining useful life prediction using joint loss convolutional neural network. *IEEE Transactions on Industrial Informatics*, 2019. En attente de l'indication du volume, du numéro et des pages.
- [52] G. Chen L. Huang X. Xia T. Song, Y. Shi. Remaining useful life prediction of turbofan engine using hybrid model based on autoencoder and bidirectional long short-term memory. *Journal of Shanghai Jiaotong University (Science)*, 23 :85–94, 2018.
- [53] Y. Xing S. Yan T. Li N. Guo, L. Lei. Deep convolutional transfer learning network : A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, 66(9) :7316–7325, 2018.
- [54] X. Tang S. Meng F. Pan, H. He. An improved bearing fault diagnosis method using one-dimensional cnn and lstm. *Strojnicki Vestnik/Journal of Mechanical Engineering*, 64, 2018.

- [55] Kamal Habib, Maki K. Mohamed. Machine learning-based predictive maintenance : Using cnn-lstm network. In *Proceedings of 2023 IEEE International Conference on Mechatronics and Automation*. IEEE, 2023.
- [56] Y. Chen Q. Tang Z. Li L. Gui W. Peng, C. Chen. A remaining useful life prognosis of turbofan engine using temporal and spatial feature fusion. *Sensors*, 21(2) :418, 2021.
- [57] Zhiqiang Cui, Hongli Jia, Qi Gao, and Haowen Song. Maintenance spare parts prediction based on multilevel migration learning cnn-ise-attention-bilstm. *IEEE Access*, 12, 2024.
- [58] Tiago Togo. Predictive useful life based into telemetry, 2024. Accessed : 2024-05-01.
- [59] Prognostics HSE. Preventive to predictive maintenance, 2024. Accessed : 2024-05-01.
- [60] Google. Google Colaboratory, 2024. Consulté le 4 avril 2024.
- [61] Seaborn Development Team. Seaborn : statistical data visualization, 2024. Consulté le 25 avril 2024.
- [62] PyTorch Contributors. PyTorch : An open source machine learning framework, 2024. Consulté le 1 mai 2024.
- [63] Y. Ran, S. Yang, S. Zhang, and H. Xi. Dynamic iaas computing resource provisioning strategy with qos constraint. *IEEE Transactions on Services Computing*, 10(2) :190–202, March 2017.

Résumé

La maintenance prédictive basée sur l'analyse des données vise à résoudre les problèmes liés aux défaillances soudaines des machines et à réduire les temps d'arrêt. Nous avons réalisé une revue exhaustive de la littérature en analysant plusieurs articles de recherche dans ce domaine. Cette étude nous a permis de comparer nos pratiques actuelles et de développer une approche innovante en utilisant une architecture basée sur les techniques de l'intelligence artificielle. Notre étude s'est concentrée sur l'application d'une méthode d'apprentissage profond DL notamment le modèle LSTM, implémenté en langage Python. Nous avons testé notre méthode sur deux ensembles de données afin de prédire la durée de vie restante des équipements avant leur panne. En comparant les résultats de nos prédictions, nous avons observé que notre architecture LSTM surpassait d'autres recherches dans le domaine en termes de précision et de taux d'erreur.

Mots clés : maintenance prédictive; LSTM; apprentissage profond; intelligence artificielle.

Abstract

Predictive maintenance based on data analysis aims to address issues related to sudden machine failures and reduce downtime. We conducted a comprehensive literature review by analyzing several research articles in this field. This study allowed us to compare our current practices and develop an innovative approach using an architecture based on artificial intelligence techniques. Our study focused on applying a deep learning (DL) method, specifically the Long Short-Term Memory (LSTM) model, implemented in Python. We tested our method on two datasets to predict the remaining useful life of equipment before failure. By comparing our prediction results, we observed that our LSTM architecture outperformed other research in the field in terms of accuracy and error rate.

Key words : predictive maintenance; LSTM; deep learning; artificial intelligence.