

# Utilisation des dispositifs MEMS pour la maintenance intelligente des machines industrielles

Par

Mohamed JEMLI

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE  
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE  
AVEC MÉMOIRE EN GÉNIE ÉLECTRIQUE  
M. Sc. A.

MONTRÉAL, LE 16 MARS 2023

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE  
UNIVERSITÉ DU QUÉBEC



Mohamed Jemli, 2023



Cette licence [Creative Commons](#) signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

**PRÉSENTATION DU JURY**

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Ricardo Izquierdo, directeur de mémoire  
Département de génie électrique à l'École de technologie supérieure

M. Frédéric Nabki, président du jury  
Département de génie électrique à l'École de technologie supérieure

M. Antoine Tahan, membre du jury  
Département de génie électrique à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 06 MARS 2023

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE



## **REMERCIEMENTS**

Un grand merci envers mon directeur de recherche, le professeur Ricardo Izquierdo pour m'avoir acceptée dans le cadre de mon mémoire. Je salue vivement ses qualités, je lui adresse toute ma gratitude pour ses disponibilités et ses précieux conseils.

Je remercie également tous les techniciens du département électrique de m'avoir accompagnée tout au long de mon mémoire. J'ai beaucoup appris avec eux et je suis très reconnaissant pour leurs soutiens, et leurs encouragements.

Je tiens à remercier profondément mes chers parents, mon frère et mes sœurs. Aucun langage ne saurait exprimer mon respect et ma considération pour leurs soutiens et encouragements.

Je tiens à exprimer ma plus grande gratitude à ma copine qui rien n'aurait été possible sans ses conseils et son amour inconditionnel tout au long de ma maîtrise, malgré la distance.

Mes remerciements sincères à tous mes amis pour leurs encouragements et leurs conseils durant ma maîtrise.



# **Utilisation des dispositifs MEMS pour la maintenance intelligente des machines industrielles**

Mohamed JEMLI

## **RÉSUMÉ**

L'anticipation des défaillances et l'estimation de l'état de santé des composants sont deux missions essentielles de la maintenance prédictive. Ce mémoire fait l'objet d'une étude pour prédire un certain nombre de défauts en utilisant des signaux capturés sur des roulements par des capteurs accéléromètres MEMS. Ces composants des machines sont très utilisés dans le monde industriel, notamment dans les moteurs électriques. Avec la naissance de l'Internet industriel des objets d'Intelligence Artificielle, la collecte des signaux a été grandement amélioré. L'utilisation des signaux enregistrés dans la prédiction des pannes et dans l'estimation de la durée de vie restante sont désormais un sujet de recherche brûlant dans le domaine de la maintenance industrielle.

Ce projet se décompose en deux parties, une première partie de l'étude sera faite sur une base de données Open Source afin de bien tester les modèles pour la détection et la classification des anomalies, ainsi que pour prédire l'état de santé des roulements. Une deuxième partie est consacrée pour concevoir un banc de tests pour enregistrer des données de vibration de deux roulements par les deux MEMS accéléromètres Adxl335 et l'Adxl 1002Z afin d'évaluer ses performances dans contexte de maintenance prédictive.

Nous sommes arrivés à prédire les défauts avant leurs apparitions en utilisant les signaux enregistrés par les deux accéléromètres avec un F-score maximal de 98.26% pour la classification des défauts par le modèle SVM, de plus une démarche méthodologique est adaptée à ce problème pour l'estimation de l'état de santé de ces composants.

**Mots-clés :** Maintenance prédictive, MEMS accéléromètres, RUL, Apprentissage Machine, Roulements, Enregistrement des données



# **The use of MEMS devices in the intelligent maintenance of industrial machines**

Mohamed JEMLI

## **ABSTRACT**

The prediction of failures and the estimation of the health status of components are two critical functions of predictive maintenance. This thesis is the subject of an investigation that aims to predict several defects using vibration signals captured from rotors bearing by MEMS accelerometers sensors. These machine components are widely used in the industrial world, particularly in electric motors. Moreover, artificial intelligence and signal collection have greatly improved, especially with the advent of the Industrial Internet of Things. The use of recorded signals in the prediction of failures and the estimation of the remaining useful life span are currently trending subjects of research in the field of industrial maintenance.

This project is divided into two parts. The first part of the research is conducted on an open-source data base to thoroughly evaluate the models for anomaly detection and classification, as well as to predict the health of the bearings. A second section is devoted to designing a test bench to record vibration data from two bearings using two MEMS accelerometers Adxl335 and Adxl1002Z in order to evaluate their performance in the context of predictive maintenance.

We were able to predict defects before their appearance using the signals recorded by the two accelerometers with a maximum F-score of 98.26% for fault classification using the SVM model. Additionally, a methodological approach has been adapted to this problem for estimating the state of health of these components.

**Keywords:** Predictive maintenance, MEMS accelerometers, RUL, Machine Learning, Bearings, and data logging



## TABLE DES MATIÈRES

	Page
INTRODUCTION .....	1
CHAPITRE 1 OBJECTIFS ET MÉTHODOLOGIE DE RECHERCHE .....	3
1.1 Introduction.....	3
1.2 Contexte et Problématique de recherche : .....	3
1.3 Objectifs :.....	4
1.4 Contribution : .....	5
CHAPITRE 2 REVUE DE LITTÉRATURE.....	7
2.1 Introduction.....	7
2.2 Maintenance 4.0.....	7
2.3 Stratégie de la maintenance prédictive .....	8
2.3.1 Acquisition et traitement des données .....	8
2.3.2 Détection et localisation de la panne .....	9
2.3.3 Pronostics et prévisions.....	9
2.3.4 Actions correctives par le diagnostic, l'optimisation et le contrôle .....	9
2.4 Surveillance des moteurs électriques .....	10
2.5 Surveillance de l'état des machines .....	10
2.6 Roulements mécaniques.....	13
2.6.1 Les défauts des roulements .....	14
2.6.1.1 Fatigue.....	14
2.6.1.2 Manque ou trop de lubrification : .....	14
2.6.1.3 Corrosion.....	15
2.6.2 Systèmes de prédictions des défauts des roulements.....	15
2.7 Accéléromètres MEMS :.....	16
2.7.1 Principe de fonctionnement : .....	17
2.7.2 Utilisations des Accéléromètres MEMS .....	17
2.8 Conclusion .....	18
CHAPITRE 3 IMPLÉMENTATION DES MODÈLES D'INTELLIGENCE ARTIFICIELLE .....	21
3.1 Introduction.....	21
3.2 Base des données .....	21
3.3 Traitement des données et extraction des descripteurs .....	23
3.3.1 Analyse des composants principales PCA.....	23
3.3.2 Descripteurs statistiques dans le domaine temporel .....	25
3.3.2.1 Valeur efficace RMS.....	25
3.3.2.2 L'aplatissement .....	25
3.3.2.3 Valeur maximale et la valeur crête à crête.....	26
3.3.2.4 Valeur Moyenne.....	26
3.3.2.5 Déviation standard ou l'écart type .....	26

3.3.2.6	Asymétrie.....	26
3.3.2.7	Entropie.....	27
3.4	Machine Learning .....	27
3.4.1	Méthodologie : Super Learner .....	27
3.4.2	Bibliothèques des algorithmes : .....	29
3.4.2.1	Xgboost (Extreme Gradient Boosting): .....	30
3.4.2.2	Linear Regression .....	31
3.4.2.3	Logistic Regression.....	31
3.4.2.4	Decision Tree .....	32
3.4.2.5	Support Vector Machine SVM .....	32
3.4.2.6	K Nearest Neighbour algorithm KNN .....	32
3.4.2.7	Random Forest.....	33
3.4.2.8	Extra Trees .....	33
3.4.2.9	Adaboost .....	33
3.4.2.10	Bagging Regressor.....	34
3.4.3	Mesures de la performance .....	34
3.4.3.1	R2 Score.....	34
3.4.3.2	Racine de l'erreur moyenne carrée (Root Mean Square Error) .	34
3.4.3.3	Erreur absolue moyenne (Mean Absolute Error).....	35
3.4.3.4	F1 score, accuarcy & precision.....	35
3.4.4	Régression.....	36
3.4.4.1	Méthodologie de simulation .....	36
3.4.4.2	Résultats de régression.....	37
3.4.5	Classification.....	39
3.4.5.1	Résultats : .....	41
3.4.6	Clustering .....	44
3.4.6.1	Résultats : .....	45
3.4.7	Détection d'anomalies et classification des défauts.....	49
3.4.7.1	Isolation Forest.....	49
3.4.7.2	Détection d'anomalies par l'algorithme Isolation Forest.....	50
3.4.7.3	Classification des anomalies .....	51
3.4.7.4	Résultats : .....	52
3.4	Conclusion .....	56
 CHAPITRE 4 ENREGISTREMENT DES DONNÉES ET LA VALIDATION DES RÉSULTATS..... 57		
4.1	Introduction.....	57
4.2	Banc de test .....	57
4.2.1	Système et processus d'acquisition.....	58
4.2.2	Composants mécaniques : .....	60
4.2.3	Fixation des capteurs.....	61
4.2.4	Enregistrement des données.....	63
4.3	Analyse des Résultats .....	65
4.3.1	Régression.....	67
4.3.2	Classification.....	70

4.3.3	Clustering.....	72
4.3.4	Détection des anomalies et leurs classifications .....	75
	4.3.4.1 Détection des anomalies .....	75
	4.3.4.2 Classification de l'anomalie.....	79
4.4	Analyse des résultats.....	83
4.5	Conclusion .....	87
	CONCLUSION.....	89
	LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....	91



## LISTE DES TABLEAUX

	Page	
Tableau 3.1	Caractéristiques des données .....	22
Tableau 3.2	Abréviation des probabilités .....	35
Tableau 3.3	Résultats de régression des roulements 1 et 2 (Suite).....	37
Tableau 3.4	Résultats de régression des roulements 3 et 4.....	38
Tableau 3.5	Résultats de classification avec les données des PCA des roulements 1 et 2.....	41
Tableau 3.6	Résultats de classification avec les données des PCA .....	41
Tableau 3.7	Résultats de classification avec les caractéristiques statiques .....	42
Tableau 3.8	Résultats de classification avec les caractéristiques statiques .....	43
Tableau 3.9	Résultats de clustering avec les données des PCA des roulements 1 et 2.....	46
Tableau 3.10	Résultats de clustering avec les données des PCA des roulements 3 et 4.....	47
Tableau 3.11	Résultats de clustering avec les caractéristiques statiques.....	47
Tableau 3.12	Résultats de clustering avec les caractéristiques statiques.....	48
Tableau 3.13	Tableau des défauts des roulements.....	51
Tableau 3.14	Performance du modèle Décision trees.....	53
Tableau 3.15	Performance du modèle Gaussian NB .....	53
Tableau 3.16	Performance du modèle Gaussian NB .....	55
Tableau 3.17	Performance du modèle Extra trees .....	56
Tableau 4.1	Caractéristiques des capteurs MEMS .....	58
Tableau 4.2	Côtes géométriques des roulements.....	61
Tableau 4.3	Résultats de régression avec les caractéristiques statiques .....	67
Tableau 4.4	Résultats de classification avec les caractéristiques statiques (Suite) .....	70

Tableau 4.5	Résultats de clustering avec les caractéristiques statiques.....	73
Tableau 4.6	Performance des modèles pour Adxl335 .....	80
Tableau 4.7	Performance des modèles pour Adxl1002Z (Suite).....	81

## LISTE DES FIGURES

	Page
Figure 2.1	Les différents types de maintenance.....
Figure 2.2	Défauts des moteurs à induction.....
Figure 2.3	Les six blocs fonctionnels de la référence .....
Figure 2.4	Les éléments des roulements .....
Figure 2.5	Principe de base de la deuxième .....
Figure 3.1	Illustration du montage des roulement.....
Figure 3.2	Composants principaux du Roulement 1 .....
Figure 3.3	Flux des données de l'Algorithme Super Learner.....
Figure 3.4	Bibliographie des algorithmes .....
Figure 3.5	Classification des données .....
Figure 3.6	Chevauchement des classes .....
Figure 3.7	Décomposition des signaux en trois phases.....
Figure 3.8	Anomalies détectées dans le roulement N°4.....
Figure 3.9	Matrice de confusion du Décision Trees .....
Figure 3.10	Matrice de confusion du Gaussian NB .....
Figure 3.11	Matrice de confusion du Gaussian NB .....
Figure 3.12	Matrice de confusion.....
Figure 4.1	Moteur DC ABB Drives .....
Figure 4.2	a) Adxl 335, b) Adxl1002Z.....
Figure 4.3	a) Arduino Uno, b) Raspberry Pi 4 .....
Figure 4.4	Montage de test.....
Figure 4.5	a) Roulement de type BC, b) Les dimensions de roulements .....

Figure 4.6	Colle résine époxyde.....	61
Figure 4.7	Fixation des capteurs.....	62
Figure 4.8	Effet de la colle époxy sur la bande passante .....	62
Figure 4.9	Installation du montage pour la collecte .....	63
Figure 4.10	Perçage de l'élément.....	64
Figure 4.11	Signaux de vibrations enregistrés avec l'accéléromètre Adxl 335.....	64
Figure 4.12	Signaux de vibrations enregistrés avec l'accéléromètre Adxl 1002Z .....	65
Figure 4.13	RMS des données des deux tests enregistrés par le capteur Adxl 335 .....	66
Figure 4.14	RMS des données des deux tests enregistrés par le capteur Adxl 1002Z .....	66
Figure 4.15	Prédiction du RUL par le modèle XG Boost .....	68
Figure 4.16	Prédiction du RUL par le modèle SVM.....	69
Figure 4.17	Prédiction du RUL par le modèle XG Boost .....	69
Figure 4.18	Prédiction du RUL par le modèle SVM.....	70
Figure 4.19	Matrice de confusion.....	71
Figure 4.20	Matrice de confusion.....	72
Figure 4.21	Matrice de confusion.....	73
Figure 4.22	Matrice de confusion du KNN .....	74
Figure 4.23	Clustering des données de l'Adxl 335 des deux tests .....	74
Figure 4.24	Clustering des données de l'Adxl 1002Z des deux tests.....	75
Figure 4.25	Structure de l'AE .....	76
Figure 4.26	Courbe de la perte d'entraînement et de validation .....	77
Figure 4.27	Distribution de la perte.....	78
Figure 4.28	Détection d'anomalie Test1.....	78
Figure 4.29	Détection d'anomalies Test2 .....	79

Figure 4.30	Matrice de confusion.....	80
Figure 4.31	Matrice de confusion.....	81
Figure 4.32	Matrice de confusion.....	82
Figure 4.33	Matrice de confusion de l'SVM .....	83
Figure 4.34	Variation du R2 score des deux roulements .....	83
Figure 4.35	Variation des erreurs des algorithmes.....	84
Figure 4.36	Variation du F-score de classification.....	85
Figure 4.37	Variation du F-score de clusering.....	86
Figure 4.38	Variation du F-score de clasification des anomalies.....	87



## **LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES**

RUL	Remaining Useful Life
MEMS	Micro Electro Mechanical System
ML	Machine Learning
Pdm	Prédiktive maintenance
IA	Intelligence Artificielle
CBM	Condition Based Maintenance
CM	Maintenance Corrective
PM	Maintenance Préventif
IOT	Internet Of Things
CPS	Cyber Physical Systems
OSA-CBM	Open System Architecture for Condition-Based Monitoring



## INTRODUCTION

L'industrie 4.0 consiste à améliorer la productivité, établir l'intercommunication des machines à travers les technologies IT et la digitalisation des processus de production, tout en gardant la qualité des produits et en respectant les délais imposés par les clients. La disponibilité des équipements joue un rôle crucial dans le maintien des processus de production, l'absence d'un élément peut provoquer des arrêts de production indésirables qui ont un impact négatif direct sur les coûts des retards de fabrication. Au cours des dernières années, le concept de la maintenance 4.0, qui dérive de la quatrième révolution industrielle, est né dans la conférence internationale des technologies industriels de Hanover en 2011. Elle consiste à la surveillance des machines en temps réel en se basant sur des anciens données pour prendre des décisions (Nissoul et al., 2020). Cet énorme développement technologique qui ont subi les processus de production industriel, intégrant les capteurs micro électromécanique MEMS, les machines modernes, le Big Data, l'utilisation des systèmes cyber-physiques CPS et l'Internet des Objets IOT ont été considérer parmi les éléments nécessaires dans les nouvelles approches de maintenance 4.0 (Tran et al., 2021). Cependant, la complexité de ces nouveaux systèmes rend la supervision et la maintenance traditionnelles des composants des chaînes de production irréalisable dans les délais et à moindre coût pour les entreprises (Tiddens', 2018).

Le mémoire est structuré comme suit : Au chapitre 1, nous définirons la problématique de recherche et les objectifs que nous ciblerons. Le chapitre 2 est une revue de littérature qui met en évidence les travaux d'autres chercheurs et présente un résumé de l'état de l'art. Le chapitre 3 détaille les outils utilisés, l'évaluation des modèles d'IA pour la détection et la classification des anomalies, et l'estimation de l'état de santé des roulements. Le chapitre 4 décrit le dispositif de test et le choix des composants et interprète les résultats obtenus. Une conclusion résume le travail effectué.



## CHAPITRE 1

### OBJECTIFS ET MÉTHODOLOGIE DE RECHERCHE

#### 1.1 Introduction

L'un des grands défis dans les industries modernes est l'optimisation des stratégies de maintenance pour augmenter la disponibilité des équipements, diminuer les temps d'arrêts de fabrication tout en minimisant les coûts. La question qui se pose, Commet peut-on réaliser un système de prédition des pannes des machines industrielles à travers les nouvelles technologies et avec un moindre coût. Ce projet de recherche propose une étude sur l'utilisation des capteurs Micro Électromécaniques MEMS en combinaison avec des modèles d'intelligences artificielles pour la détection et la classification des pannes, l'évaluation des performances de ces modèles ainsi que la performance des MEMS à faible coût dans le contexte de maintenance prédictive.

#### 1.2 Contexte et Problématique de recherche :

La maintenance est une activité essentielle dans les industries, son impact est très éloquent sur la compétitivité des entreprises dans le marché, elle exerce une grande influence sur la qualité des produits, le respect des délais de préparation des commandes, ainsi que sur les coûts de production. Suite à l'évolution technologique mondiale et l'apparition de nouvelles techniques et technologies (MEMS, IOT, IA, CPS...), les stratégies de maintenance industrielles ont subi une transition de la maintenance corrective CM à la maintenance préventif PM et enfin à la maintenance prédictive PdM, cette dernière est basée sur la prédition des pannes en temps réel avant leur apparition, et ça sert à diminuer la fréquence des interventions inutiles. Selon l'enquête réalisée par (Ran et al., 2019), les temps d'arrêt imprévus causés par une mauvaise stratégie de maintenance réduisent la capacité de production globale d'une usine jusqu'à 20 % et coûtent environ 50 milliards de dollars par an. Parmi les éléments les plus utilisés dans les sites industriels on trouve les moteurs électriques où l'on constate que plus de 90% des industries utilisent ces moteurs (Kavana & Neethi, 2018), puisqu'ils ont une large gamme de

puissances nominales. Mais malgré leur flexibilité et robustesse, ils sont sujet à de nombreuses pannes catastrophiques notamment au niveau des roulements. Identifier ces défauts au plus tôt est très important, sinon ils entraînent d'énormes pertes de production et des dégâts financières. Le problème qui surgit est que la maintenance prédictive ou la surveillance de l'état de l'équipement à la base de sa condition sont très coûteuses à implémenter pour les industries, pour cela notre objectif de recherche est d'étudier la possibilité de concevoir un système de maintenance prédictive à base des capteurs MEMS les moins chère dans le marché en combinaison avec les modèles prédictifs d'Intelligence artificielle, pour la surveillance et la détection des pannes des roulements avant son apparition, également la classification de ces dernières selon leurs types.

### 1.3 Objectifs :

La maintenance prédictive est la stratégie la plus adapté par les industries pendant les dernières années bien qu'elles ne soient pas totalement satisfaites des solutions de PdM (Ag, 2018), où qu'elles installent un système de maintenance afin de prédire les problèmes avant leur apparition. Dans ce projet de recherche, on vise l'utilisation des accéléromètres MEMS afin de mesurer la vibration de l'équipement. Dans un environnement industriel, la classification des défauts est une tâche qui permettra de réduire les pannes futures en organisant des opérations de maintenance uniquement lorsque la situation l'impose. Ainsi, les principaux objectifs de notre travail de recherche sont :

- Évaluation des modèles d'intelligence artificiels pour la prédiction de l'état de santé des roulements et la classification des anomalies qui occurrent.
- Estimation de la durée de vie utile (RUL) des roulements.
- Évaluation de la performance des accéléromètres MEMS pour la détection et la classification des pannes des roulements en combinaison avec les modèles IA.

#### **1.4 Contribution :**

La première partie vise à évaluer la performance des modèles prédictifs pour la détection des défauts des séquences temporelles et leurs classifications selon leurs types, ainsi que l'évaluation de l'état de l'équipement. Les tests de cette étape seront exécutés sur une base de données enregistrée par le NSF I/UCR Centre des systèmes de maintenance intelligents avec le soutien de Rexnord Corp. à Milwaukee, WI.

La deuxième partie de notre projet de recherche consiste à concevoir un montage composé d'un moteur électrique, d'un accouplement, d'un arbre de transmission, de roulements et d'accéléromètres pour l'enregistrement des données de vibration, ainsi que d'un microcontrôleur pour l'échange des informations avec l'unité de calcul. En se basant sur les résultats de la première partie, nous testerons les modèles prédictifs avec nos propres données dans le but d'évaluer la performance des accéléromètres les moins chère du marché et d'examiner leurs capacités à fournir des données fiables pour les utiliser dans un système de maintenance prédictive.



## **CHAPITRE 2**

### **REVUE DE LITTÉRATURE**

#### **2.1        Introduction**

La maintenance industrielle joue un rôle primordial dans la réduction des coûts de réparation et d'intervention dissipés par les entreprises sur les biens, dans ce chapitre on va discuter le développement de cette dernière en mettant les travaux qui ont été réalisés par la communauté des chercheurs pour la contribution à l'évolution de la maintenance pour bien cadrer notre projet de recherche.

#### **2.2        Maintenance 4.0**

Selon La norme AFNOR NF X 60 010 la maintenance est « l'ensemble des actions permettant de maintenir ou de rétablir un bien dans un état spécifié ou en mesure d'assurer un service déterminé ». Il existe trois principaux types de maintenance : la maintenance corrective ou réactive, la maintenance préventive et la maintenance prédictive, la première est établie lorsque la panne apparaît, la seconde est mise en place en pour éviter l'apparition de la panne et la troisième est utilisée à fin de prédire la défaillance avant qu'elle n'apparaisse (Pr, n.d.). De nos jours, les systèmes de maintenance prédictive est devenu plus populaire grâce à la quatrième révolution industrielle tel que la maintenance à base de l'état de l'équipement CBM (Condition Based Maintenance), où la surveillance des biens est faite en temps réel pour prédire la panne avant qu'elle ne persiste. La figure 2.1 nous montre la différence entre ces trois types.

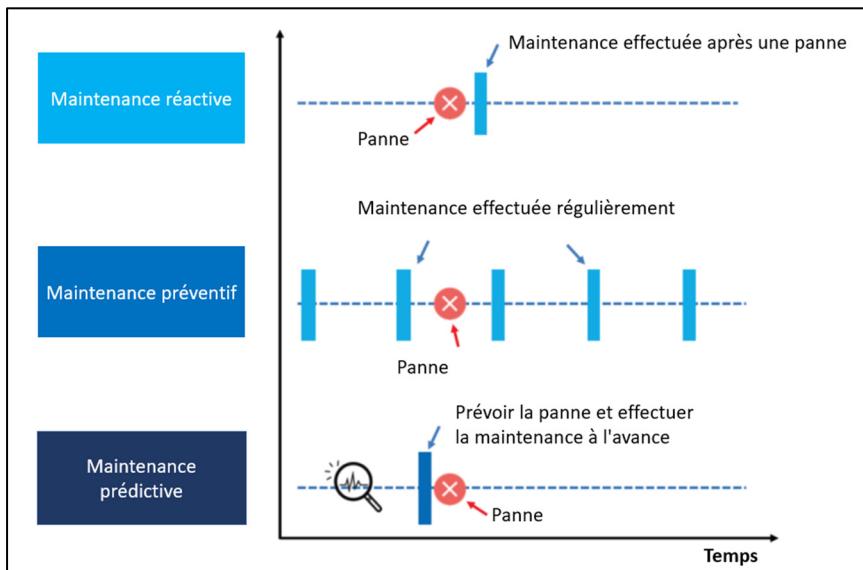


Figure 2.1 Les différents types de maintenance  
Tirée de Ran (2019, p.5)

Cependant, l'utilisation des circuits intégrés et des micro-capteurs électro mécaniques ont permis de digitaliser la maintenance, d'interconnecter les machines, de visualiser plusieurs paramètres physiques et de faciliter l'intégration des termes tels que le Big Data et l'accès aux données en temps réel.

### 2.3 Stratégie de la maintenance prédictive

La maintenance prédictive se décompose en plusieurs modules : D'abord on trouve l'unité d'acquisition et de traitement des données, ensuite on choisit les indicateurs de santé que l'on va utiliser pour la surveillance, puis on passe à la détection de début de l'anomalie afin de pouvoir prédire la nature de la panne et prendre les actions correctives nécessaires (Lughofer & Sayed-Mouchaweh, 2019).

#### 2.3.1 Acquisition et traitement des données

L'acquisition des données consiste à enregistrer les valeurs des paramètres physiques générés par des capteurs placés sur les machines tels que les vibrations, la température, l'état de surface, la variation du courant...etc. De même, l'historique des évènements de réparation et des

interventions sur les machines peut contribuer de manière efficace à la prédition des pannes des équipements. Par la suite le traitement des données enregistrées vise à extraire les caractéristiques et les indicateurs de défaillance.

### **2.3.2 Détection et localisation de la panne**

Cette phase est la plus importante, où il faut détecter tout changement dans les données enregistrées en temps réel, qui ça se peut le début de la défaillance, après cela on peut localiser et identifier le type de la défaillance afin d'intervenir au bon moment et au bon endroit.

### **2.3.3 Pronostics et prévisions**

Dans cette phase, on va parler du pronostic des évènements susceptibles de se produire dans le futur. La précision et la qualité du pronostic et des prévisions jouent un rôle essentiel dans le modèle de maintenance prédictive pour donner un temps de réaction suffisamment long pour intervenir significativement avant l'endommagement de l'un des composants de la machine. Généralement, les modèles de pronostic et de prévisions se basent sur des indicateurs de l'état de santé de la machine et sur des indicateurs d'usure des pièces afin de nous garantir des anticipations fiables des anomalies. Par exemple l'estimation de la durée de vie utile d'un bien (RUL : Remaining Useful Life) est l'une des variantes spécifiques du pronostic et de la prévision.

### **2.3.4 Actions correctives par le diagnostic, l'optimisation et le contrôle**

La manifestation et la détection des défauts nécessitent la planification des actions correctives pour augmenter la disponibilité des processus et réduire les coûts de maintenance. La programmation de ces actions nécessite la vérification de la maintenabilité de l'intervention, la disponibilité des composants et la gravité de la défaillance, où on peut classés selon leur gravité et leur criticité pour qu'on puisse choisir les bons moments d'intervention.

## 2.4 Surveillance des moteurs électriques

Les moteurs électriques sont les types de moteurs les plus utilisés dans les machines et les équipements industriels tels que les convoyeurs, les compresseurs, les presses, etc... et ils consomment plus que 60% de l'énergie électrique produite dédiée aux industries (Sayed-Mouchaweh, 2018), nonobstant le fait qu'ils sont beaucoup utilisés, ils sont exposés à plusieurs types de défauts qui peuvent être catastrophiques s'ils ne sont pas détecter dès les premiers stade, par conséquent les moteurs peuvent être endommagés, il y aura des arrêts de production

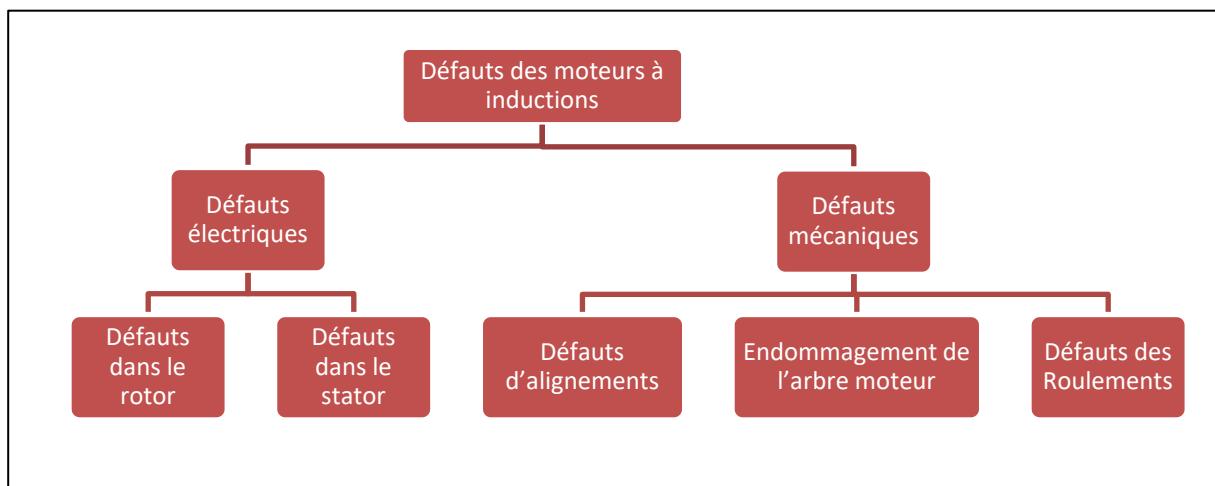


Figure 2.2 Défauts des moteurs à induction

non planifié et des coûts de maintenance élevés, les principaux défauts des moteur à induction sont décomposés en deux catégories : les défauts mécaniques et les défauts électrique comme il est montré dans la figure 2.2, avec une distribution de 69% des pannes dans les roulements, 21% dans les bobines de stator, 7% au niveau de l'arbre du rotor et 3% dans le couplement du moteur (Wen, 2011).

## 2.5 Surveillance de l'état des machines

La surveillance de l'état de santé des équipements d'une machine industrielle est indispensable, la défaillance d'un seul composant peut entraîner un arrêt de production et de longs délais pour effectuer le diagnostic et la détection de l'emplacement du défaut, ces arrêts de production

entraînent des dégâts financiers dû au retard qu'ils occasionnent et aux frais supplémentaires pour la réparation. Durant les dernières années, la maintenance prédictive est devenue la stratégie la plus utilisée dans les industries et elle représente un composant essentiel dans l'industrie 4.0 (Murphy, n.d.), en particulier dans les machines tournantes où les moteurs, les roulements, les engrenages et d'autres composants sont exposés à des pannes sur une base quotidienne, par conséquence la maintenance à la base de la condition des machines représente la méthode la plus adéquate à cette situation. Cette technique nous aide à prendre les décisions de maintenance selon le changement des paramètres physiques tels que la vibration, la température et le courant électrique. La vibration représente le principal paramètre qui donne des informations exactes sur les conditions des équipements, la CBM offre beaucoup des avantages aux industries : l'amélioration de l'efficacité opérationnelle, la bonne planification de la maintenance aux bons moments et la réduction des ces coûts, ainsi que l'augmentation de la disponibilité des équipements et de la durée de vie des biens. Néanmoins elle à des points négatifs qui affectent la décision des entreprises à intégrer ce type de maintenance y inclus le coût des équipements nécessaires, l'installation et la collecte des données qui prend beaucoup de temps, la formation des employées...etc. (Ali et al., 2019). Notamment, des architectures ont été proposées comme des références dans le but de standardiser les étapes de la maintenance conditionnelle. Dans notre projet de recherche, la référence Open System Architecture for Condition-Based Monitoring (OSA-CBM) est adapté comme modèle conceptuel. L'OSA-CBM est une architecture normée par l'organisation internationale de normalisation ISO sous le standard ISO-13374 (ISO 13374, 2019). OSA-CBM simplifie le processus d'intégration d'une large variété de composants logiciels et matériels en mettant une architecture standard et un cadre pour la mise en place de maintenance conditionnelle. Il décrit les six blocs fonctionnels des systèmes CBM illustrés à la figure 2.3.

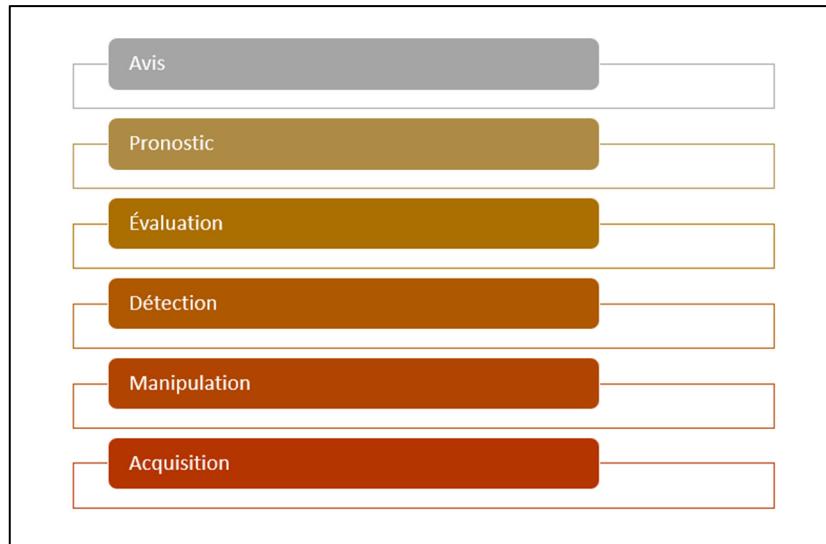


Figure 2.3 Les six blocs fonctionnels de la référence  
OSA-CBM

Les principales fonctions du premier bloc « Acquisition » représentent le processus de collecte et de stockage des données de l'installation en réparation. Les données entrantes sont principalement divisées en deux catégories : les données d'événements et les données de surveillance de l'état. Les données d'événements décrivent ce qui est arrivé à une installation, comme les défauts, les réparations et les détériorations. Ces données sont souvent saisies manuellement, qui prédispose à l'influence du facteur subjectif. Les mesures liées à l'état actuel de l'installation sont incluses dans les données de surveillance à partir des observations de l'état de l'installation. Ces informations sont généralement saisies automatiquement, mais dans certaines situations, elles sont saisies manuellement dans le système d'information CBM, ce qui peut entraîner des imprécisions subjectives (Campos et al., 2020). Le deuxième bloc « Manipulation » sert à transformer les données et extraire les caractéristiques utiles pour la surveillance de l'état du bien. Ensuite en comparant les données actuelles avec les données historiques on peut déterminer l'état de système, cette étape définit le bloc « Détection ». En se basant sur l'état du système dans le bloc « Évaluation », on peut identifier le régime de fonctionnement (Régime normal, en dégradation, en défaut, etc.). En outre, on trouve le bloc « Pronostic » qui est responsable de produire une prévision sur l'évolution temporelle de l'état opérationnel des machines. Au cours du bloc « Avis », des décisions et des recommandations seront prises pour mettre en place les bonnes activités de maintenance (Cachada et al., 2018).

## 2.6 Roulements mécaniques

« Le roulement est un organe mécanique chargé de permettre la rotation entre deux arbres ou entre un arbre et un logement dans de bonnes conditions de guidage et avec un minimum de déperdition énergétique. La rotation est permise par des corps roulants (1) séparés par une cage, qui roulent sur les bagues intérieure (2) et extérieure (3) de la figure 2.4 » (Djebili, 2013). Ils permettent aussi de réduire les frottements mécaniques et supportent les charges radiales et axiales.

Pendant les dernières années les roulements étaient les éléments les plus investigus dans la maintenance dans les industries à cause de nombre élevé de pannes qu'ils provoquent. Environ 69 % des pannes sont des pannes de roulements. Ce type de défaut peut affecter non seulement affecter les roulements, mais aussi d'autres composants voisins qui peuvent établir des dégâts catastrophiques (Wen, 2011).

Les roulements sont composés de quatre éléments principaux : Une bague extérieure, une bague intérieure, des éléments tournants supportés par une cage, comme la montre dans la figure 2.4 suivante :

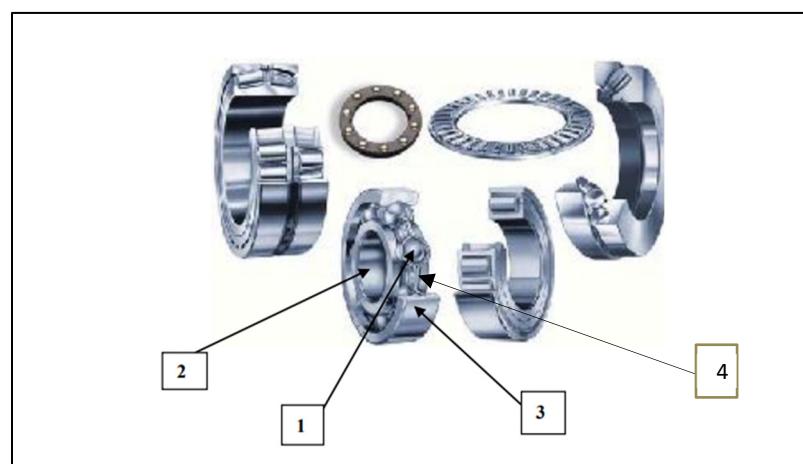


Figure 2.4 Les éléments des roulements mécaniques  
tirée de Djebili (2013, p. 49)

Dans l'ensemble, les roulements ont un faible frottement au début du fonctionnement et la température de rotation ne les affecte pas, ainsi avec le bon établissement de la maintenance et l'entretien des roulements, on peut éviter la majorité des. Une défaillance de roulement peut être causée par plusieurs facteurs tels que la fatigue, un manque ou trop de lubrifiant ou de la corrosion. Nous allons détailler ces mécanismes et leurs effets.

### **2.6.1 Les défauts des roulements**

En se basant sur le travail de (Aliwan, 2016) on va présenter des défauts fréquent des roulements.

#### **2.6.1.1 Fatigue**

Au bout d'un certain temps, tout roulement soumis à des charges alternées peut céder en raison de la fatigue du matériau. L'apparition de petites fissures sous la surface du roulement est le premier signe d'un endommagement par fatigue. Les fissures s'étendent à la surface au fur et à mesure, entraînant la rupture du matériau dans les zones de contact. La présence de piqûres ou d'écaillles sur les bagues du roulement ou sur les parties roulantes peut indiquer une véritable défaillance. Si le roulement est utilisé, la détérioration s'accélère, notamment dans la zone du défaut en raison de la concentration des contraintes. En introduisant des chocs de courte durée répétés aux fréquences caractéristiques des défauts des éléments roulants, la surface endommagée perturbe le mouvement nominal des éléments roulants.

#### **2.6.1.2 Manque ou trop de lubrification :**

Une lubrification inadaptée est l'une des principales causes de défaillance prématuée des roulements, provoquant le dérapage, le glissement et la rupture des roulements. Elle accélère le processus d'usure des adhésifs. Une mauvaise lubrification augmente aussi la température des composants du roulement, ce qui accélère le processus de détérioration.

### 2.6.1.3 Corrosion

Lorsque l'eau ou d'autres impuretés présentes dans l'huile atteignent les composants du roulement, la corrosion se produit. Des joints endommagés, des lubrifiants acides ou la condensation qui se produit lorsque les roulements sont rapidement refroidis à partir d'une température de fonctionnement plus élevée dans un air humide peuvent tous en être la cause. Comme les particules de rouille interfèrent avec la lubrification, cela provoque de la rouille sur les surfaces de roulement, ce qui entraîne un fonctionnement irrégulier et bruyant. Les particules de rouille ont également un impact abrasif, entraînant l'usure des roulements.

### 2.6.2 Systèmes de prédictions des défauts des roulements

On se rappellent que l'objectif de notre travail est de développer un système de détection et de classification des défauts de roulements à l'aide des modèles prédictifs en se basant sur des données de vibration enregistrées par des MEMS accéléromètres. La modélisation prédictive est un processus mathématique utilisé pour prédire des événements ou des résultats futurs en analysant des données d'entrée. Il s'agit d'un composant crucial de l'analyse prédictive, un type d'analyse qui utilise des données actuelles et historiques pour prévoir l'activité, le comportement et les tendances. Les modèles prédictifs sont décomposés en deux parties, une pour la détection des caractéristiques et l'autre pour la classification des données. Dans les travaux de recherche les défauts des roulements ont été identifiés et classés à l'aide de diverses techniques. Les approches de diagnostic des défauts basées sur des modèles et sur des données sont les deux types d'approches les plus utilisées (Zhang et al., 2021).

Les techniques basées sur des modèles visent à comprendre et à prévoir le comportement réel d'un système en modélisant mathématiquement le processus physique. L'analyse vibratoire utilisant les fréquences de défaillance des roulements est un exemple de technique basée sur un modèle. Ces stratégies nécessitent une expertise professionnelle ainsi qu'une compréhension approfondie du système.

Les approches basées sur les données ont fait l'objet de nombreuses recherches au fil du temps. Les modèles de séries temporelles tels que les modèles ARMA (modèles autorégressifs et moyenne mobile), le Kurtosis spectral et la décomposition en valeur singulière sont des techniques traditionnelles du domaine temporel (Meng et al., 2022). En raison du large éventail de conditions opérationnelles, la majorité de ces procédures sont spécifiques à une application qu'elles ne peuvent être généralisées à plusieurs types de défaillances à la fois. En contrepartie, les technologies basées sur l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning) sont de plus en plus populaires pour répondre à ce désir d'une plus grande performance dans une variété de circonstances opérationnelles. Les méthodes de réduction de la dimensionnalité telles que l'analyse en composantes principales (PCA) (Buchaiah & Shakya, 2022) ou la décomposition en valeurs singulières (SVD) sont utilisées comme extracteurs de caractéristiques dans l'apprentissage automatique. Ensuite, pour classer les types de défaillance, des techniques telles que les Random Forests (Roy et al., 2020), les KNearest Neighbours (Sharma et al., 2018) et les Support Vector Machines (Han et al., 2021) sont largement utilisées. Les réseaux neuronaux convolutifs (purwahyudi, 2018) se sont imposés dans le domaine de Deep Learning en raison de la puissance des couches d'extraction de caractéristiques. La durée d'utilisation restante des roulements a également été prédite à l'aide de réseaux neuronaux récurrents RNN et de réseaux de neurones à mémoire longue et à court terme LSTM (Long Short-Term Memory Neural Networks) (Qiao et al., 2020).

## 2.7 Accéléromètres MEMS :

Un accéléromètre est un capteur mécanique qui surveille les accélérations constantes (gravité), variables dans le temps (vibrations) et quasi-statiques (inclinaison). Les accéléromètres MEMS sont classés comme piézo-résistifs, piézoélectriques, capacitifs, à effet tunnel, résonants, électromagnétiques, thermocouples, optiques, inductifs et autres en fonction de leurs principes de fonctionnement.

### 2.7.1 Principe de fonctionnement :

D'une manière simple, un accéléromètre MEMS peut être considéré comme un processus basé sur la deuxième loi de Newton dans lequel une masse est attachée à un ressort dans un cadre de référence figure 2.5. Les accéléromètres MEMS fonctionnent selon deux principes principaux : l'un mesure le déplacement de la masse et l'autre mesure la fréquence d'un élément vibrant (masse) à la suite de la variation de tension. L'accéléromètre mesure l'accélération linéaire, et nous pouvons calculer la position en intégrant deux fois le signal enregistré. (Arun Faisal et al., 2019).

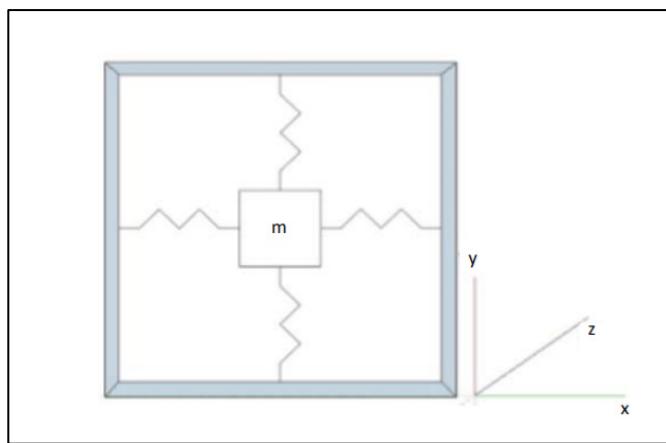


Figure 2.5 Principe de base de la deuxième  
Loi de Newton  
Tirée de Arun Faisal (2019, p. 827)

### 2.7.2 Utilisations des Accéléromètres MEMS

Pendant les dernières années, l'utilisation des capteurs accéléromètres MEMS a augmenté dans les applications d'ingénierie grâce à leurs petites tailles et leurs prix peu coûteux par rapport aux accéléromètres piézoélectrique. De plus ils sont plus performants en travaillant dans les bases fréquences. Ces micro-capteurs ont trouvé beaucoup de succès dans plusieurs applications tels que : la récupération d'énergie (Varanis et al., 2018), la surveillance des systèmes (Korkua et al., 2010), système de navigation inertiel (Kavitha et al., 2016), ainsi comme le travail présenté par (Villacorta et al., 2021) sur le contrôle de santé intégrée. D'autre

part (Wang et al., 2021) ont proposer une méthode de détection des défauts des roulements à base d'un modèle de deep learning CNN en utilisant des données des signaux de vibrations, cette méthode a été validé au début sur des bases des données scientifiques présentés par des centres de recherche et par la suite par ses propres bases des données enregistrée par des capteurs MEMS accéléromètres ADXL335. D'autre part, (Kumar & Hati, 2021) ont développés une autre méthode de détection des anomalies des roulements, c'est un optimiseur de gradient adaptatif basé sur les neurones à convolution profonde ADG-dCNN, qui fait l'extraction automatique des descripteurs afin de les utiliser pour la classification des défauts. Ces deux chercheurs ont utilisé une base des données enregistrée aussi par des capteurs MEMS ADXL326 donnant une précision de classification 99,70%. Aussi (Larranaga et al., 2020) ont fait une recherche sur les microcontrôleurs et les capteurs MEMS accéléromètres les moins chères dans le marché qu'on peut les utiliser pour l'acquisition des données afin de les exploiter dans un contexte de maintenance prédictive, à la fin ils ont choisi le Raspberry Pi3 comme un microcontrôleurs avec un accéléromètre ADXL1002Z et un convertisseurs ADC (Analogique-Digital) ADS131A04, à la fin ce système a arriver à détecter les fréquences de défauts de roulement mais le montage a rencontré des problèmes avec le Raspberry pour l'échantillonnage des fréquences égales ou supérieures à 16 kHz, ça nécessite plus de ressources de calcule et d'exécution. De surcroît, (Adli & Rusmin, 2020) dans son travail ont utilisé l'accéléromètre ADXL345 pour l'enregistrement des données de vibration d'une machine rotatif, par la suite ces données ont été transformées sous la forme de mm/s pour qu'ils puissent se référer au tableau de limite de vibration selon la norme ISO standard 10816-1. Ce capteur a montré sa capacité de détection des caractéristiques vibratoire de la machine, ainsi que son état de fonctionnement.

## 2.8 Conclusion

Ce chapitre donne un bref aperçu sur l'évolution de la maintenance au sein des industries en arrivant à la maintenance 4.0. La surveillance des moteurs électriques est discutée, arrivant aux défauts liés aux roulements et de leur diagnostic puisqu'ils présentent le pourcentage le plus

élevé des défauts de ce types des moteurs. Enfin, une présentation des capteurs MEMS accéléromètres a été faite, qui seront utilisés par la suite dans notre projet de recherche.



## CHAPITRE 3

### IMPLÉMENTATION DES MODÈLES D'INTELLIGENCE ARTIFICIELLE

#### 3.1 Introduction

Ce chapitre a pour but de fournir une description des modèles qui seront utilisés dans le projet, l'approches d'intelligence artificielle utilisée est la Machine Learning. Ces modèles seront évalués dans un premier temps avec une base des données open source qui sera décrite dans ce chapitre.

#### 3.2 Base des données

Les signaux de vibration utilisés dans la première partie de notre travail sont fournis par le Centre pour les systèmes de maintenance intelligents (IMS). Les ensembles de données expérimentales sont générés à partir d'essais de fonctionnement jusqu'à la défaillance de roulements avec des conditions de charge constante sur un banc d'essai spécialement conçu, comme le montre la figure 3.1. Le banc d'essai des roulements accueille quatre roulements de type Rexnord ZA-2115 à double rangée de billes montées sur un arbre. L'arbre est entraîné par un moteur à courant alternatif et couplé par des courroies de transmission. Une charge radiale de 6 000 livres est appliquée sur l'arbre et au roulement par un mécanisme à ressort. La vitesse de rotation est maintenue constante à 2000 tr/min. Un accéléromètre PCB 353B33 à quartz ICP à haute sensibilité est installé sur chaque logement de roulement. La collection des données est effectuée à l'aide d'un programme LabVIEW de National Instruments de National Instruments. Le taux d'échantillonnage des données est de 20 kHz.

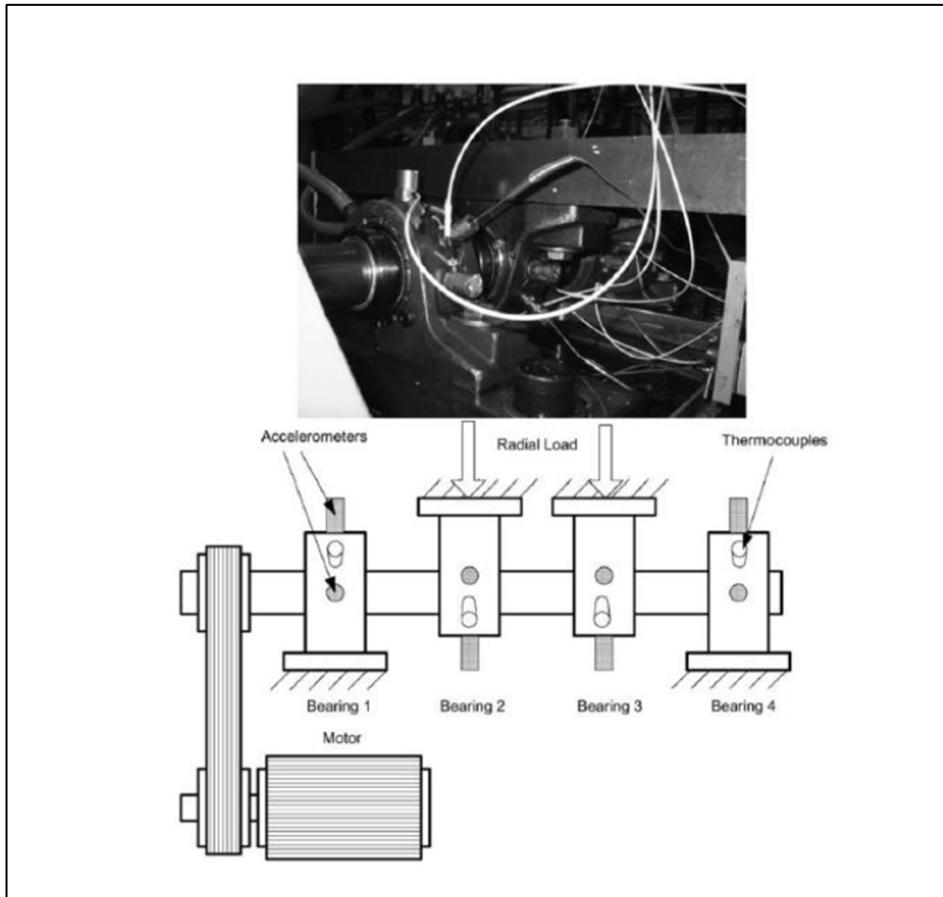


Figure 3.1 Illustration du montage des roulement et l'emplacement des Accéléromètres

Le tableau 3.1 représente les propriétés des données enregistrés dans chaque expérience.

Tableau 3.1 Caractéristiques des données

Test	Nombres d'échantillons	Taille d'échantillon	Défaillance des roulements	Durée a la défaillance
1	2156	20480	Roulements 3 & 4	35 jours
2	984	20480	Roulements 1	8 jours
3	6324	20480	Roulements 3	31 jours

L'accéléromètre utilisé pendant les tests est un accéléromètre piézoélectrique PCB353B33, sa fréquence de résonance est supérieure ou égale à 22 kHz. L'accéléromètre a été monté sur les roulements avec un support de goujon, le courbe dans la figure 3.2 nous montre l'effet des méthodes de montage sur la bande passante, ou la méthode utilisé qui est la première dans notre cas nous offre jusqu'à 60 kHz de bande passante ce qui est largement suffisant.

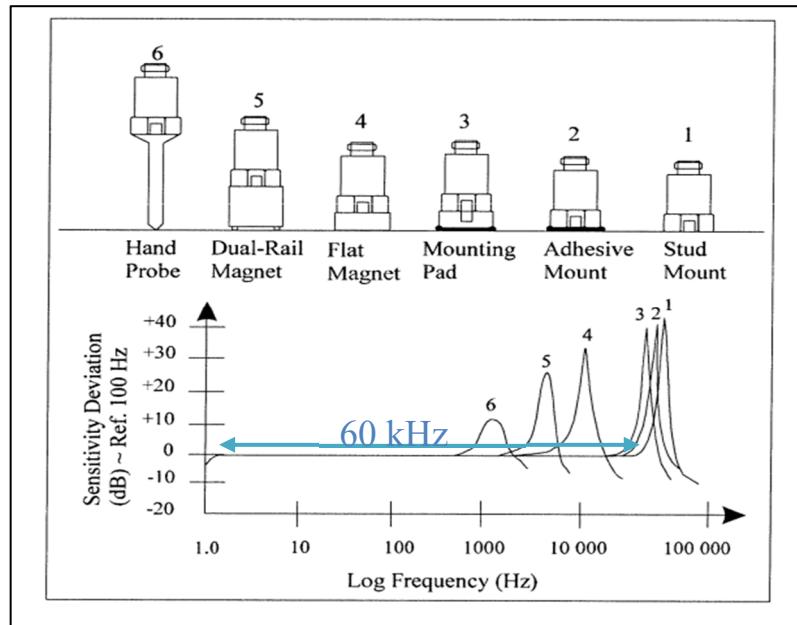


Figure 3.2 Effet des méthodes de montage sur la bande passante de l'accéléromètre

### 3.3 Traitement des données et extraction des descripteurs

Pour l'évaluation et la validation des modèles qu'on va les utiliser, nous avons sélectionné deux types de méthodes d'extraction des caractéristiques des données, l'une pour réduire les dimensions en raison de la taille énorme des données que nous avons et la seconde pour extraire des caractéristiques statiques dans le domaine temporel.

#### 3.3.1 Analyse des composants principaux PCA

Selon (Rehab et al., 2021), PCA convient généralement à la réduction de la dimensionnalité. Les données sont transposées dans un nouvel espace appelé espace propre à l'aide d'une

transformation linéaire orthogonale, de sorte que la variance maximale soit préservée dans les composantes primaires.

La PCA convertit  $u$  en  $v$  :  $v = Wu$  ( $u$  étant le vecteur d'observation). Les valeurs propres et les vecteurs propres sont calculés à l'aide de la décomposition d'une matrice en éléments propres. En permettant à la matrice de covariance des données d'être  $R_u$  :

$$R_u = U_u D_u U_u^T \quad (3.1)$$

Où  $U_u$  est la matrice des vecteurs propres de  $R_u$  et  $D_u$  est la matrice diagonale correspondante des valeurs propres. Par conséquent, si  $W = U_u^T$ , la transformation des données est réalisée par

$$v = Wu \quad (3.2)$$

La réduction de la dimension est obtenue en ordonnant les valeurs propres et en sélectionnant les  $p$  colonnes correspondantes de  $R_u$  où  $p < M$ .  $W$  est reconstruit en la matrice  $p \times M$  des  $p$  vecteurs de colonnes dominants de  $U_u$ .

L'application de cette méthode sur nos données nous a donné les résultats suivants :

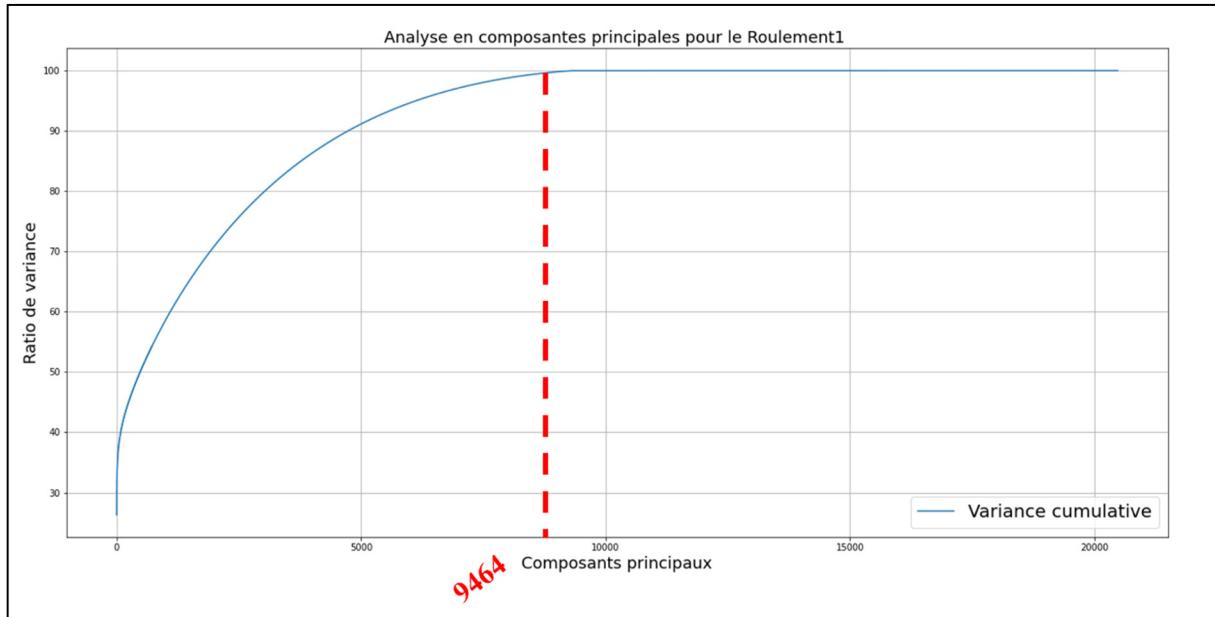


Figure 3.3 Composants principaux du Roulement 1

Le courbe de résultat obtenus nous montre qu'on peut représenter la totalité des données en utilisant les 9464 composants principaux seulement, qui donnent une variance de 100% ou une représentation totale des données.

### 3.3.2 Descripteurs statistiques dans le domaine temporel

Plusieurs caractéristiques uniques pouvant être extraites des signaux de vibration des roulements. Cette section donne un bref aperçu sur la manière de générer des caractéristiques originales à partir de domaine temporel, qui seront utilisés tout au long de notre travail.

#### 3.3.2.1 Valeur efficace RMS

La première caractéristique que nous avons calculée est la racine carrée moyenne calculée à partir de la moyenne. Il s'agit d'une caractéristique bien connue qui représente le signal en termes de son contenu énergétique. Elle est donnée par l'équation 3.3 :

$$\text{RMS} = \sqrt{\frac{1}{N} \times \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.3)$$

Où N est le nombre d'observation acquis,  $x_i$  est la valeur de chaque échantillon et  $\bar{x}$  est la moyenne des valeurs du signal.

#### 3.3.2.2 L'aplatissement

Kurtosis ou l'aplatissement est un signal statistique très sensible à l'irrégularité ou les chocs (Lamraoui et al., 2011) présentent dans les données. Cette caractéristique est donnée par l'équation 3.4 :

$$\text{Kurtosis} = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{(N-1) \times \sigma^4} \quad (3.4)$$

### 3.3.2.3 Valeur maximale et la valeur crête à crête

Les troisièmes et quatrièmes caractéristiques sont respectivement la valeur maximale de crête et la valeur de crête à crête, qui sont définies comme l'amplitude maximale et l'intervalle entre l'amplitude maximale et l'amplitude minimale, données par les équations suivantes :

$$\text{Max} = \max(x) \quad (3.5)$$

$$\text{PTP} = \text{Abs}(\max(x)) + \text{Abs}(\min(x)) \quad (3.6)$$

### 3.3.2.4 Valeur Moyenne

Cette caractéristique est donnée par l'équation suivante :

$$\text{Moyenne} = \frac{\sum_{i=1}^N X_i}{N} \quad (3.7)$$

### 3.3.2.5 Déviation standard ou l'écart type

En statistique, l'écart-type est une mesure de l'importance de la variation ou de la dispersion d'un ensemble de valeurs.

$$\text{STD} = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}} \quad (3.8)$$

### 3.3.2.6 Asymétrie

L'asymétrie est une mesure de l'asymétrie de la distribution d'une variable (X. Chen et al., 2019). C'est un outil statistique qui a l'avantage d'être indépendant de la moyenne des données, sa formule est donnée par l'équation suivante :

$$\text{Asymétrie} = \frac{\sum_i^N (X_i - \bar{X})^3}{(N-1) \times \sigma^3} \quad (3.9)$$

### 3.3.2.7 Entropie

L'entropie est un terme fréquemment utilisé dans la détection des défauts mécaniques. Elle est utilisée pour caractériser la complexité et les variations dynamiques des signaux de vibrations dans l'étude des signaux de vibrations mécaniques non linéaires et non stationnaires (Li et al., 2019). Où P est la probabilité.

$$\text{Entropie} = - \sum_{i=1}^N P(x_i) \log P(x_i) \quad (3.10)$$

## 3.4 Machine Learning

Cette section sera divisée en deux parties, l'une sera dédiée à l'estimation de l'état de santé des roulements et la deuxième pour la détection et la classification des défauts des roulements.

- **Partie 1 : État de santé des roulements**

### 3.4.1 Méthodologie : Super Learner

(Van Der Laan et al., 2007) ont proposé et évalué théoriquement le super apprentissage, qui est un système d'apprentissage générique basé sur les pertes. Nous allons examiner le super apprentissage pour la prédiction de l'état de santé des roulements. Cette dernière est une approche de prédiction qui utilise une combinaison de plusieurs algorithmes de prédiction pour déterminer le meilleur résultat. L'architecture du super learner est basée sur la théorie de la validation croisée ou ‘K-Fold Cross Validation’ et permet à l'ensemble d'explorer un large éventail d'algorithmes de prédiction.

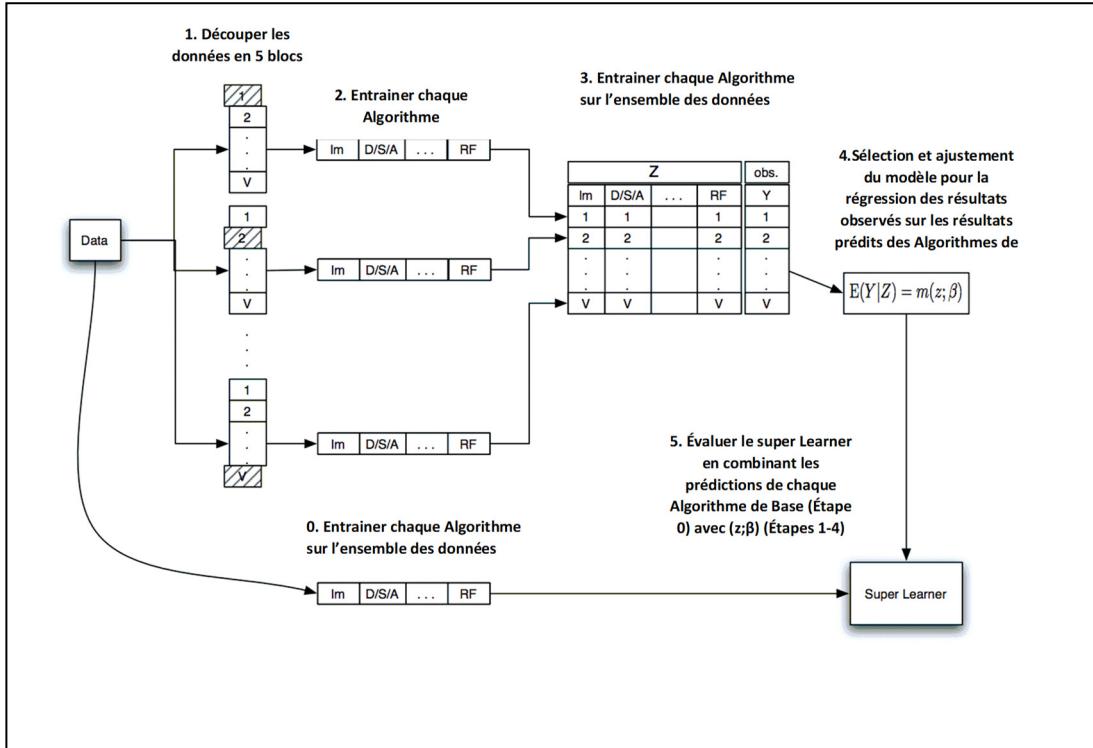


Figure 3.4 Flux des données de l'Algorithme Super Learner  
Tirée de Van Der Laan (2007, p. 6)

(Polley et al., 2010) ont investigu   le Super Learner pour les utilisations dans les probl  mes de pr  dition, en donnant des recommandations d'impl  mentation. Une biblioth  que d'algorithmes de pr  dition peut  tre propos  e pour un sujet donn  . Cette biblioth  que doit  tre bas  es sur des informations contextuelles et un grand nombre d'algorithmes par d  faut. Pour mieux comprendre son principe de fonctionnement, nous allons d  cortiquer ces  tapes comme indiqu   dans la figure 3.4 :

- La premi  re  tape consiste   entra  ner chaque algorithme de base sur la totalit   des donn  es  $\mathcal{X} = \{X_i : i = 1, \dots, n\}$
- Deuxi  mement, en utilisant l'approche de validation crois  e   5 niveaux, divisez l'ensemble de donn  es  $\mathcal{X}$  en un  chantillon d'entra  nement et un  chantillon de

validation : divisez les  $n$  observations ordonnées en 5 groupes de taille égale, le 5-ième groupe servant d'échantillon de validation et le groupe restant servant d'échantillons d'entraînement.

- Troisièmement, entraîner chaque algorithme de base et enregistrer les prédictions sur la base des données de validation.
- Quatrièmement, empiler les prédictions de chaque algorithme pour former une matrice  $n$  par  $K$ .
- À la fin, pour créer l'entraînement final du Super Learner, combinez les prédictions des algorithmes de base sur les données entière et toutes les autres prédictions par la méthode K-fold cross validation.

### **3.4.2 Bibliothèques des algorithmes :**

Plusieurs algorithmes de régression, classification et de clustering sont disponibles dans la boîte à outils de Python. Dans ce travail, plusieurs algorithmes seront utilisés pour évaluer leurs performances et leur adéquation au diagnostic des défauts des roulements. Les algorithmes utilisé sont affiché dans le diagramme suivant :

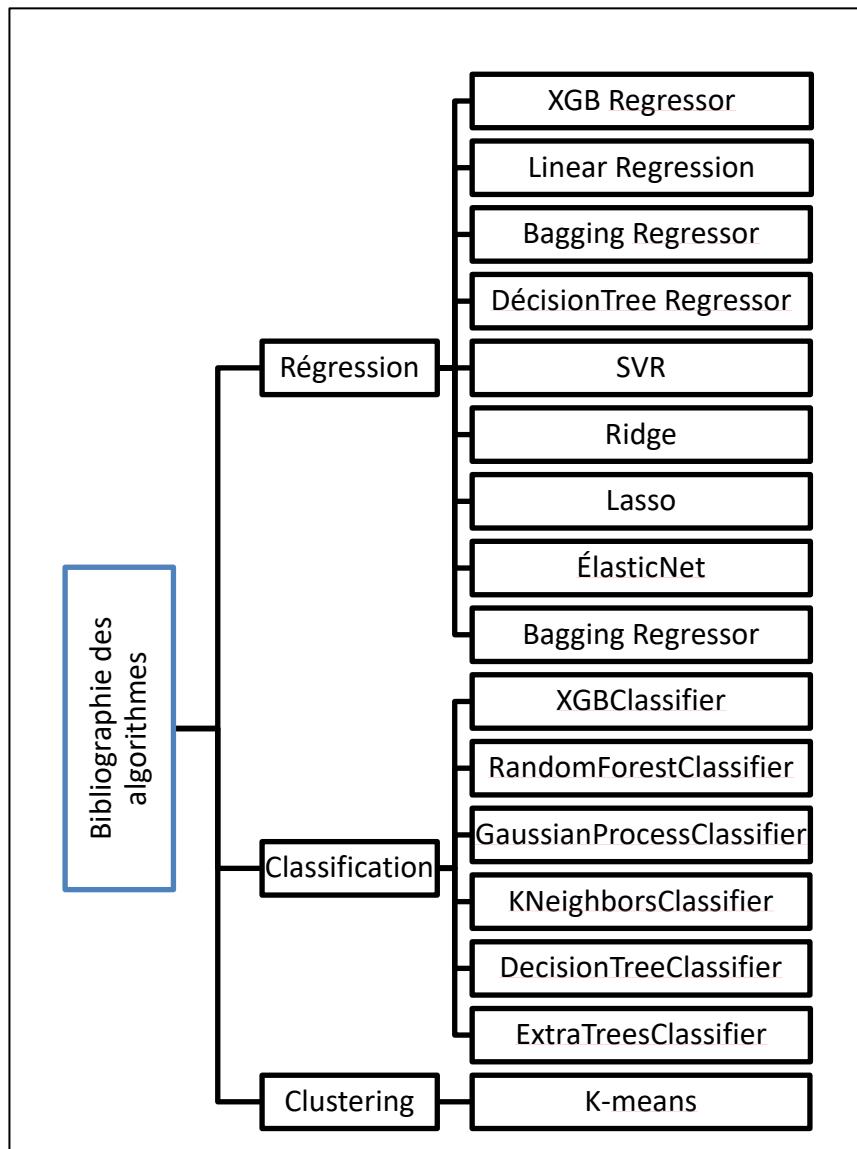


Figure 3.5 Bibliographie des algorithmes

### 3.4.2.1 Xgboost (Extreme Gradient Boosting):

The Gradient Boosting Decision Tree (GBDT) développé par (Friedman, 2001), est un algorithme d'apprentissage automatique fortement adaptatif qui est couramment utilisé dans les tâches de régression et de classification de l'exploration de données. Basé sur le concept de boosting, le GBDT construit plusieurs évaluateurs de base successivement en plusieurs itérations et résume les résultats de tous les évaluateurs de base pour améliorer les performances de régression ou de classification.

(T. Chen & Guestrin, 2016) ont développé XGboost qui est un algorithme GBDT optimisé où la sélection de ces paramètres dépend des données de training utilisées dans le modèle car il s'agit d'un modèle d'apprentissage supervisé non paramétrique. La principale différence avec GBDT est que lors du calcul de la fonction objective, la fonction de perte n'utilise que la dérivée première de la fonction de perte. XGBoost utilise l'expansion de Taylor du second ordre pour approximer la fonction de perte.

### 3.4.2.2 Linear Regression

La régression est une méthode d'apprentissage supervisé. Elle peut être utilisée pour faire des prédictions et modéliser des variables continues. Dans la régression, les ensembles de données sont étiquetés et la valeur de la variable de sortie est déterminée par les valeurs des variables d'entrée. La régression linéaire est le type de régression le plus basique, dans lequel on tente d'ajuster une ligne droite à l'ensemble de données, ce qui est possible lorsque la relation entre les variables de l'ensemble de données est linéaire.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (3.11)$$

L'équation 3.11 représente le modèle du Linear Regression,  $x$  est la variable indépendante,  $y$  est la variable dépendante,  $\beta_0$  la pente de la droite,  $\beta_1$  l'interception et  $\varepsilon$  représente l'erreur statique.

### 3.4.2.3 Logistic Regression

Logistique Regression (Zou et al., 2019) est utilisé pour traiter des problèmes de classification. Il fournit un résultat binomial en donnant la probabilité qu'un événement se produise ou non (en termes de 0 et 1) en fonction des valeurs des variables d'entrée. La différence la plus significative entre lui et le Linear Regression c'est que ses points de données ne sont pas organisés en rangées.

### 3.4.2.4 Decision Tree

L'arbre de décision (Izza et al., 2020) est une approche d'apprentissage automatique supervisé utilisée pour des problèmes de classification et de régression qui les résout en divisant continuellement les données en fonction d'un paramètre spécifique. Les décisions sont prises dans les feuilles, et les données sont divisées en nœuds. La variable de décision dans un arbre de classification est catégorique (un résultat de Oui/Non), alors que la variable de décision dans un arbre de régression est continue.

### 3.4.2.5 Support Vector Machine SVM

Les SVM (Vapnik & Izmailov, 2021) sont capables de traiter à la fois les problèmes de classification et de régression. L'objectif des SVM est de classer correctement les objets en fonction des exemples de l'ensemble de données d'apprentissage. Les avantages des SVM sont les suivants : ils peuvent traiter des données structurées et semi-structurées, et ils peuvent traiter des fonctions complexes si la fonction noyau appropriée peut être dérivée.

### 3.4.2.6 K Nearest Neighbour algorithm KNN

L'algorithme KNN (Lu et al., 2021) est un algorithme de classification. Il utilise une base de données avec des points de données organisés en plusieurs groupes, et la méthode tente de classer un point de données échantillon en tant que problème de classification. L'algorithme KNN est non-paramétrique, il s'agit d'une technique simple qui est facile à implémenter, au début on commence à chercher les K échantillons d'entraînement les plus proches de l'échantillon de test, c'est-à-dire le K-voisin le plus proche. Si une classe a le nombre maximal de voisins les plus proches, l'échantillon de test peut être déterminé comme appartenant à cette classe. Une distance euclidienne  $d$  est utilisée pour calculer la distance entre l'échantillon de test et tous les échantillons d'apprentissage. La formule est définie comme suit :

$$d(X, Y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (3.12)$$

### 3.4.2.7 Random Forest

Selon (Jin et al., 2020) Random Forest est un algorithme d'apprentissage automatique qui combine plusieurs algorithmes d'apprentissage automatique. Ceux-ci sont combinés à une série d'arbres classificateurs, chaque arbre votant unitairement pour la classe la plus populaire, et le résultat final du tri est obtenu en combinant ces résultats.

### 3.4.2.8 Extra Trees

Extra trees (Geurts et al., 2006) sont une technique d'apprentissage automatique relativement nouvelle qui a été développée comme une extension de l'algorithme Random Forest et qui est moins susceptible du surapprentissage. L'Extra trees (ET) utilise le même principe que le Random Forest et entraîne chaque algorithme de base avec un sous-ensemble aléatoire de caractéristiques (Ahmad et al., 2018). Cependant, il choisit la meilleure caractéristique au hasard, ainsi que la valeur correspondante. ET entraîne chaque arbre de régression en utilisant l'ensemble des données d'entraînement.

### 3.4.2.9 Adaboost

Le boosting est une approche d'apprentissage automatique qui repose sur l'idée de combiner plusieurs règles de prédiction relativement faibles et imprécises pour créer une règle de prédiction très précise. L'algorithme AdaBoost de (Freund & Schapire, 1997) a été le premier algorithme de boosting pratique et reste l'un des plus utilisés et des plus étudiés, avec des applications dans un large éventail de domaines.

### 3.4.2.10 Bagging Regressor

C'est un méta-estimateur qui ajuste les algorithmes de base sur des sous-ensembles aléatoires de l'ensemble de données originales, puis agrège leurs prédictions individuelles pour former une prédition finale (Khiari & Olaverri-Monreal, 2020).

## 3.4.3 Mesures de la performance

Pour l'évaluation de la performance des modèles nous allons utiliser des différents indicateurs :

$N$  est le nombre de données d'un échantillon.

### 3.4.3.1 R2 Score

Le  $R^2$  est une mesure statistique qui permet d'évaluer la qualité de prédition (Cameron & Windmeijer, 1997). La formule est définie comme suit :

$$R^2 = 1 - \frac{\sum_{i=1}^N (X_i^{actuel} - X_i^{prédict})^2}{\sum_{i=1}^N (X_i^{actuel} - \bar{X}^{prédict})^2} \quad (3.13)$$

### 3.4.3.2 Racine de l'erreur moyenne carrée (Root Mean Square Error)

Le RMSE est un métrique très utilisé pour les problèmes de régression qui permet de calculer la variance entre les valeurs réelles et les valeurs prédictes (Kardani et al., 2020).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i^{actuel} - X_i^{prédict})^2} \quad (3.14)$$

### 3.4.3.3 Erreur absolue moyenne (Mean Absolute Error)

Erreur absolue moyenne représente la moyenne de la somme des valeurs absolues entre les valeurs réelles et les valeurs prédites, plusieurs études privilégient le MAE par rapport au RMSE pour l'évaluation des modèles (Karunasingha, 2022).

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_i^{actuel} - X_i^{prédict}| \quad (3.15)$$

### 3.4.3.4 F1 score, accuarcy & precision

Étant donné des données d'apprentissage de la forme  $\{<x_i, y_i>, \dots, <x_n, y_n>\}$  où chaque  $x_i$  est un vecteur de caractéristiques de dimension d et chaque  $y_i$  est un vecteur binaire d'étiquettes vraies de dimension m, un classificateur probabiliste produit un modèle qui spécifie la probabilité conditionnelle (Lipton et al., 2014). Le tableau 3.2 nous affiche les abréviations des probabilités.

Tableau 3.2 Abréviaction des probabilités

Types des prédictions	Actuel Positive	Actuel Négative
Prédiction Positive	$tp$	$fp$
Prédiction négative	$fn$	$tn$

La formule du score F1 est comme suit :

$$F1 score = \frac{2tp}{2tp + fp + fn} \quad (3.16)$$

La formule de précision est comme suit :

$$Précision = \frac{tp}{tp + fp} \quad (3.17)$$

La formule de l'exactitude est comme suit :

$$Exactitude = \frac{tp + tn}{tp + tn + fp + fn} \quad (3.18)$$

### 3.4.4 Régression

À partir d'un état neuf, une pièce de machine qui a été correctement fabriquée et installée dans un système sans aucune erreur fonctionnera à un niveau de performance particulier, qui correspond idéalement à son exigence de conception. Au cours du temps, des dégradations se produisent. Pour cela l'étude de régression de l'état de santé du bien est nécessaire afin d'identifier rapidement les composants qui se détériorent et prévoir leur durée de vie pour que les ingénieurs de maintenance aient connaissance de la RUL du composant, qui constitue une opportunité pour programmer les procédures de maintenance, de réparation et de remplacement avant la défaillance. Selon (Maulud & Abdulazeez, 2020) la régression est une technique utilisée pour tester deux théories. Premièrement, les analyses de régression sont généralement utilisées pour la prévision et la prédiction, avec des chevauchements importants avec le domaine de l'apprentissage automatique. Deuxièmement, l'analyse de régression peut être utilisée pour déterminer les relations causales entre les variables indépendantes et dépendantes dans certains cas. Les régressions, en revanche, ne montrent que les relations entre une variable dépendante et un ensemble de données fixe de différentes variables.

#### 3.4.4.1 Méthodologie de simulation

Pour la prédiction de la durée de vie utile restante des composants de la machine, nous nous appuyons sur plusieurs algorithmes qu'ils ont été décrits dans la section 3.4.2, les tests de simulation seront effectués sur trois types des données, la première sur la totalité des données

par la suite sur les données récupérées par la méthode PCA et finalement sur les caractéristiques statiques extraites dans la section 3.3.

#### 3.4.4.2 Résultats de régression

Dans cette section nous allons présenter les résultats de régression pour la prédiction des RULs des quatre roulements pour les deux types de données :

Par la méthode PCA :

Après avoir entraîner et tester l'ensemble des modèles de régression avec les données obtenues par la méthode des composants principaux, la majorité des modèles ont donné un R2 score négative et des grandes erreurs, tous ces facteurs peuvent s'expliquer par les difficultés rencontrées par les modèles pour comprendre et suivre la tendance des données, avec la possibilité que les signaux présentent des bruits indésirables qui affecte l'aptitude de prédiction des algorithmes utilisés. C'est pourquoi nous allons à nouveau tester notre base d'algorithmes par des caractéristiques statiques.

- **Caractéristiques classiques :**

Tableau 3.3 Résultats de régression des roulements 1 et 2 (Suite)

Les roulements	Roulement 1			Roulement 2		
	Évaluation Modèle	R2 score	MAE	RMSE	R2 score	MAE
<b>Xg_boost</b>	61.22%	12.46	2.87	52.13%	12.61	2.89
<b>LinearRegression</b>	10.58%	20.93	3.29	23.56%	22.79	3.28
<b>SVM</b>	5.52%	23.78	3.32	-0.02%	25.18	3.37
<b>Ridge</b>	4.69%	21.23	3.32	6.11%	23.03	3.29
<b>Lasso</b>	10.76%	24.42	3.34	-0.03%	24.81	3.34

Les roulements	Roulement			Roulement		
	1		2			
Évaluation	R2	MAE	RMSE	R2	MAE	RMSE
Modèle	score			score		
<b>DecisionTree</b>	33.50%	23.78	3.51	22.44%	14.79	3.15
<b>ExtraTrees</b>	61.25%	19.5	3.2	50.96%	12.44	2.9
<b>AdaBoost</b>	53.21%	21.17	3.22	45.83%	15.26	2.98
<b>BaggingRegressor</b>	55.70%	19.62	3.23	46.18%	12.79	2.94
<b>RandomForest</b>	60.24%	19.09	3.19	52.30%	12.29	2.9
<b>Super learner</b>	62.16%	18.76	3.16	52.47%	12.36	2.88

Tableau 3.4 Résultats de régression des roulements 3 et 4

Les roulements	Roulement			Roulement		
	3		4			
Évaluation	R2 score	MAE	RMSE	R2 score	MAE	RMSE
Modèle						
<b>Xg_boost</b>	44.02%	14.08	2.9	70.83%	13.58	2.9
<b>LinearRegression</b>	23.37%	21.87	3.26	29.82%	20.08	3.28
<b>SVM</b>	9.63%	25.01	3.36	6.64%	25.04	3.36
<b>Ridge</b>	13.44%	22.37	3.28	14.48%	21.27	3.27
<b>Lasso</b>	13.04%	24.29	3.33	12.00%	23.03	3.29
<b>DecisionTree</b>	-8.80%	18.08	3.23	40.21%	18.97	3.28
<b>ExtraTrees</b>	40.94%	13.88	2.92	72.59%	13.74	2.93
<b>AdaBoost</b>	31.42%	15.86	2.98	42.51%	14.52	2.94
<b>BaggingRegressor</b>	36.65%	14.36	2.96	69.37%	14.11	2.97
<b>RandomForest</b>	41.90%	13.86	2.92	71.07%	13.75	2.94
<b>Super learner</b>	44.55%	13.8	2.89	72.85%	13.58	2.91

À la suite des résultats négatifs obtenus par les modèles de régressions avec les données réduites par la méthode des composants principaux, on a passé aux caractéristiques statiques décrites dans le paragraphe 3.1.2. Effectivement les modèles ont donné des meilleurs résultats avec ces données par rapport aux autres, en effet ont a passé des résultats négatifs qui consomment beaucoup du temps a des résultats qui ont un R2 score de 72.85% pour le Super Learner pour le roulement quatre. Les valeurs des erreurs ont été baissées mais elles restent encore un peu élevées pour cela nous allons passer dans notre étude pour une deuxième phase de classification.

### **3.4.5 Classification**

La classification est le processus consistant à classer chaque élément d'une collection de données dans l'une des classes ou l'un des groupes prédéfinis. Dans la classification des données, un modèle ou un classificateur est construit pour prédire la catégorie d'un échantillon. La classification est une fonction d'exploration de données qui catégorise ou classe les éléments d'une collection. L'objectif de la classification est de prédire avec précision la classe cible pour chaque cas dans les données (Kesavaraj & Sukumaran, 2013). Bien qu'il existe une variété de techniques disponibles pour l'apprentissage automatique, la classification est la technique la plus largement utilisée (Singh et al., 2013). Pour donner suite aux résultats de régression discutés dans le paragraphe 3.4.4.2, qui nous donnent la réflexion de traiter ce problème en tant qu'une étude de classification. Sur ce point, nous allons diviser nos données en dix classes comme c'est indiqué dans la figure 3.6, en attribuant à chacun un pourcentage de RUL, les tests seront conduits sur deux types des données identiquement a la régression.

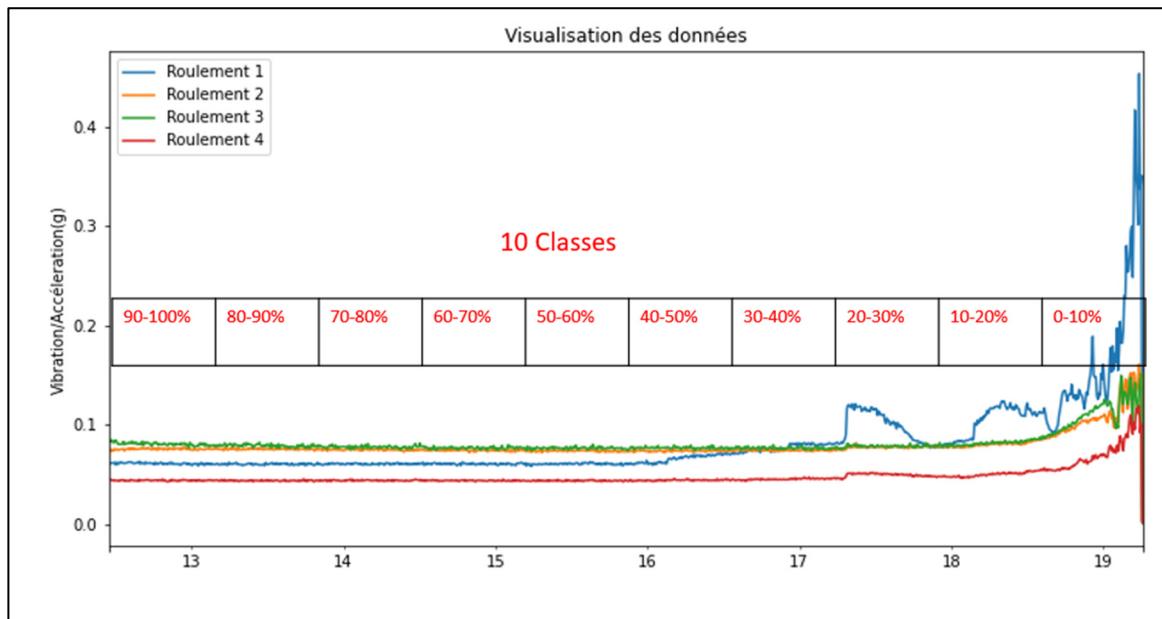


Figure 3.6 : Classification des données

Les algorithmes de classification qui seront utilisés sont :

- Logistic regression
- K-nearest neighbors KNN Classifier
- RandomForest Classifier
- Multilayer perceptron MLP Classifier
- ExtraTrees Classifier
- BaggingClassifier
- GaussianNB
- DecisionTree Classifier
- Xg\_boost Classifier

### 3.4.5.1 Résultats :

- Par la méthode PCA :

Tableau 3.5 Résultats de classification avec les données des PCA des roulements 1 et 2

Les roulements	Roulement 1			Roulement 2		
	F-score	Précision	Exactitude	F-score	Précision	Exactitude
Évaluation Modèle						
Xg_boostClassifier	7.16%	15.41%	14.61%	5.34%	10.6%	8.69%
LogisticRegression	11.61%	12%	12%	9.13%	9.17%	9.22%
RandomForestClassifier	2.91%	15.17%	11.42%	2.74%	1.61%	11.23%
KNNClassifier	1.80%	0.99%	9.93%	6.28%	5.12%	10.06%
MLPClassifier	10.62%	12.73%	6.3%	5.97%	7.29%	9.53%
ExtraTreesClassifier	9.32%	12.16%	15.18%	5.32%	11.62%	10.91%
BaggingClassifier	10.88%	14.77%	14.84%	8.01%	9.7%	10.06%
AdaBoost	8.55%	12.77%	14.61%	6.75%	9.59%	9.64%
GaussianNB	9.16%	8.52%	19.29%	1.41%	1.45%	5.83%
DecisionTreeClassifier	11.43%	12.22%	8.16%	8.74%	9.57%	9.64%

Tableau 3.6 Résultats de classification avec les données des PCA des roulements 3 et 4 (Suite)

Les roulements	Roulement 1			Roulement 2		
	F-score	Précision	Exactitude	F-score	Précision	Exactitude
Évaluation Modèle						
Xg_boostClassifier	2.74%	2.33%	4.4%	7.09%	8.71%	9.03%
LogisticRegression	10.23%	10.17%	10.36%	9.95%	9.91%	10.15%
RandomForestClassifier	2.23%	1.24%	2.5%	8.51%	13.97%	15.79%

Les roulements	Roulement			Roulement			
	1		2				
Modèle	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>MLPClassifier</b>	9.44%	9.33%	9.84%	9.60%	10.07%	10.46%	
<b>ExtraTreesClassifier</b>	3.21%	8.26%	4.6%	<b>11.90%</b>	<b>12.16%</b>	<b>13.95%</b>	
<b>BaggingClassifier</b>	3.25%	3.96%	11.03%	9.68%	9.57%	11.09%	
<b>AdaBoost</b>	5.94%	10.52%	10.88%	9.48%	10.75%	13.44%	
<b>GaussianNB</b>	2.22%	1.23%	3.4%	7.94%	19.15%	11.18%	
<b>DecisionTreeClassifier</b>	8.25%	10.67%	11.19%	9.17%	9.12%	9.64%	

Pour les modèles testés avec les données extraits par la méthode PCA, nous avons constaté que les résultats ont été amélioré par rapport aux résultats de régression avec cette méthode en passant des valeurs négatives de score R2 à un score F1 maximal de 11,9 %, toutefois ces valeurs restent toujours insuffisantes et inacceptables.

### Caractéristiques statiques :

Tableau 3.7 Résultats de classification avec les caractéristiques statiques des roulements 1 et 2 (Suite)

Les roulements	Roulement			Roulement			
	1		2				
Modèle	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>Xg_boostClassifier</b>	38.46%	39.2%	39.7%	38.28%	40.8%	39.52%	
<b>LogisticRegression</b>	20.24%	18.71%	23.75%	14.43%	12.2%	19.85%	
<b>RandomForestClassifier</b>	42.38%	45.68%	44.53%	36.74%	38.71%	38.59%	
<b>KNNClassifier</b>	28.19%	29.92%	27.83%	24.48%	26.6%	24.68%	
<b>MLPClassifier</b>	33.28%	35.17%	33.58%	17.18%	15.7%	20.96%	
<b>ExtraTreesClassifier</b>	40.58%	41.73%	41.93%	36.38%	37.67%	37.29%	

Les roulements		Roulement 1			Roulement 2		
Modèle	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>AdaBoost</b>		18.16%	23.67%	21.89%	20.44%	26.52%	27.27%
<b>GaussianNB</b>		40.62%	48.12%	43.6%	36.46%	46.04%	39.33%
<b>DecisionTreeClassifier</b>		33.88%	35.52%	32.84%	33.48%	34.16%	33.21%

Tableau 3.8 Résultats de classification avec les caractéristiques statiques des roulements 3 et 4

Les roulements		Roulement 1			Roulement 2		
Modèle	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>Xg_boostClassifier</b>		44.14%	46.15%	44.16%	73.77%	74.3%	74.03%
<b>LogisticRegression</b>		32.67%	34.13%	35.25%	57.75%	58.11%	60.11%
<b>RandomForestClassifier</b>		45.64%	48.02%	46.38%	<b>75.02%</b>	<b>76.26%</b>	<b>75.14%</b>
<b>KNNClassifier</b>		34.67%	36.17%	34.51%	65.82%	66.63%	66.23%
<b>MLPClassifier</b>		31.85%	33.96%	33.77%	61.22%	61.68%	62.52%
<b>ExtraTreesClassifier</b>		42.31%	44.3%	42.49%	74.34%	75.04%	74.77%
<b>BaggingClassifier</b>		43.55%	45.91%	43.78%	74.55%	74.98%	74.58%
<b>AdaBoost</b>		23.99%	24.95%	32.28%	40.63%	41.86%	44.16%
<b>GaussianNB</b>		42.64%	46.74%	44.53%	58.20%	61.27%	59.93%
<b>DecisionTreeClassifier</b>		38.98%	39.66%	38.96%	66.46%	66.6%	66.79%

Pour la classification utilisant les caractéristiques statiques, les résultats bien meilleurs qu'avec les autres données en atteignant un score F1 maximal de 75.02 %, une précision maximale de 76.26% et une exactitude égale à 75.14% pour les quatre roulements avec le modèle RandomForest Classifier. Lors de cette expérience de classification, nous avons constaté que les longueurs des données des trois tests sont différentes, de sorte que cela pourrait avoir un impact négatif sur ce test. En s'appuyant sur la figure 3.7, qui présente les trois cycles de vies des roulements en fonction des nombres de mesures pour chaque test, on peut constater que la

différence des nombres des mesures peut engendrer une confusion de prédition des modèles, prenant l'exemple d'un échantillon appartenant à la classe six du troisième cycle, sa classe équivalente dans le deuxième cycle est la septième ou la huitième classe, de même pour le première cycle de vie il appartient à la sixième ou la septième classe, ce qui explique les résultats obtenus pendant cette expérimentation et ce qui nous acheminons vers une troisième phase de Clustering.

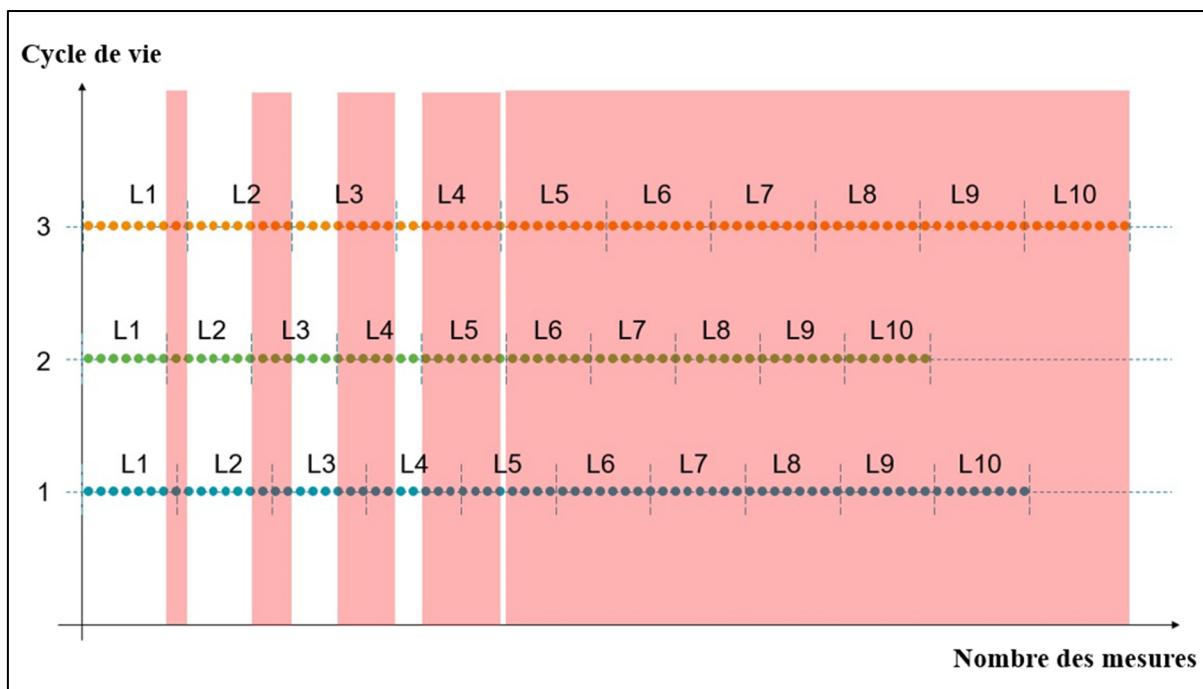


Figure 3.7 Chevauchement des classes

### 3.4.6 Clustering

Selon (Omran et al., 2007) le clustering est le problème d'apprentissage non supervisé le plus important ; comme tous les autres problèmes de ce type, il consiste à trouver une structure dans une collection de données non étiquetée. Un cluster est donc une collection d'objets qui sont similaires entre eux mais dissemblables aux objets des autres clusters. Pour donner suite aux résultats de classification et le problème de chevauchement des classes causé par la différence des nombres des mesures pour les trois cycles de vies, nous avons décidé de passer au clustering en mettant l'hypothèse suivante :

**Hypothèse :** On suppose que les signaux enregistrés pendant les trois tests des cycles de vies des roulements se décomposent en trois phases comme c'est indiqué dans la figure 3.8, une première phase où le roulement fonctionne en pleine performance, une deuxième phase où on constate le début de l'apparition de la panne suite aux changements des vibrations, et enfin une troisième phase qui représente l'éclosion de la panne.

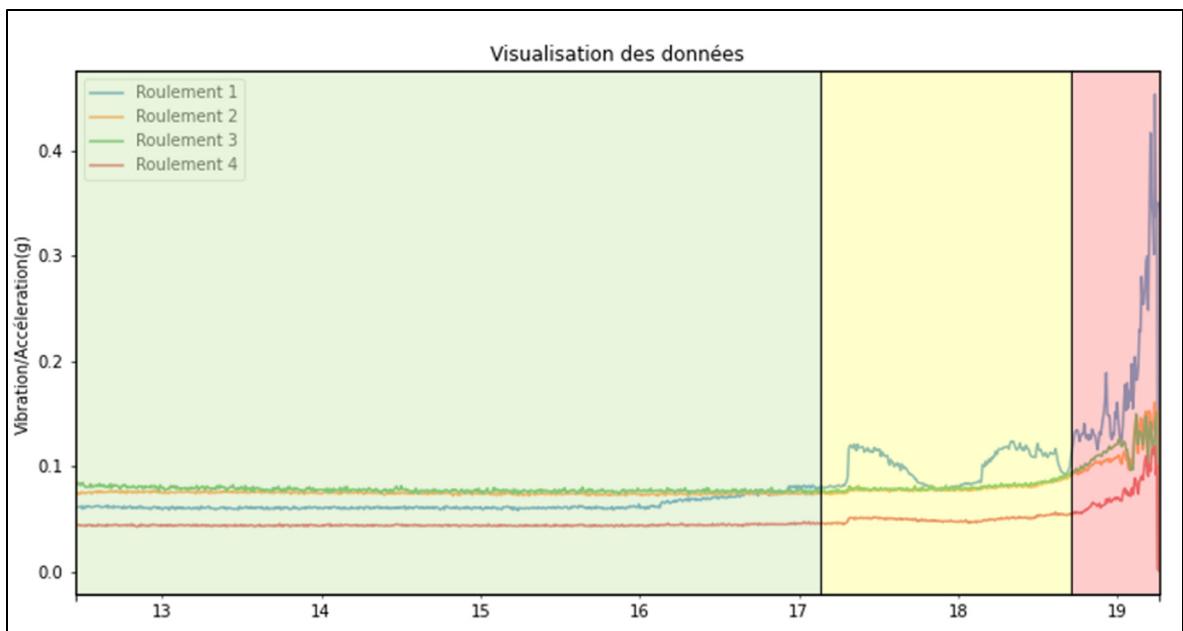


Figure 3.8 Décomposition des signaux en trois phases

#### 3.4.6.1 Résultats :

Pour commencer la section de clustering il fallait choisir un algorithme de clustering, le K-Means est l'une des techniques d'apprentissage non supervisé les plus fondamentales pour traiter le problème bien connu du clustering. L'approche suit une méthode simple et peu compliquée pour classer un ensemble de données donné à l'aide d'un nombre défini de clusters en supposant  $k$  clusters (Likas et al., 2003).

L'algorithme se décompose en plusieurs étapes :

- Choisir le nombre des clusters (3 dans notre cas)

- L'algorithme choisi par la suite 3 centroïdes pour les 3 clusters
- Il assigne chaque point du data au centroïde de cluster le plus proche.
- Ensuite il recalcule les centroïdes des clusters récemment créés.
- À la fin le Kmeans continue à répéter les étapes 3 et 4 jusqu'à les centroïdes des clusters ne changent pas.

- **Par la méthode PCA :**

Tableau 3.9 Résultats de clustering avec les données des PCA des roulements 1 et 2

Les roulements	Roulement 1			Roulement 2		
	Évaluation	F-score	Précision	Exactitude	F-score	Précision
Modèle						
<b>Xg_boostClassifier</b>	13.97%	15.2%	13.93%	29.10%	30%	31.67%
<b>LogisticRegression</b>	23.08%	26.51%	21.46%	42.66%	42.84%	45.23%
<b>RandomForestClassifier</b>	9.06%	9.02%	10.05%	16.81%	31.59%	19.28%
<b>KNNClassifier</b>	21.02%	25.91%	19.18%	20.04%	18.23%	22.67%
<b>SVM</b>	19.80%	19.77%	21.35%	24.63%	26.34%	24.36%
<b>MLPClassifier</b>	15.73%	17.79%	14.38%	19.16%	18.93%	19.6%
<b>ExtraTreesClassifier</b>	32.66%	33.85%	32.42%	36.75%	37.54%	37.29%
<b>BaggingClassifier</b>	10.57%	11.38%	10.73%	<b>52.63%</b>	<b>57.74%</b>	<b>55.83%</b>
<b>AdaBoost</b>	17.25%	18.75%	18.04%	38.10%	46.62%	41.63%
<b>GaussianNB</b>	33.78%	49.06%	37.79%	18.27%	27.54%	33.16%
<b>DecisionTreeClassifier</b>	15.57%	16.07%	15.18%	47.57%	52.98%	51.8%

Tableau 3.10 Résultats de clustering avec les données des PCA des roulements 3 et 4

Les roulements	Roulement			Roulement		
	3			4		
Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>Xg_boostClassifier</b>	33.94%	35.23%	38.34%	41.71%	37.69%	49.85%
<b>LogisticRegression</b>	32.76%	33.43%	32.33%	30.27%	29.54%	32.21%
<b>RandomForestClassifier</b>	28.41%	34.45%	31.5%	34.88%	36.27%	41.3%
<b>KNNClassifier</b>	37.70%	46.93%	38.03%	30.11%	30.72%	31.18%
<b>SVM</b>	16.99%	24.28%	29.33%	27.35%	26.39%	30.77%
<b>MLPClassifier</b>	35.33%	35.74%	36.79%	29.30%	26.98%	33.54%
<b>ExtraTreesClassifier</b>	32.39%	32.63%	33.47%	30.60%	30.59%	30.67%
<b>BaggingClassifier</b>	27.28%	29.28%	35.23%	46.40%	40.34%	57.44%
<b>AdaBoost</b>	36.60%	37.02%	36.99%	35.36%	33.46%	37.95%
<b>GaussianNB</b>	15.85%	21.37%	32.12%	23.15%	34.38%	31.79%
<b>DecisionTreeClassifier</b>	30.41%	31.14%	33.16%	43.20%	42.57%	48.41%

En comparant les résultats de clustering présentés dans les tableaux 3.9 et 3.10 des quatre roulements avec les résultats de classification en utilisant les données de PCA, le score F1 maximal a été augmenté de 11,9% par ExtraTreesClassifier à 52,63 % par le BaggingClassifier, respectivement la précision maximale de 12.16% à 57.74% et une exactitude maximale de 13.95% à 55.83%.

- **Caractéristiques statiques :**

Tableau 3.11 Résultats de clustering avec les caractéristiques statiques des roulements 1 et 2 (Suite)

Les roulements	Roulement			Roulement		
	1			2		
Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>Xg_boostClassifier</b>	97.48%	97.54%	97.59%	97.21%	97.2%	97.22%
<b>LogisticRegression</b>	79.02%	76.39%	85.16%	83.15%	81.62%	86.46%
<b>RandomForestClassifier</b>	96.77%	97.13%	97.03%	97.39%	97.39%	97.4%

Les roulements		Roulement 1		Roulement 2			
Modèle	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>KNNClassifier</b>	97.52%	97.66%	97.88%	95.28%	96.77%	95.12%	
<b>SVM</b>	95.36%	94.39%	95.66%	93.29%	93.67%	93.43%	
<b>MLPClassifier</b>	90.52%	93.31%	92.76%	87.41%	87.15%	88.13%	
<b>ExtraTreesClassifier</b>	96.32%	96.79%	96.66%	97.38%	97.39%	97.4%	
<b>BaggingClassifier</b>	97.71%	97.72%	97.77%	97.01%	97.01%	97.03%	
<b>AdaBoost</b>	94.71%	95.41%	95.36%	91.89%	92.87%	92.39%	
<b>DecisionTreeClassifier</b>	97.15%	97.13%	97.22%	95.38%	95.41%	95.36%	

Tableau 3.12 Résultats de clustering avec les caractéristiques statiques des roulements 3 et 4

Les roulements		Roulement 3			Roulement 4		
Modèle	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>Xg_boostClassifier</b>	98.87%	98.87%	98.89%	97.95%	97.96%	97.96%	
<b>LogisticRegression</b>	96.71%	96.67%	96.85%	95.46%	95.64%	95.55%	
<b>RandomForestClassifier</b>	98.67%	98.68%	98.7%	98.13%	98.14%	98.14%	
<b>KNNClassifier</b>	97.82%	97.96%	97.96%	97.18%	97.24%	97.22%	
<b>SVM</b>	96.36%	96.39%	96.66%	94.29%	94.67%	94.43%	
<b>MLPClassifier</b>	97.63%	97.68%	97.77%	<b>98.89%</b>	<b>98.89%</b>	<b>98.89%</b>	
<b>ExtraTreesClassifier</b>	98.85%	98.87%	98.89%	98.69%	98.71%	98.7%	
<b>BaggingClassifier</b>	98.30%	98.31%	98.33%	98.32%	98.33%	98.33%	
<b>AdaBoost</b>	96.42%	96.64%	96.29%	94.16%	94.4%	94.06%	
<b>GaussianNB</b>	96.90%	96.89%	97.03%	96.05%	96.2%	96.1%	
<b>DecisionTreeClassifier</b>	98.16%	98.2%	98.14%	98.14%	98.14%	98.14%	

Pour le test des modèles de Clustering avec les données statiques les résultats sont présentés dans les tableaux 3.13 et 3.14, le score F1 maximal a été augmenté de 75,02 % par le Random Forest Classifier à 98,89 % par le MLPClassifier, respectivement la précision maximale de

76.26% à 98,89 % et une exactitude maximale de 75.14% à 98,89 %, ce qui nous confirme l'hypothèse mis au début de l'expérience, et ça nous aide à identifier la phase ou un échantillon appartient avec précision et exactitude maximales de 98,89% avec le modèle MLP, aussi les résultats de la majorité des modèles étaient satisfaisants ou ils ont arrivés à bien classifier les 3 clusters des données.

- **Partie 2 : Identification des anomalies**

#### 3.4.7 Détection d'anomalies et classification des défauts

La détection des anomalies est un problème critique qui a été étudié et mis en œuvre dans de nombreux domaines. Plusieurs méthodes de détection des anomalies ont été développées pour des applications spécifiques, tandis que d'autres sont plus génériques. Par exemple, (Prasad et al., 2009) ont effectué un examen approfondi des techniques et des applications de détection des anomalies. Ils ont procédé à une enquête détaillée de diverses techniques d'apprentissage automatique et non automatique, telles que les méthodes de détection statistique et spectrale. En outre, l'étude couvre une variété d'applications de détection d'anomalies. La détection des cyber-intrusions, la détection des fraudes, la détection des anomalies médicales, la détection des dommages industriels... etc. Dans notre étude, nous allons étudier la détection des anomalies des roulements pour procéder par la suite à leurs classifications selon leurs types.

##### 3.4.7.1 Isolation Forest

Isolation Forest ou iForest est une combinaison de iTrees, qui est une méthode efficace de détection des anomalies. Il s'agit d'un algorithme efficace et populaire qui peut traiter de grandes données. Il se diffère des méthodes traditionnelles de détection des anomalies par sa grande précision, sa complexité temporelle linéaire et son faible coût en mémoire. Selon (Liu et al., 2008) le principe de fonctionnement consiste à traiter des données sous-échantillonées de manière aléatoire dans une structure arborescente basée sur des caractéristiques sélectionnées de manière aléatoire. Les échantillons qui se déplacent plus profondément dans l'arbre sont moins susceptibles d'être des anomalies, car il a fallu plus de coupes pour les isoler.

De même, les échantillons qui aboutissent dans des branches plus courtes indiquent des anomalies car il a été plus facile pour l'arbre de les séparer des autres observations. Pour mieux comprendre nous allons décortiquer l'algorithme en étapes :

- A partir d'un ensemble de données, on choisit un sous-ensemble aléatoire de celui-ci et on l'affecte à un arbre binaire.
- L'arbre est d'abord ramifié en sélectionnant une caractéristique aléatoire (parmi l'ensemble des N caractéristiques). La ramification est ensuite effectuée à un seuil aléatoire (toute valeur comprise entre les valeurs minimale et maximale de la caractéristique sélectionnée).
- Si la valeur d'un point de données est inférieure au seuil sélectionné, il est acheminé vers la branche gauche ; sinon, il est acheminé vers la branche droite. Par conséquent, un nœud est divisé en branches gauche et droite.
- Ce processus est effectué de manière répétitive jusqu'à ce que chaque point de données soit complètement isolé.
- La formation du modèle est terminée lorsqu'un ensemble d'iTrees (forêt d'isolation) est créé. Pendant la notation, un point de données est parcouru par tous les arbres précédemment formés. Chaque point de données reçoit maintenant un "score d'anomalie" basé sur la profondeur de l'arbre nécessaire pour atteindre ce point. Ce score est une moyenne des profondeurs obtenues à partir de chaque iTree. En fonction du paramètre de contamination (pourcentage d'anomalies présentes dans les données), un score d'anomalie de -1 est attribué aux anomalies et un score de 1 aux points normaux.

#### **3.4.7.2 Détection d'anomalies par l'algorithme Isolation Forest**

La détection d'anomalies a été appliquée sur les signaux de vibrations des quatre roulements :

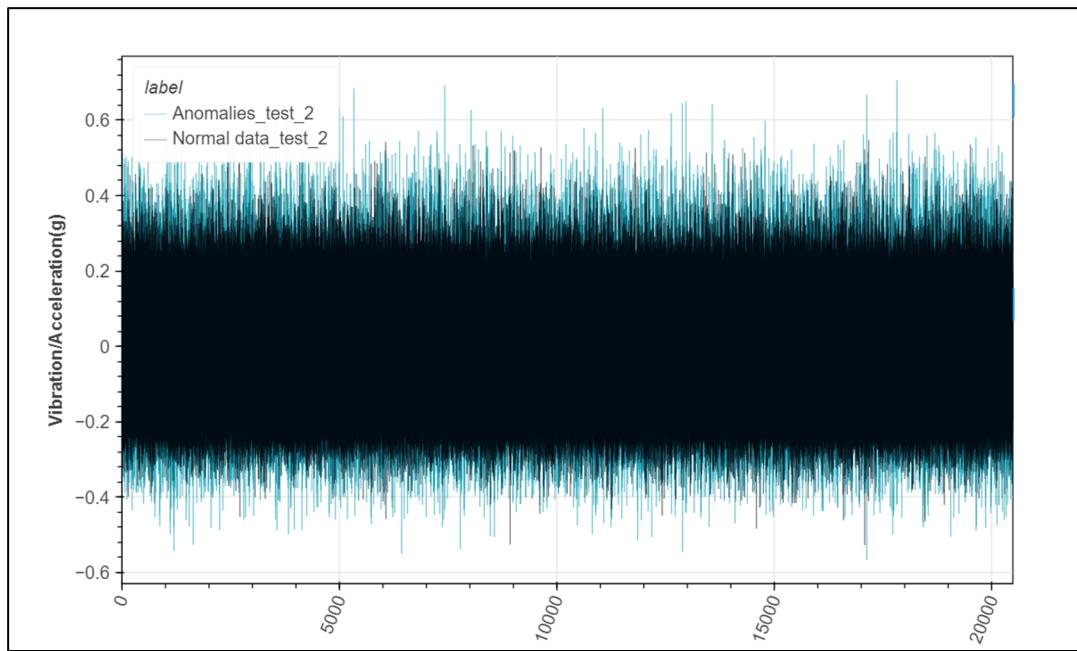


Figure 3.9 Anomalies détectées dans le roulement N°4 pendant le deuxième test

La figure 3.9 nous montre le signal sain en noir et les anomalies détectées en bleu.

### 3.4.7.3 Classification des anomalies

Avant de commencer la classification des défauts, il fallait donner des classes aux anomalies détectées dans chaque roulement, en se basant sur le document de la base des données étudiées nous avons extrait le tableau des défauts suivant :

Tableau 3.13 Tableau des défauts des roulements

Roulements	Test1	Test2	Test3
<b>Roulement 1</b>	X	Défaut dans la bague extérieure	X
<b>Roulement 2</b>	X	X	X
<b>Roulement 3</b>	Défaut dans la bague intérieure	X	Défaut dans la bague extérieure
<b>Roulement 4</b>	Défaut dans l'élément roulant	X	X

Pour classifier ces défauts à l'aide des modèles d'intelligence artificielle nous avons attribué des classes aux défauts :

- Classe 0 : Défaut dans la bague extérieur
- Classe 1 : Défaut dans la bague intérieur
- Classe 2 : Défaut dans l'élément roulant
- Classe 3 : Autres défauts détecter par l'algorithme Isolation Forest

#### 3.4.7.4 Résultats :

La performance en classification de modèles est évaluée à l'aide des deux types des données. Les figures suivantes présentent les matrices de confusion qui nous permettent de visualiser les classes « faciles » et « difficiles » pour un modèle à prédire.

Données originales :

- **Classification avec les données extraites par la méthode PCA :**

Pour le modèle Decision Trees, la matrice de confusion résultante est donnée dans la Figure 3.10.

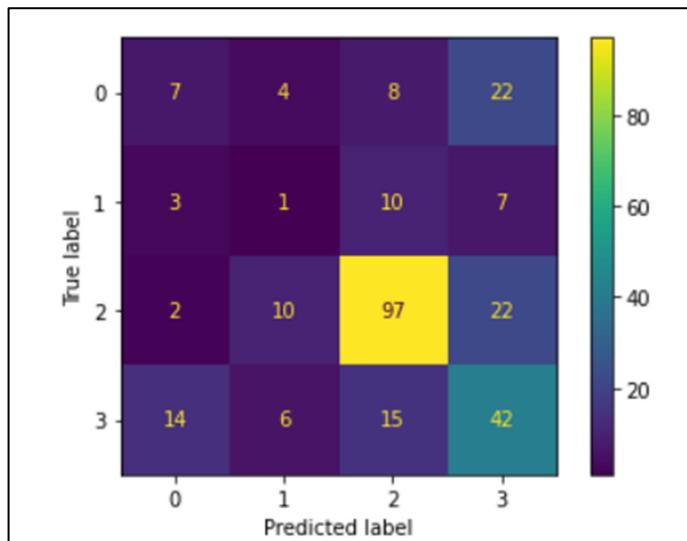


Figure 3.10 Matrice de confusion du Décision Trees

La performance en classification est résumée dans le Tableau 3.14, ce modèle n'était pas capable de bien connaître les défauts de classe 0,1.

Tableau 3.14 Performance du modèle Décision trees

<b>Exactitude = 147/270 (54%)</b>				
	<b>Classe 0</b>	<b>Classe 1</b>	<b>Classe 2</b>	<b>Classe 3</b>
<b>Précision</b>	0.27	0.05	0.75	0.45
<b>F1 score</b>	0.21	0.05	0.74	0.49
<b>Recall</b>	0.17	0.05	0.74	0.55

Pour le modèle Gaussien NB la matrice de confusion résultante est donnée dans la Figure 3.11.

Tableau 3.15 Performance du modèle Gaussian NB

<b>Exactitude = 210/270 (78%)</b>				
	<b>Classe 0</b>	<b>Classe 1</b>	<b>Classe 2</b>	<b>Classe 3</b>
<b>Précision</b>	0.49	0.60	0.89	0.72
<b>F1 score</b>	0.46	0.23	0.93	0.75
<b>Recall</b>	0.44	0.14	0.98	0.78

La performance en classification est résumée dans le Tableau 3.15, c'est le meilleur modèle avec les données des PCA, mais il est incapable de bien détecter les défauts de classe 1.

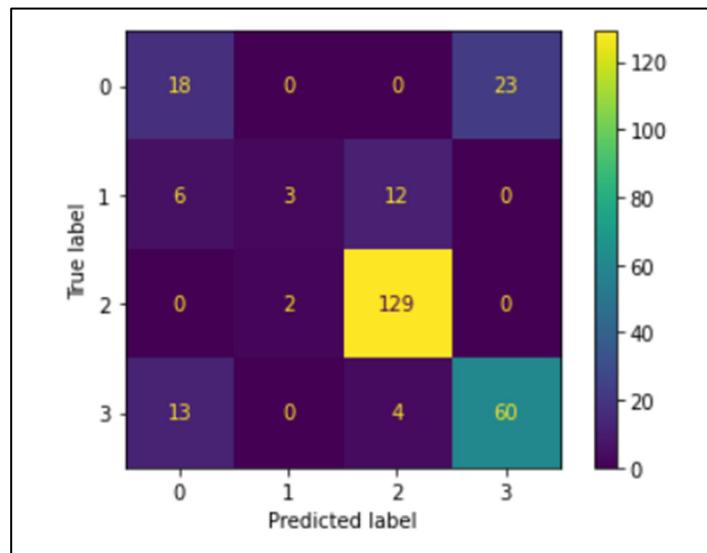


Figure 3.11 : Matrice de confusion du Gaussian NB

- **Classification avec les données des caractéristiques statiques :**

Pour le modèle Gaussien NB la matrice de confusion résultante est donnée dans la Figure 3.12.

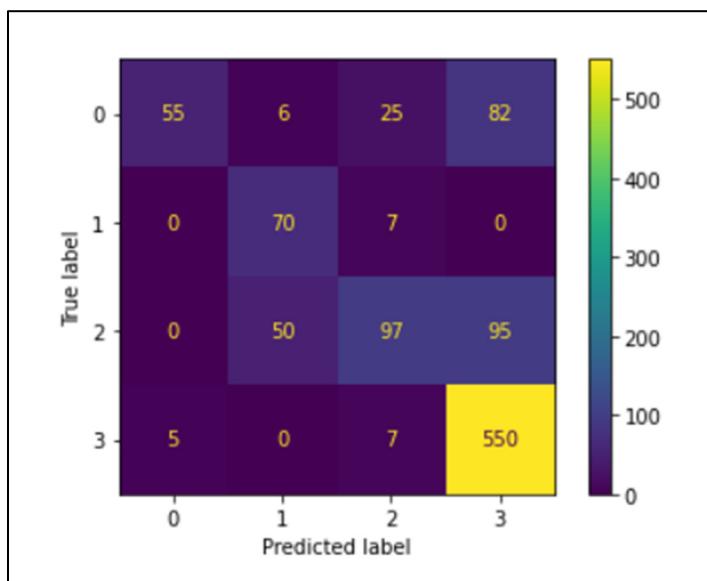


Figure 3.12 : Matrice de confusion du Gaussian NB

La performance en classification est résumée dans le Tableau 3.16, étant donné que ce modèle était le meilleur avec les données PCA, il a donné des mauvais résultats et est apparu non capable de bien détecter les défauts de classes 0,1 et 2.

Tableau 3.16 Performance du modèle Gaussian NB

Exactitude = 726/1049 (69%)				
	Classe 0	Classe 1	Classe 2	Classe 3
<b>Précision</b>	0.70	0.67	0.62	0.71
<b>F1 score</b>	0.45	0.56	0.49	0.81
<b>Recall</b>	0.33	0.48	0.41	0.81

Pour le modèle Extra Trees, la matrice de confusion résultante est donnée dans la Figure 3.13.

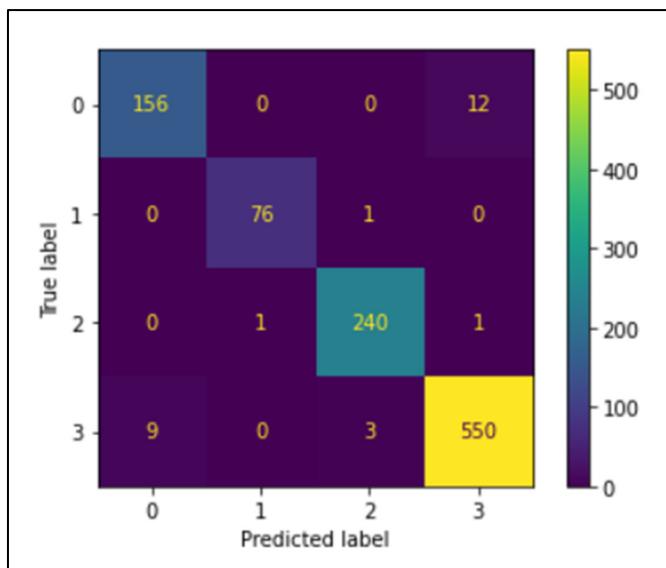


Figure 3.13 Matrice de confusion de l'Extra Trees

La performance en classification est résumée dans le Tableau 3.17, notamment ce modèle nous a donné une précision 97,42% de classification des défauts selon leurs types, ce qui lui mette au « top » de la liste des modèles. De plus, d'autres modèles ont donné des bons résultats tel

que le Random Forest avec 96.62% de précision, Xgboost 96.26% et le Bagging Regressor 95.59%.

Tableau 3.17 Performance du modèle Extra trees

<b>Exactitude = 1022/1049 (97,42%)</b>				
	<b>Classe 0</b>	<b>Classe 1</b>	<b>Classe 2</b>	<b>Classe 3</b>
<b>Précision</b>	0.95	0.99	0.98	0.98
<b>F1 score</b>	0.94	0.99	0.99	0.98
<b>Recall</b>	0.93	0.99	0.99	0.98

### 3.4 Conclusion

Dans ce chapitre, une base d’algorithmes pour l’évaluation de l’état de santé des roulements a été présentés. Deux types d’extractions des caractéristiques ont été discutés pour les utiliser par la suite dans le test des modèles. Une démarche basée sur les résultats et sur la nature des données est adoptée pour pouvoir classifier les données en trois phases selon leurs états de santé. Ensuite la détection des anomalies des roulements est faite, de surcroit leurs classifications.

Dans le chapitre suivant, la même démarche sera appliquée sur les données qui seront enregistrées par des capteurs MEMS accéléromètres dans notre laboratoire, afin d’évaluer ses performances envers leur utilisation pour la maintenance prédictive.

## CHAPITRE 4

### ENREGISTREMENT DES DONNÉES ET LA VALIDATION DES RÉSULTATS

#### 4.1 Introduction

Dans ce chapitre, nous allons présenter le banc de test utilisé pour la collection des données de vibrations des roulements, ensuite les résultats des tests de validation seront discutés.

#### 4.2 Banc de test

Ce projet se base sur l'enregistrement des signaux des vibrations des roulements en utilisant des capteurs MEMS. Partant de ce fait et selon la disponibilité des équipements dans notre département électronique, nous avons choisi au début la source d'entrainement de notre montage qui est un moteur DC de la marque ABB Drives de type 71S – 2GVWTG avec une vitesse maximale de 3000 RPM et une puissance maximale de 0.25 KW présenté dans la figure 4.1.



Figure 4.1 Moteur DC ABB Drives  
71S – 2GVWTG

#### 4.2.1 Système et processus d'acquisition

Cette section a pour but de mettre en évidence les différents composants du système et la procédure de collecte des données pour les roulements. Nous présentons tout d'abord les différents composants du système, puis la méthode de collecte, et enfin nous décryptons le jeu de données collectées.

Deux types de capteurs sont inclus dans le système d'acquisition ainsi qu'un module d'acquisition. Le choix des capteurs a été basé sur plusieurs facteurs, selon (Murphy, 2020) les capteurs les mieux adaptés pour la surveillance de l'état des roulements sont les accéléromètres MEMS capacitifs et les accéléromètres Piezo-électrique, de surcroit les MEMS sont moins chères et plus petits que les Piezo. Pour ces raisons, nous allons choisir deux capteurs de type MEMS pour la collection des données et afin de valider ses compétences à fournir des données fiables qu'on peut les utiliser dans un contexte de maintenance prédictive.

Tableau 4.1 Caractéristiques des capteurs MEMS

Capteur MEMS	Portée	Bandé passante	Sensibilité	Sensibilité sur l'axe transversal	Prix
<b>MMA7455</b>	$\pm 2g$	62.5/125 Hz	54 LSB/g	$\pm 5\%$	14 \$
<b>MMA 7361</b>	$\pm 1.5g$	400 Hz	800 mV/g ou 6400 LSB/g	$\pm 5\%$	10 \$
<b>MMA8452</b>	$\pm 2g$	0.78-400 Hz	1024 LSB/g	-	14 \$
<b>ADXL 335</b>	$\pm 3g$	<b>1600 Hz</b>	<b>300 mV/g or 2400 LSB/g</b>	$\pm 1\%$	<b>10 \$</b>
<b>MPU 6050</b>	$\pm 2g$	500 Hz	16384 LSB/g	$\pm 2\%$	10 \$
<b>BMI160</b>	$\pm 2g$	5.06-684 Hz	16384 LSB/g	$\pm 1\%$	20 \$
<b>ADXL1002Z</b>	$\pm 50g$	<b>11 kHz</b>	<b>40mV/g</b>	$\pm 1\%$	<b>85\$</b>

Le tableau 4.1 nous montre des différentes caractéristiques des capteurs MEMS accéléromètres disponible dans le marché. Dans l'ensemble, on trouve que l'ADXL 335 (Figure 4.2.a) présente le meilleur choix puisqu'il donne une meilleure portée et sensibilité pour un coût le plus faible. Aussi, nous avons choisi le capteur ADXL1002Z (Figure 4.2.b) de la société Analog Devices étant donné qu'il est fabriqué spécialement pour la maintenance prédictive en 2018 et qu'il n'est pas très utilisé dans les travaux de recherche, ce dernier nous offre une très bonne portée de  $\pm 50\text{g}$ , une bande passante de 11 kHz avec une faible sensibilité de 40mV/g pour un prix de 85\$.

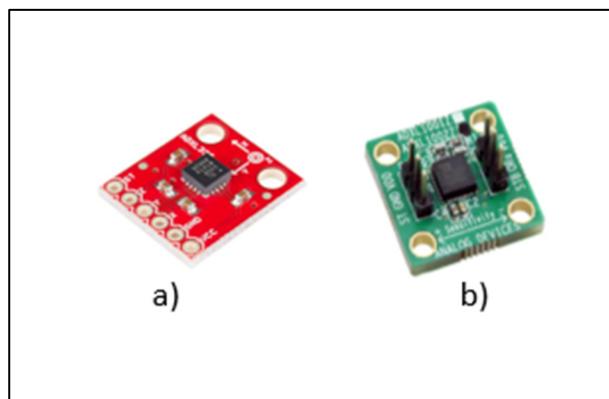


Figure 4.2 a) Adxl 335, b) Adxl1002Z

Les signaux de vibration analogique captés par ces deux capteurs seront acquis par une carte Arduino UNO (Figure 4.3.a) basée sur le microcontrôleur ATmega328P, transmis par la suite à une carte Raspberry pi 4 (Figure 4.3.b) via un câble USB afin de les mettre dans un fichier CSV.

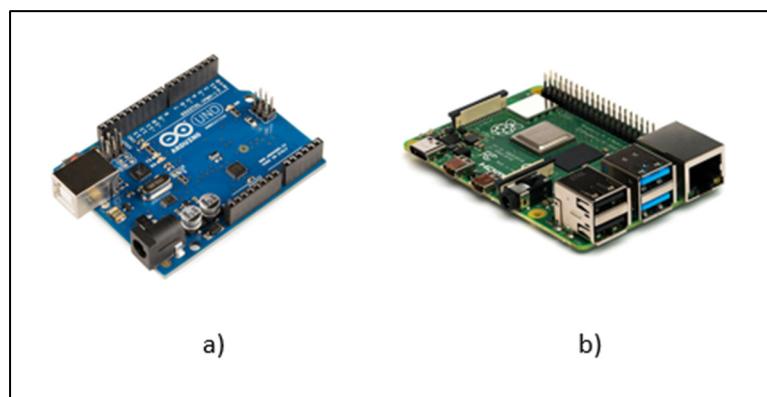


Figure 4.3 a) Arduino Uno, b) Raspberry Pi 4

#### 4.2.2 Composants mécaniques :

Le montage de test (figure 4.4) est composé d'un arbre acier carbone de diamètre 30mm et de longueur 500mm, accouplé avec le moteur par un accouplement élastique.

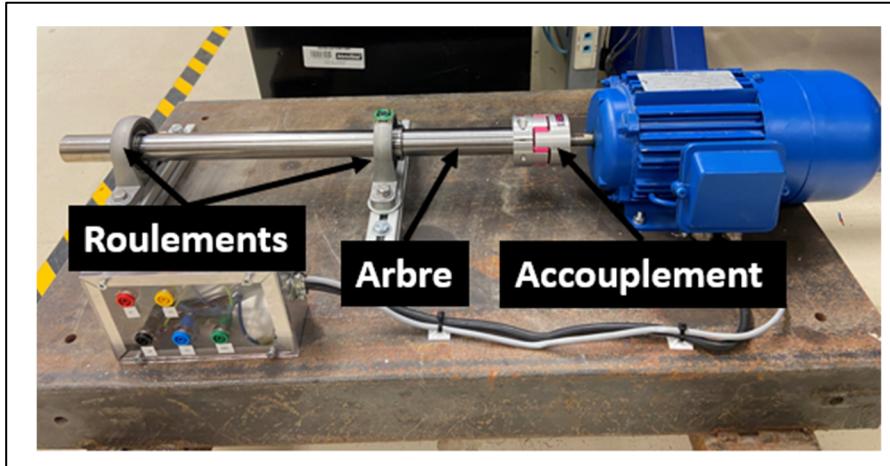


Figure 4.4 Montage de test

Le Guidage en rotation est assuré par deux roulements de type BC faits en alliage d'acier monté dans une boîtier en fonte, et de diamètre 30mm illustré dans la figure 4.5.a.

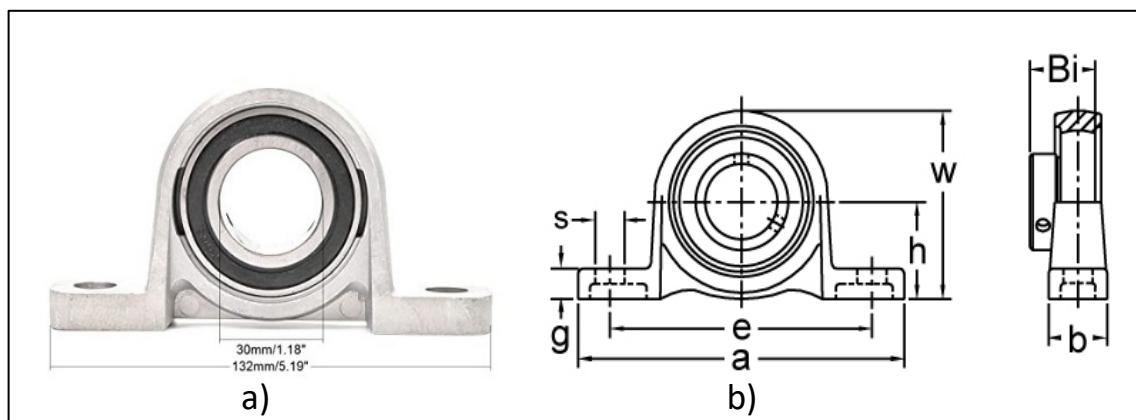


Figure 4.5 a) Roulement de type BC, b) Les dimensions de roulements

Les dimensions des roulements de la figure 4.5.b sont décrites dans le tableau 4.2 :

Tableau 4.2 Côtes géométriques des roulements

Côte	a	b	e	g	h	s	w	Bi
Valeur (mm)	132	26	106	11	36	13	70	24.5

Les roulements utilisés ont une longueur de 132 mm, une largeur de la base égale à 26 mm, la hauteur est de 70 mm, avec un diamètre de 30 mm.

#### 4.2.3 Fixation des capteurs

Après avoir préparé le montage mécanique, nous avons passé à la fixation des MEMS accéléromètres sur les roulements, une colle résine époxyde (figure 4.6) a été utilisé avec du ruban double face.



Figure 4.6 Colle résine époxyde

Par la suite nous avons fixé les capteurs sur les roulements comme c'est montré dans la figure 4.7 :

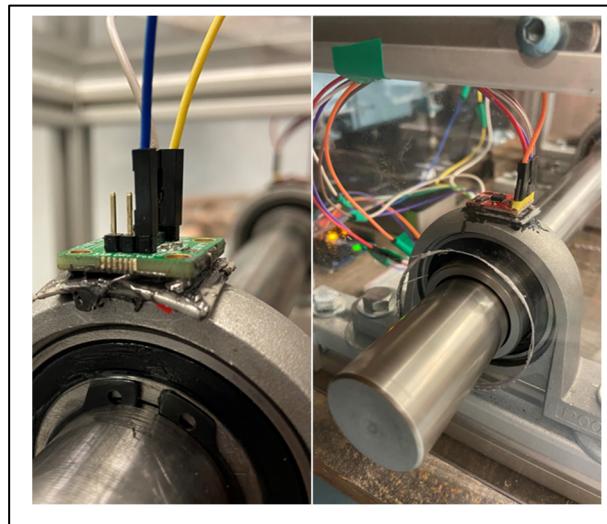


Figure 4.7 Fixation des capteurs

L'utilisation de résine époxy avec le ruban double face aide à bien adhérer les capteurs sur la surface des roulements afin d'améliorer la qualité des signaux de vibration enregistrés. Mais il fallait vérifier son effet sur la bande passante du capteur.

Prenant l'exemple du capteur Adxl1002Z sa fréquence de résonance est égale à 21kHz, sa bande passante est égale à 11kHz en utilisant de l'adhésive sa bande passante est réduite jusqu'à 7kHz qui reste toujours suffisant pour notre étude puisqu'on fait l'échantillonnage à 1 kHz comme c'est montré dans la figure 4.8.

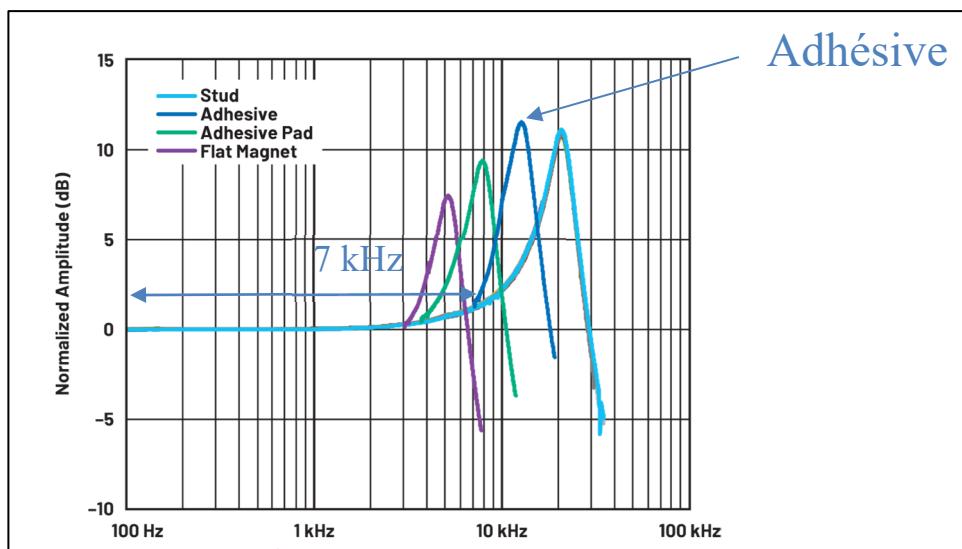


Figure 4.8 : Effet de la colle époxy sur la bande passante du capteur Adxl 1002Z

#### 4.2.4 Enregistrement des données

La figure 4.8 nous montre l'installation du montage dans le laboratoire pour l'enregistrement des données, le moteur est entraîné en rotation avec une vitesse égale à 1500 RPM, les MEMS accéléromètres sont branchés avec l'Arduino pour convertir leurs signaux analogiques en un signal numérique, ces données seront transmises par la suite à la deuxième carte Raspberry 4 via une Communication en série en utilisant un câble USB-UART.

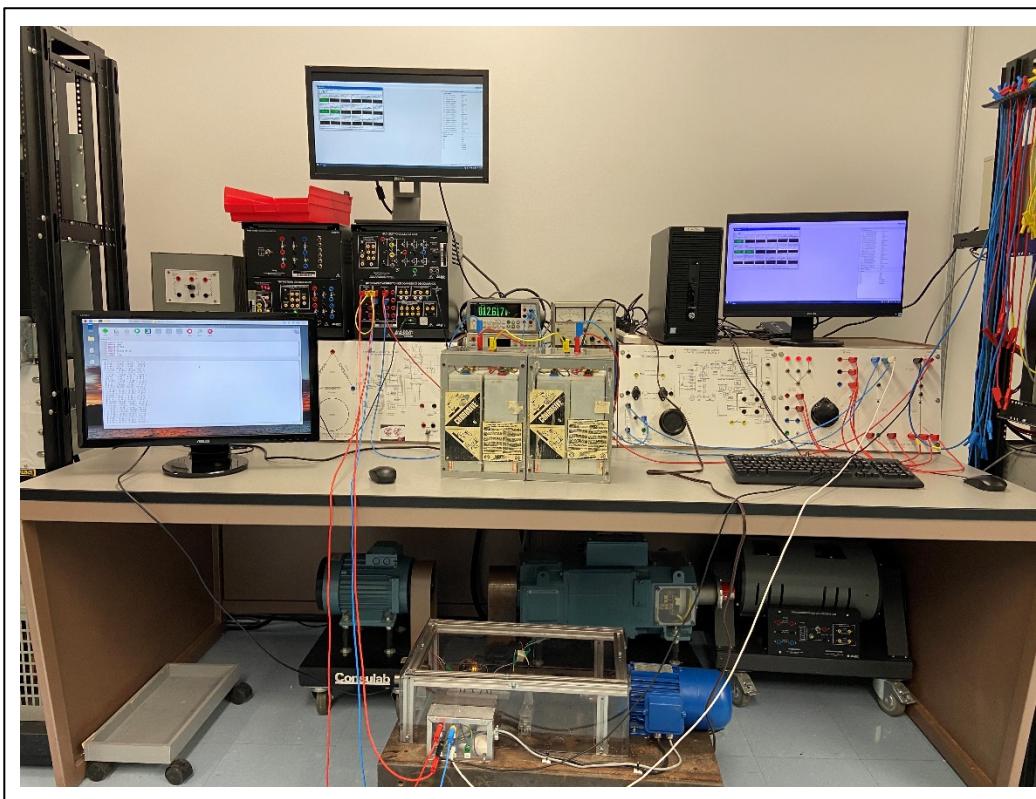


Figure 4.9 Installation du montage pour la collecte de données et l'analyse des défauts

Ce banc d'essai expérimental est construit et développé pour générer des données de vibrations des roulements. Les données seront utilisées pour valider les résultats obtenus dans le chapitre précédent, et pour examiner la performance de ces capteurs MEMS accéléromètres d'être utilisés dans un contexte de maintenance prédictive. Deux tests ont été faits sur les roulements, le premier sera utilisé pour l'entraînement des modèles et le deuxième pour tester et valider.

Les durées de vie des roulements des différents tests sont accélérées par des perçages de diamètres 4mm dans les billes comme nous montre la figure 4.9.

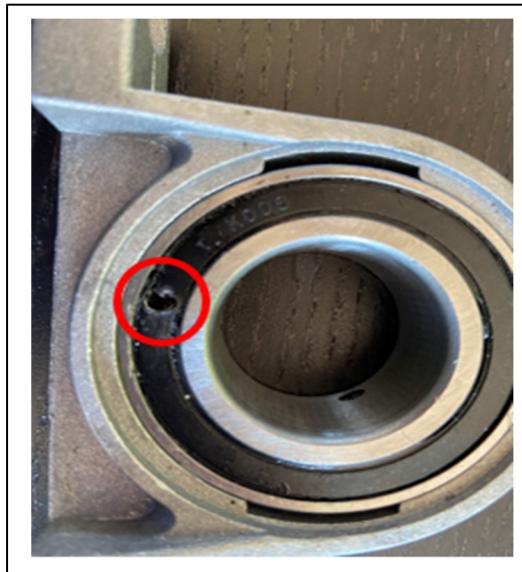


Figure 4.10 Perçage de l'élément roulant du roulement

Les figures 4.10 et 4.11 ci-dessous nous montrent les signaux de vibrations enregistrés pendant les deux tests.

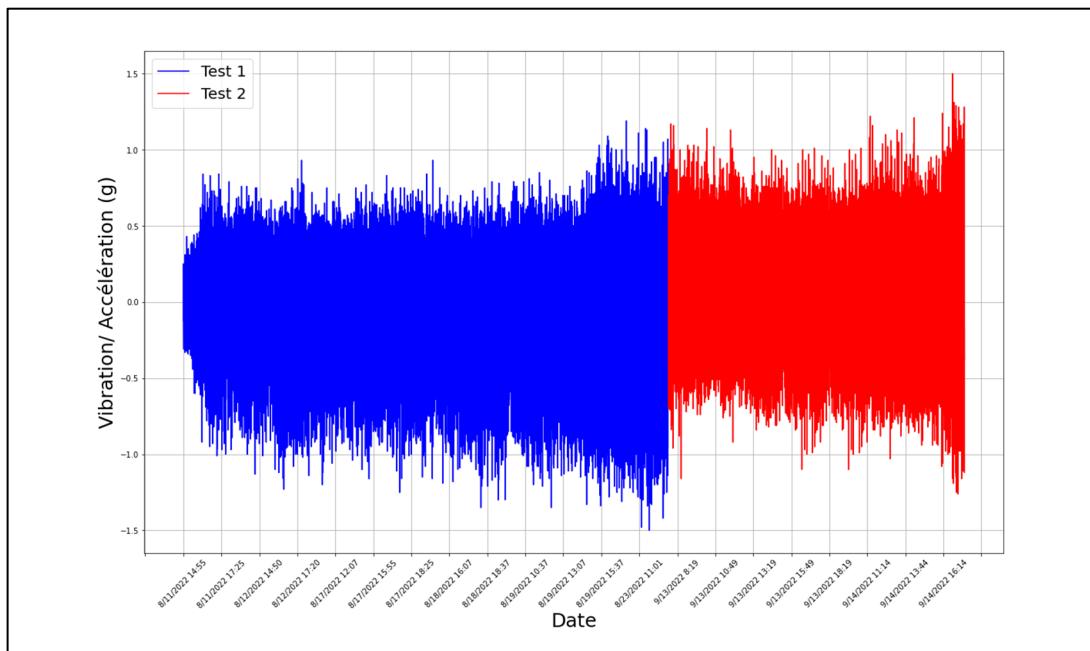


Figure 4.11 Signaux de vibrations enregistrés avec l'accéléromètre Adxl 335

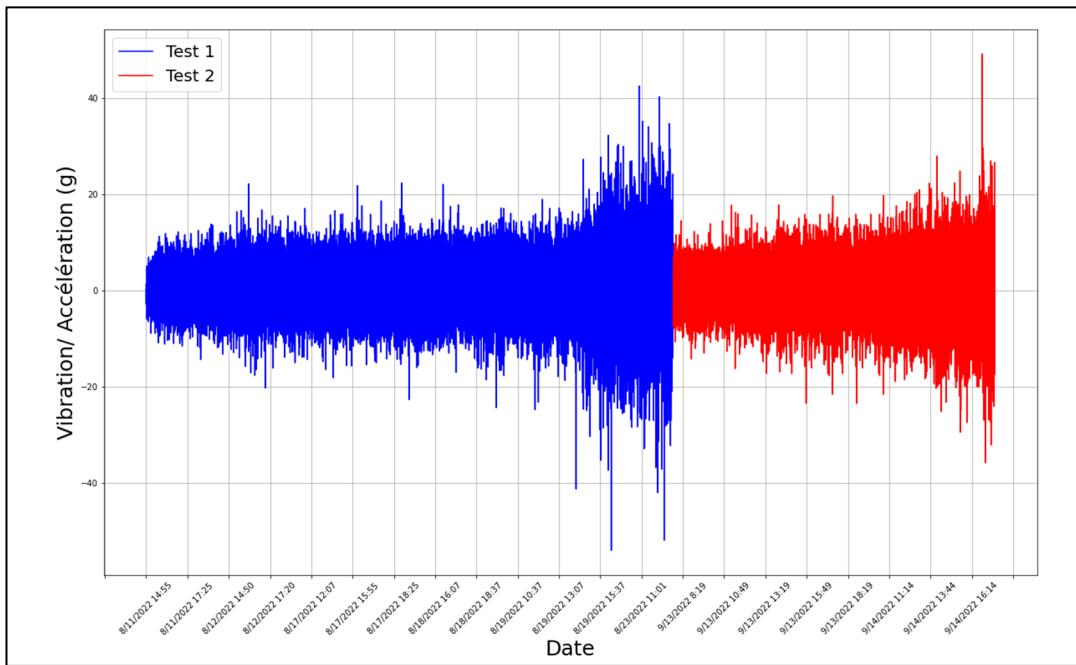


Figure 4.12 Signaux de vibrations enregistrés avec l'accéléromètre Adxl 1002Z

La figure 4.10 nous illustre les signaux enregistrés par le capteur Adxl 335 monté sur le premier roulement et la figure 4.11 présente les signaux enregistrés par le capteur Adxl1002Z monté sur le deuxième roulement. Les signaux bleus seront utilisés par la suite en tant que données d'entraînement et les signaux rouges en tant que données de test.

### 4.3 Analyse des Résultats

Dans cette section, les résultats sont présentés et discutés. Les résultats relatifs aux modèles et aux différentes mesures de performance calculées sont d'abord abordés, par la suite on va discuter la détection et la classification des anomalies.

A priori, pour donner suite aux résultats obtenus dans la première partie, nous avons choisi de mener les simulations sur mes données avec les caractéristiques statiques puis qu'ils ont donné les meilleurs résultats. Partant de ce fait les résultats étaient insuffisants et inattendus, pour ces raisons nous avons essayé de trouver la caractéristique la plus représentante à nos données et effectivement c'était la Racine moyenne carré. À cet égard, les résultats des simulations dans

les prochaines sections seront traités avec la racine moyenne carré des données enregistrés pendant les deux tests.

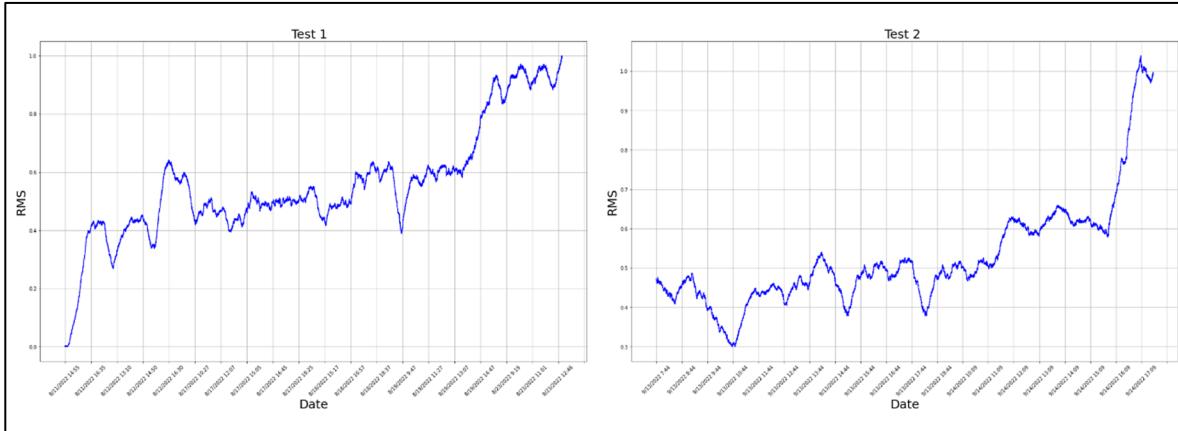


Figure 4.13 RMS des données des deux tests enregistrés par le capteur Adxl 335

Les deux figures 4.12 et 4.13 nous montrent les racines moyenne carrés des signaux enregistrés par les deux capteurs pendant les deux tests. On peut constater que l'allure des courbes des deux capteurs augmente à la fin des tests au moment de l'apparition de défaut.

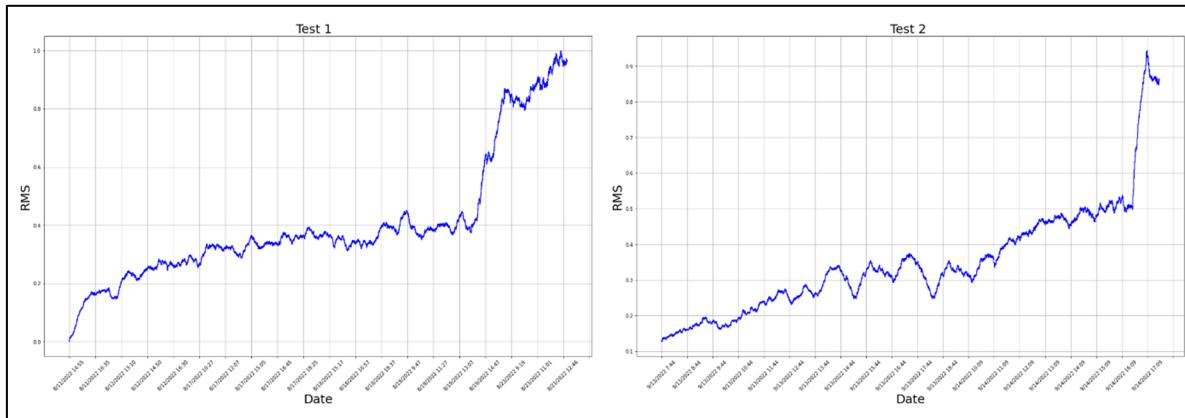


Figure 4.14 RMS des données des deux tests enregistrés par le capteur Adxl 1002Z

### 4.3.1 Régression

Pour donner suite aux résultats de régression obtenus dans la première partie dans la section 3.2.4.2, nous allons présenter les résultats avec nos propres données.

- Résultats avec la Racine moyenne carré :

Tableau 4.3 Résultats de régression avec les caractéristiques statiques

Les roulements	Roulement			Roulement		
	1			2		
Évaluation Modèle	R2 score	MAE	RMSE	R2 score	MAE	RMSE
<b>SVM</b>	74.92%	11.08	2.67	85.29%	8.09	2.4
<b>Xg_boost</b>	68.58%	12.83	2.78	<b>90.28%</b>	7	<b>2.2</b>
<b>LinearRegression</b>	57.66%	16.15	2.93	67.07%	14.41	2.81
<b>RandomForest</b>	56.45%	14.35	2.9	84.12%	8.48	2.44
<b>Ridge</b>	57.66%	16.16	2.9	67.06%	14.41	2.81
<b>Lasso</b>	51.82%	17.27	3	60.77%	15.87	2.89
<b>KNN</b>	62.08%	13.77	2.88	85.02%	8.21	2.41
<b>DecisionTree</b>	45.48%	15.5	3.06	82.33%	8.98	2.5
<b>ExtraTrees</b>	52.49%	14.79	2.99	83.46%	8.66	2.46
<b>AdaBoost</b>	58.81%	15.12	2.92	85.18%	8.1	2.39
<b>BaggingRegressor</b>	55.33%	14.47	2.96	83.92%	8.53	2.45
<b>Super learner</b>	62.27%	14.44	2.88	86.41%	7.93	2.36

Le tableau 4.3 nous affiche les résultats de régression obtenus par les modèles de ML entraînés avec les caractéristiques extraites à partir des données du premier test, et testés avec les caractéristiques extraites à partir des données du deuxième test. On constate que pour le premier capteur les 2 meilleurs modèles sont l'SVM et l'XG Boost avec un score R2 respectivement de 74.92% et 68.58% avec des erreurs minimales par rapport aux autres modèles de 2.67 et 2.78 pour l'RMSE et 11.08 et 12.83 pour le MAE. Les figures 4.14 et 4.15 nous montrent la prédiction de RUL du premier roulement par les deux modèles SVM et XG

Boost, qu'elle était mieux avec l'SVM qu'avec l'XG Boost. Ensuite, pour le deuxième capteur les résultats se sont améliorés par rapport aux résultats du premier capteur, on trouve que le score R2 a augmenté à 90.28% pour le modèle XG Boost, à 86.41% pour le Super learner et à 85.29% pour l'SVM, ainsi que pour les erreurs ont diminués 2.2, 2.36 et 2.4 pour le RMSE et 7, 7.93 et 8.09 pour le MAE.

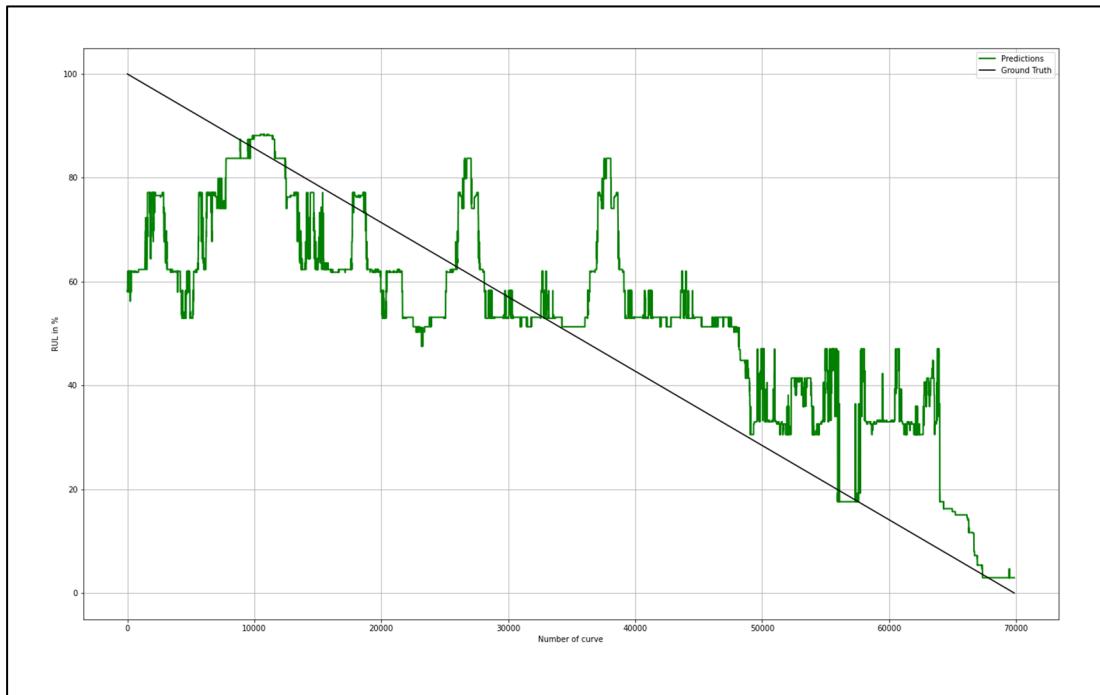


Figure 4.15 Prédition du RUL par le modèle XG Boost

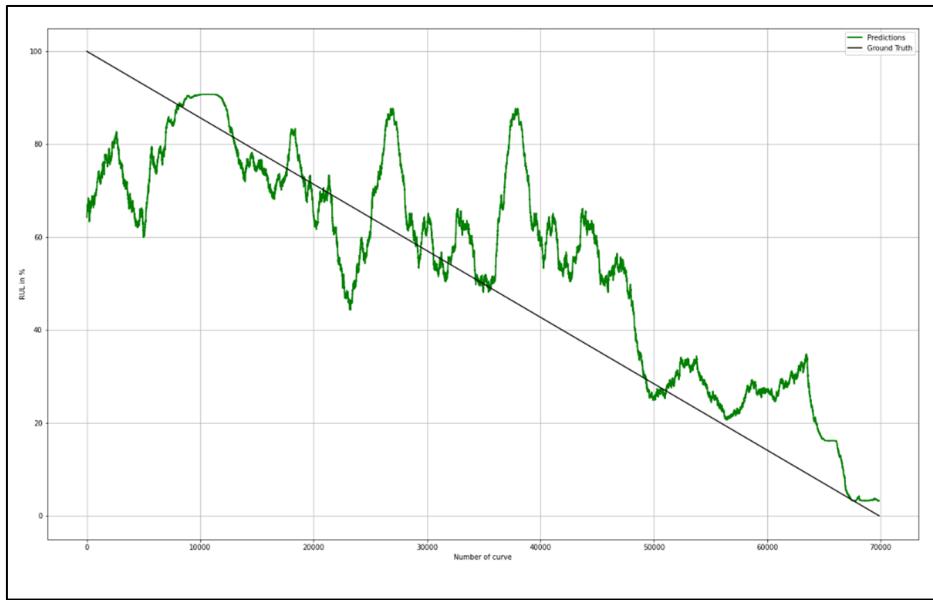


Figure 4.16 Prédiction du RUL par le modèle SVM

Les figures 4.16 et 4.17 nous montrent la prédition de RUL pour le deuxième roulement par les meilleurs trois modèles qui sont l'SVM et l'XG Boost. Les courbes étaient améliorées par rapport aux courbes de premiers roulements, on peut voir que les extrémités suivent notre cible avec des fluctuations au milieu.

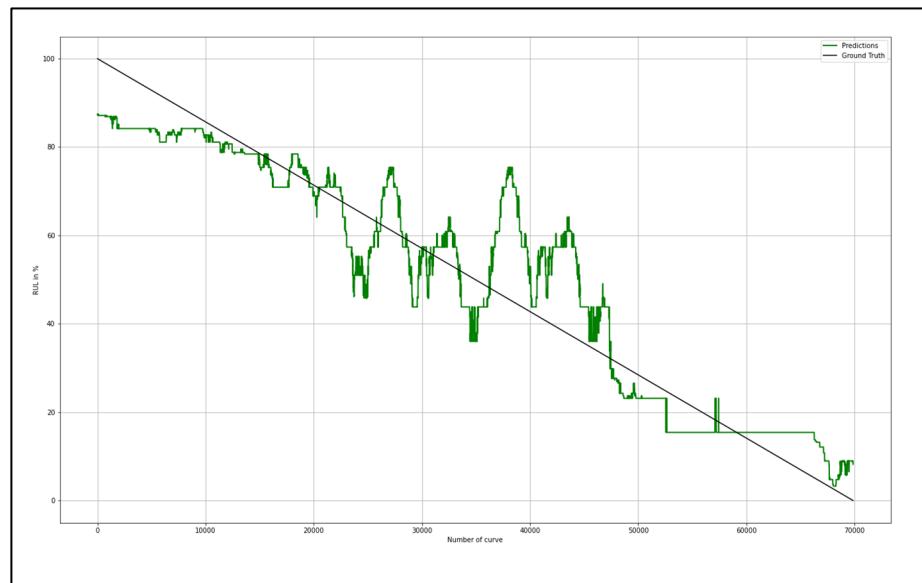


Figure 4.17 Prédiction du RUL par le modèle XG Boost

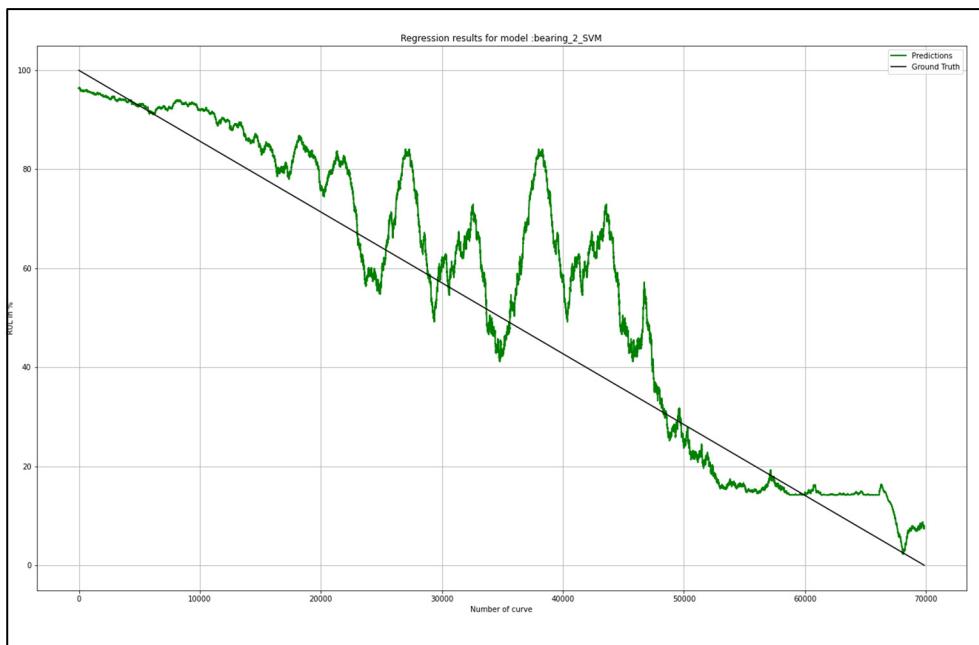


Figure 4.18 Prédiction du RUL par le modèle SVM

#### 4.3.2 Classification

Identiquement à la régression, nous allons présenter les résultats de classification avec nos propres données.

- Résultats avec les caractéristiques statiques :

Tableau 4.4 Résultats de classification avec les caractéristiques statiques (Suite)

Les roulements	Roulement 1			Roulement 2			
	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
Modèle							
Xg_boostClassifier		30.80%	32.92%	34.44%	33.28%	39.77%	37.26%
LogisticRegression		28.40%	30.72%	32.20%	<b>38.76%</b>	<b>48.46%</b>	<b>41.40%</b>

Les roulements	Roulement			Roulement		
	1			2		
Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
<b>KNNClassifier</b>	29.75%	33.48%	30.45%	34.50%	41.18%	37.81%
<b>MLPClassifier</b>	32.52%	33.19%	37.89%	36.25%	40.91%	40.10%
<b>ExtraTreesClassifier</b>	28.78%	32.97%	27.86%	33.60%	40.97%	36.86%
<b>BaggingClassifier</b>	28.94%	33.07%	28.09%	33.74%	41.06%	37.02%
<b>AdaBoost</b>	9.49%	7.14%	19.19%	5.46%	7.88%	13.37%
<b>GaussianNB</b>	24.16%	21.54%	29.75%	31.74%	42.23%	34.78%
<b>DecisionTreeClassifier</b>	28.77%	32.97%	27.85%	33.60%	40.9%	36.87%

La performance en classification des modèles est résumée dans le Tableau 4.4. Ces derniers n'étaient pas capables de bien classifier les données en 10 classes tandis que les résultats avec les données du deuxième capteur étaient mieux qu'avec les données du premier capteur. Les figures 4.19 et 4.20 représentent les matrices de confusion pour les deux meilleurs modèles pour les deux roulements.

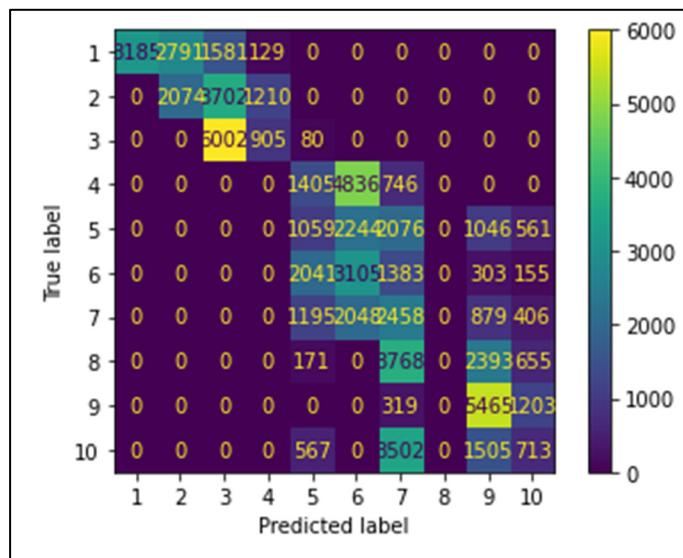


Figure 4.19 Matrice de confusion de l'XGB Classifier

Les deux matrices de confusion nous montrent que les modèles ne sont pas arrivés à bien classifier les données en 10 classes, néanmoins l'XGB Classifier est arrivé à classifier les données du 3ème et 9ème classe des données de l'Adxl 335, ainsi que pour le Logistic Regressor est arrivé à classifier les données de 2ème et 10ème classe, mais ils sont incapables de reconnaître convenablement les autres classes.

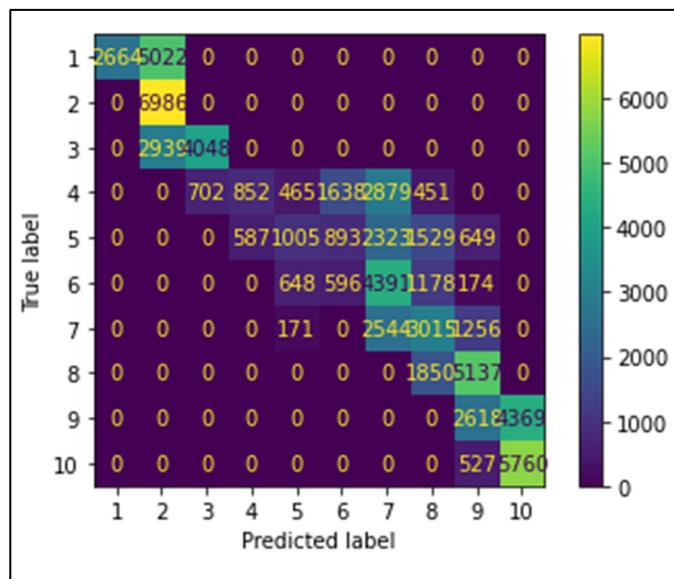


Figure 4.20 Matrice de confusion  
du Logistic Regressor

#### 4.3.3 Clustering

Troisièmement, en suivant la même démarche de la première partie, nous allons présenter les résultats de clustering des données de vibration pour les deux capteurs montés sur les deux roulements :

Tableau 4.5 Résultats de clustering avec les caractéristiques statiques

Les roulements	Roulement 1			Roulement 2			
	Évaluation	F-score	Précision	Exactitude	F-score	Précision	Exactitude
Modèle							
<b>Xg_boostClassifier</b>	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
<b>LogisticRegression</b>	98.72%	98.78%	98.76%	99.35%	99.36%	99.35%	99.35%
<b>RandomForestClassifier</b>	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
<b>KNNClassifier</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
<b>SVM</b>	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
<b>MLPClassifier</b>	99.83%	99.83%	99.83%	99.71%	99.72%	99.71%	99.71%
<b>ExtraTreesClassifier</b>	99.99%	99.99%	99.99%	100.00%	100.00%	100.00%	100.00%
<b>BaggingClassifier</b>	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
<b>AdaBoost</b>	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
<b>GaussianNB</b>	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
<b>DecisionTreeClassifier</b>	95.78%	96.47%	96.32%	97.87%	98.02%	97.83%	97.83%

Le Tableau 4.5 nous présente les résultats de performance des modèles utilisés pour faire la décomposition des signaux de vibrations en trois classes. Les modèles ont donné des résultats parfaits où ils sont arrivés à bien reconnaître les différentes phases des données.

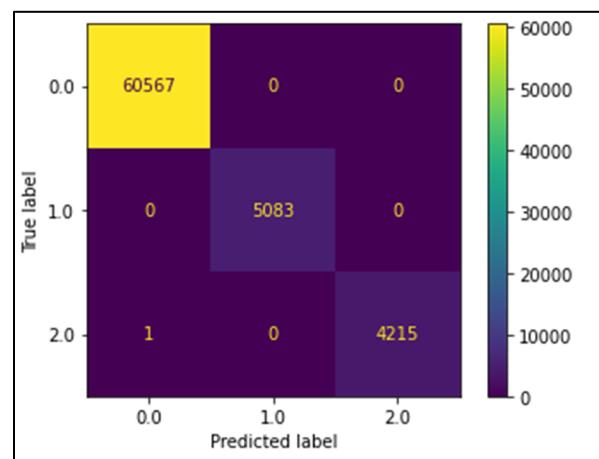


Figure 4.21 Matrice de confusion du Random Forest

Les figures 4.21 et 4.22 représentent les matrices de confusion de deux exemples des modèles pour les deux roulements, c'est clair qu'ils décomposent les données en trois phases, mais pour mieux montrer les résultats, nous allons afficher la décomposition des données sur les courbes des données.

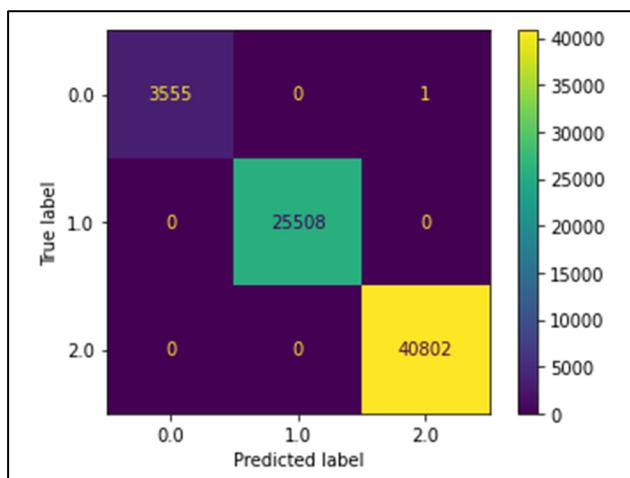


Figure 4.22 Matrice de confusion du KNN

La figure 4.20 nous illustre les signaux des deux tests enregistrés avec l'accéléromètre Adxl 335 décomposés en 3 phases, les données du premier test étaient plus stables que celles du deuxième test pour cela le modèle de clustering les a mieux décomposés.

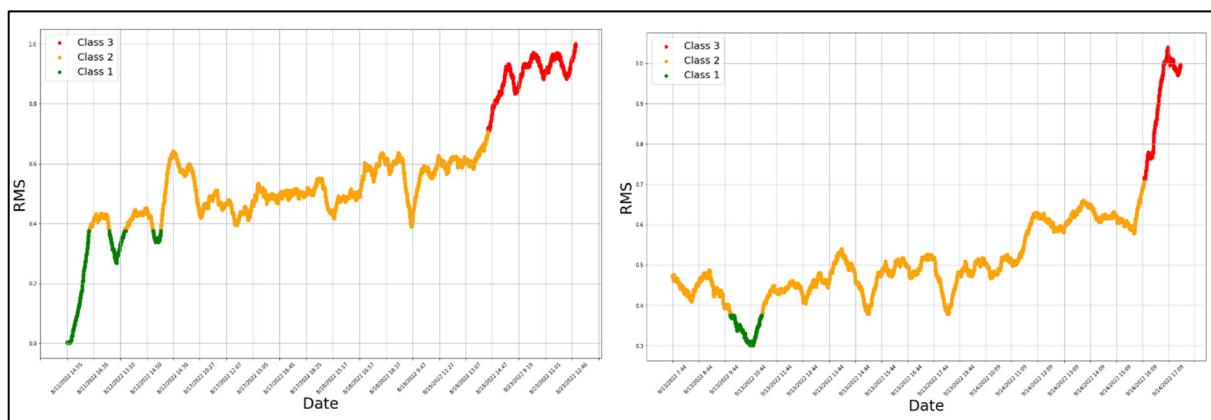


Figure 4.23 Clustering des données de l'Adxl 335 des deux tests

Les couleurs des courbes dans la figure 4.23 nous indiquent l'état du roulement, où la couleur verte présente l'état normal, la couleur jaune présente une perturbation de vibration et en fin la couleur rouge nous indique l'apparition de la panne.

La figure 4.24 nous affiche les signaux des deux tests enregistrés avec l'accéléromètre Adxl 1002Z décomposés en 3 phases par le modèle de clustering, non seulement que les données qui proviennent de ce capteur sont plus stable, mais aussi le modèle a bien décomposé les signaux en 3 phases selon l'évolution des vibrations au cours du temps, de sorte que la prédiction de RUL est plus pertinente avec ce capteur.

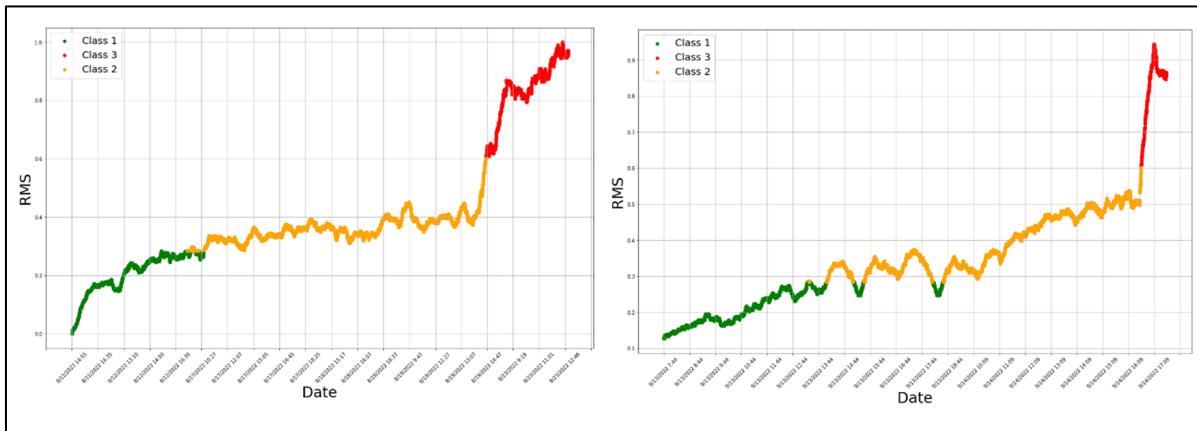


Figure 4.24 Clustering des données de l'Adxl 1002Z des deux tests

#### 4.3.4 Détection des anomalies et leurs classifications

Dans cette partie, nous allons présenter les résultats de la détection des anomalies et leur classification.

##### 4.3.4.1 Détection des anomalies

L'objectif de cette partie est de détecter l'anomalie au début du changement des signaux de vibrations. On va explorer l'habileté de l'architecture Auto encodeur AE utilisée pour la détection des défauts, selon (Nguyen et al., 2021) l'auto-encodeur est un type de réseau neuronal non supervisé qui utilise des données pour former la meilleure technique d'encodage-

décodage. Il se compose d'une couche d'entrée, d'une couche de sortie, d'un réseau neuronal encodeur, d'un réseau neuronal décodeur et d'un espace latent en général. Lorsque des données sont fournies au réseau, l'encodeur les compresse dans l'espace latent, tandis que le décodeur les décomprime dans la couche de sortie. La sortie codée-décodée est ensuite comparée aux données originales, et l'erreur est propagée à travers l'architecture pour mettre à jour les poids du réseau. L'architecture utilisée dans notre projet est illustrée dans la figure 4.25, elle est décomposée par une couche d'entrée, une partie pour l'encodage des données, un espace latent, une partie pour le décodage des données et une couche de sortie. (Nguyen et al., 2021)

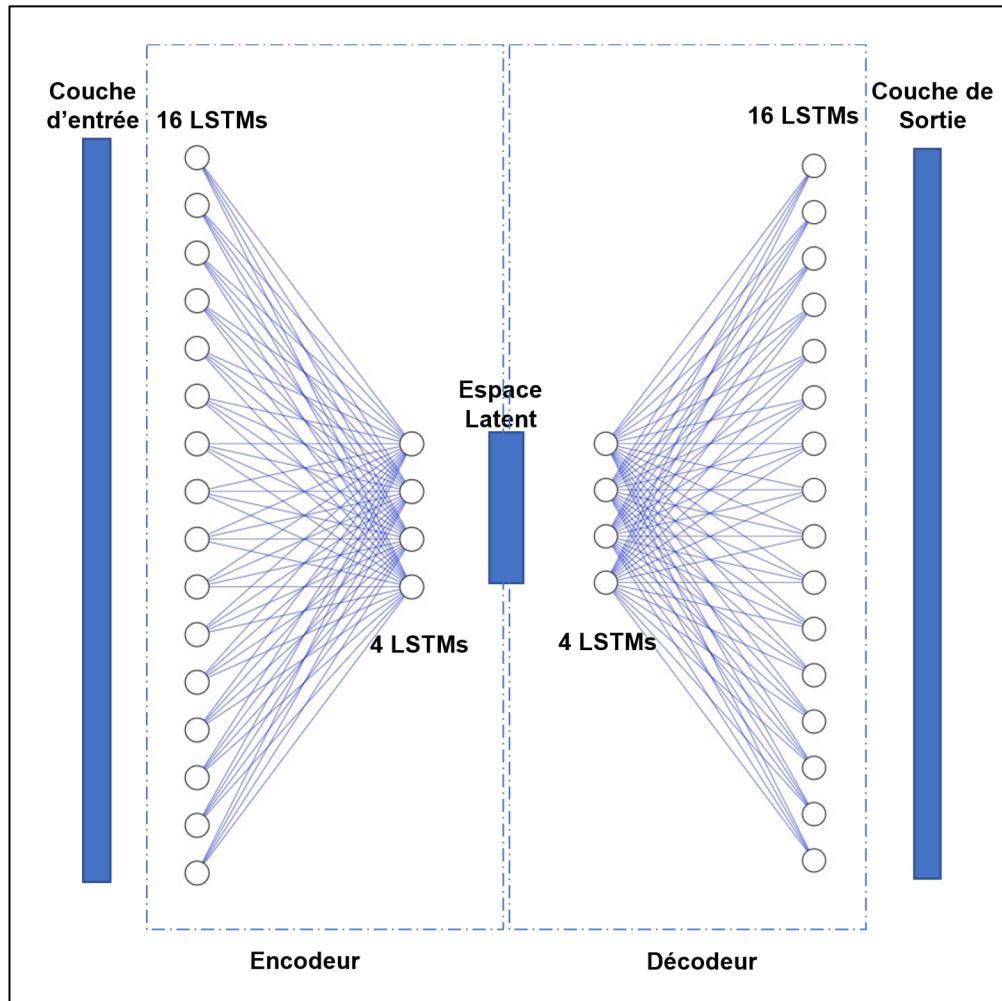


Figure 4.25 Structure de l'AE

Les parties de l'encodeur et le décodeur sont basée sur les couches LSTM, ces dernières sont aujourd'hui l'un des modèles de séquences temporelles les plus utilisés pour leurs efficacités

dans les applications pratiques. Par exemple, (Maleki et al., 2021) ont utilisé un AE basé sur des couches LSTMs pour la détection des anomalies dans des scénarios industriels réels.

- **Résultat**

Dans cette partie, nous allons vérifier la performance de l'architecture utilisée pour la détection des anomalies dans les signaux de vibrations des roulements enregistrés dans le laboratoire. L'entraînement du modèle est fait par la partie saine des données du premier test, et la validation sera faite par la partie qui présente l'anomalie.

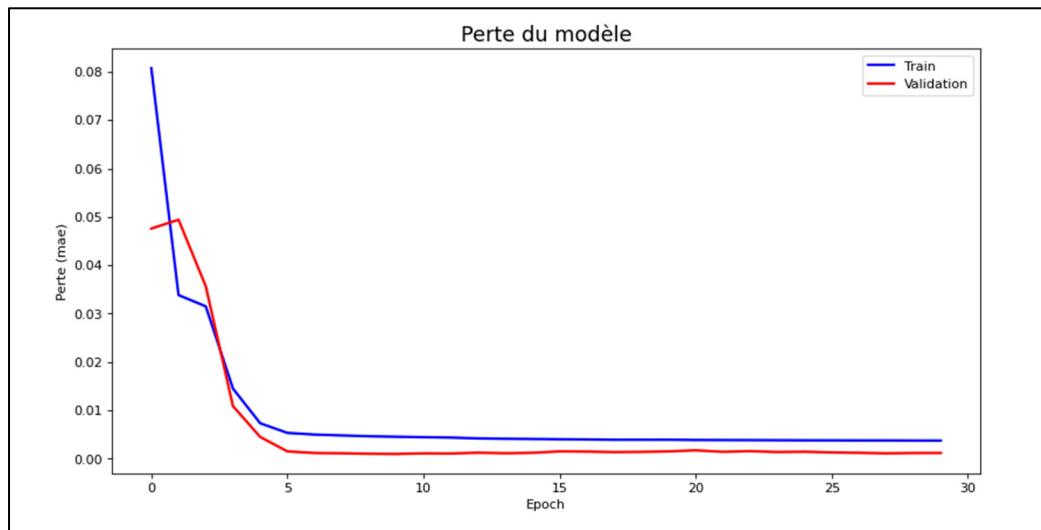


Figure 4.26 Courbe de la perte d'entraînement et de validation du modèle proposé

La courbe 4.26 nous montre que le modèle a bien compris les données d'entrée à partir de la 5<sup>ème</sup> epoch, en atteignant une valeur de perte de validation égale à 0.0012 MAE.

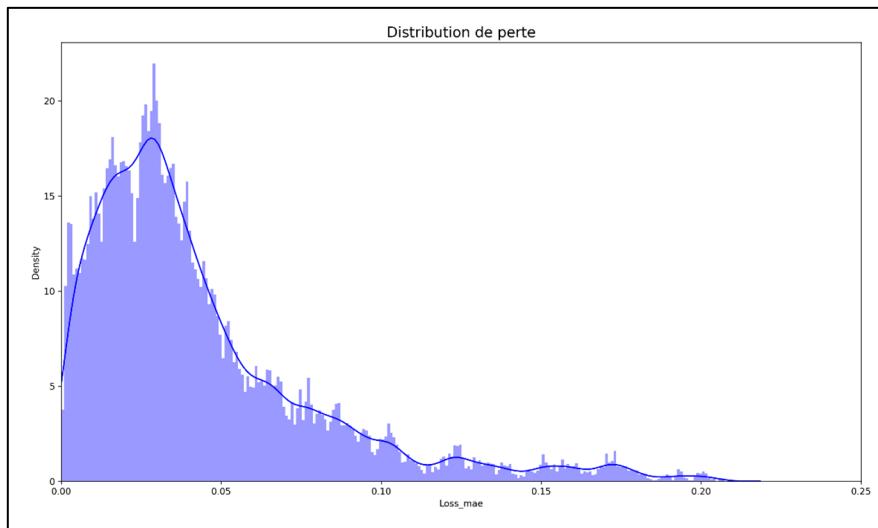


Figure 4.27 Distribution de la perte

La courbe de la figure 4.27 nous présente la distribution de la perte entre le signal d'entrée et le signal prédit, et à partir de cette courbe on peut sélectionner une valeur limite de perte pour la détection des anomalies qui sera  $K = 0.22$ .

Par la suite pour valider le modèle, nous devons le tester avec le signal entier pour vérifier sa capacité à détecter le défaut.

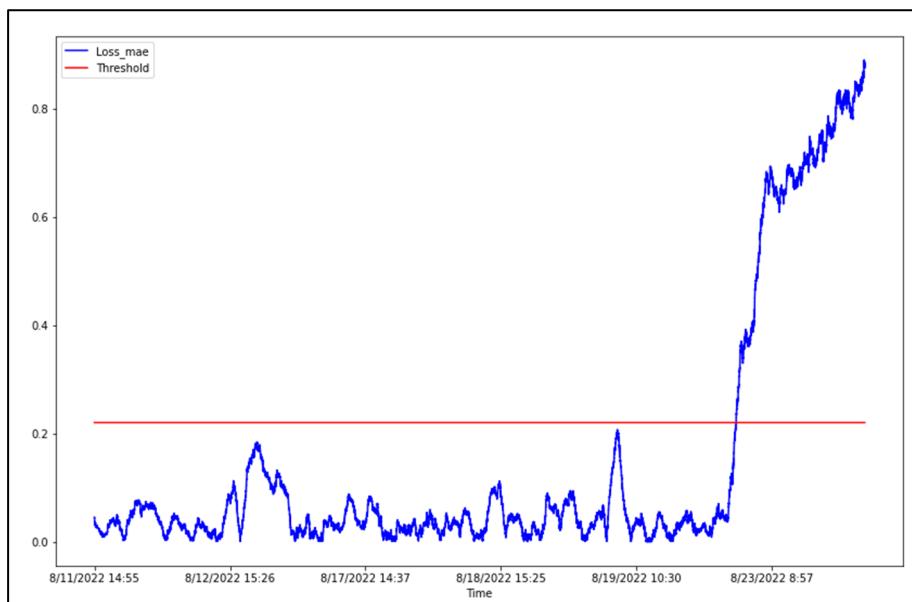


Figure 4.28 Détection d'anomalie Test1

La figure 4.28 nous présente que le modèle est arrivé à détecter l'anomalie au moment de dépassement de la limite rouge à la suite du changement des signaux de vibration provient des capteurs. La deuxième phase est de tester avec les données du deuxième test et s'assurer de la validité du modèle.

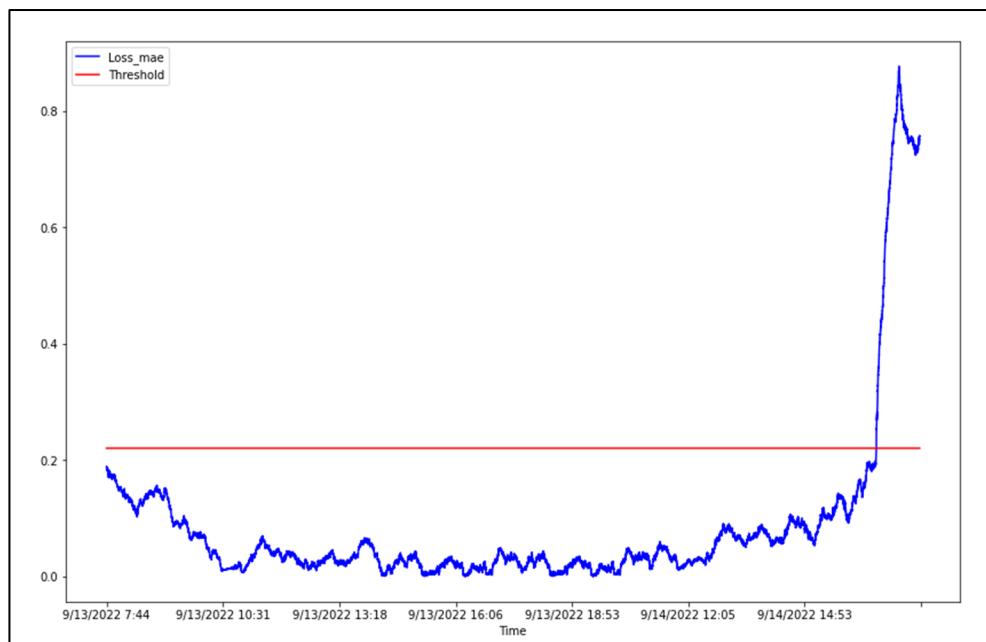


Figure 4.29 Détection d'anomalies Test2

Le résultat illustré dans la figure 4.29 nous confirme la performance du modèle est sa capacité à détecter les anomalies. La courbe de perte entre le signal d'entrée et le signal prédict a dépassé la limite ce qui nous indique automatiquement l'apparition de la panne.

#### 4.3.4.2 Classification de l'anomalie

Dans cette section, nous allons présenter les résultats de classification des anomalies des deux roulements, comme on a mentionné dans la section 4.2.4, on a un seul type de défaut effectué par un perçage dans les éléments roulants des roulements, pour cela les données seront divisées en deux parties après la détection, une avant l'effectuation de défaut et l'autre après le perçage afin de les utiliser pour évaluer les performances de classification des modèles étudiés.

Le tableau 4.6 nous présente les résultats de classification de tous les modèles pour le premier capteur Adxl 335 :

Tableau 4.6 Performance des modèles pour Adxl335

Modèle	F-Score	Précision	Exactitude	F-Score Classe 0	F-Score Classe 1
<b>Logistic regressor</b>	96.81%	96.85%	96.77%	98%	75%
<b>KNN</b>	97.28%	97.26%	97.3%	99%	78%
<b>Random Forest</b>	97.97%	97.97%	98.05%	99%	83%
<b>Extra Trees</b>	95.41%	95.9%	95.1%	97%	67%
<b>Bagging regressor</b>	95.86%	96.17%	95.65%	98%	69%
<b>Adaboost</b>	97.98%	97.98%	98.05%	99%	83%
<b>Décision Trees</b>	95.46%	95.93%	95.17%	97%	67%
<b>Xg boost</b>	<b>97.98%</b>	<b>97.99%</b>	<b>98.06%</b>	<b>99%</b>	<b>83%</b>
<b>SVM</b>	97.97%	97.98%	98.05%	99%	83%

Les modèles de classification sont bien arrivés à classifier nos données en deux classes soit un échantillon de défaut ou un échantillon qui appartient aux données normales, pour le modèle Extra trees la matrice résultante est présenté dans la figure 4.30, le modèle à un F-score générale de 95.41% mais le F-score du classe 1 est égale à 67% qui est fiable par rapport aux

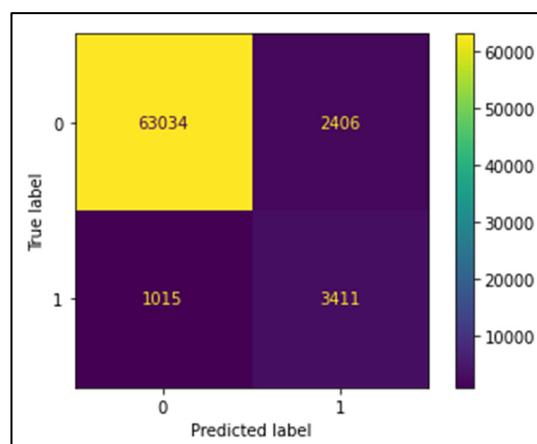


Figure 4.30 Matrice de confusion de l'Extra Trees

autres modèles affichés dans le tableau 4.6, prenant l'exemple du modèle Xg boost sa matrice résultante est présenté dans la figure 4.31, C'est le modèle qui a le plus haut F-score générale

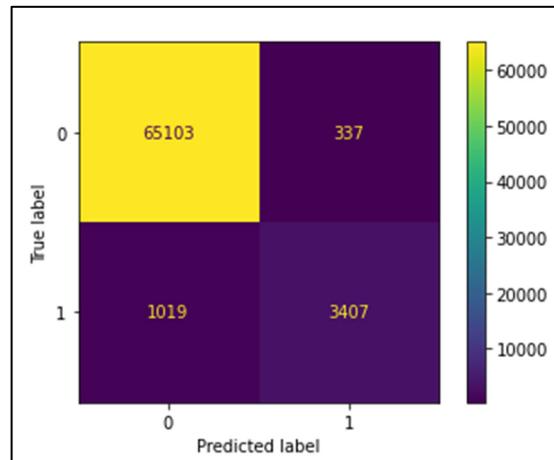


Figure 4.31 Matrice de confusion de l'Xg boost

de 97.98%, ainsi que les F-score respective des classe 0 et 1 sont égales à 99% et 83 %.

Le tableau 4.7 nous présente les résultats de classification de tous les modèles pour le deuxième capteur Adxl1002Z :

Tableau 4.7 Performance des modèles pour Adxl1002Z (Suite)

Modèle	F-Score	Précision	Exactitude	F-Score	F-Score
				Classe 0	Classe 1
<b>Logistic regressor</b>	98.11%	98.13 %	98.19 %	99%	84%
<b>KNN</b>	96.77%	96.82 %	96.72 %	98 %	75 %
<b>Random Forest</b>	96.78%	96.82 %	96.74 %	98 %	75 %
<b>Extra Trees</b>	96.66%	96.73 %	96.6 %	98 %	74 %
<b>Bagging regressor</b>	96.74%	96.79 %	96.69 %	98%	75%

Modèle	F-Score	Précision	Exactitude	F-Score Classe 0	F-Score Classe 1
<b>Adaboost</b>	97.01%	97.02 %	97%	98 %	76 %
<b>Décision Trees</b>	96.65%	96.72 %	96.59 %	98 %	74 %
<b>Xg boost</b>	96.99%	97 %	96.98 %	98 %	76 %
<b>SVM</b>	<b>98.26%</b>	<b>98.31 %</b>	<b>98.34 %</b>	<b>99%</b>	<b>85 %</b>

Les modèles de classification sont bien arrivés à classifier nos données, la matrice résultante du modèle Décision Trees est illustrée dans la figure 4.32, il présente le F-score le plus bas de 96.65% avec un F-score des classes 0 et 1 respectivement 98 % et 74 %.

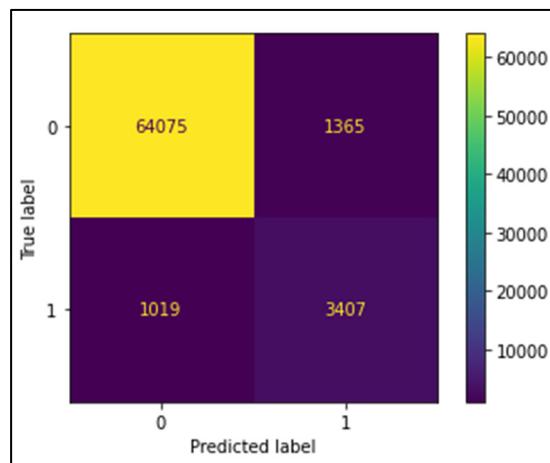


Figure 4.32 Matrice de confusion de Décision Trees

L’SVM est le meilleur Modèle pour les données des deux capteurs en se basant sur les résultats affichés dans les deux tableaux 4.6 et 4.7, sa matrice résultante est illustrée dans la figure 4.33, il présente le F-score de 98.26% avec un F-score des classes 0 et 1 respectivement 99 % et 85 %.

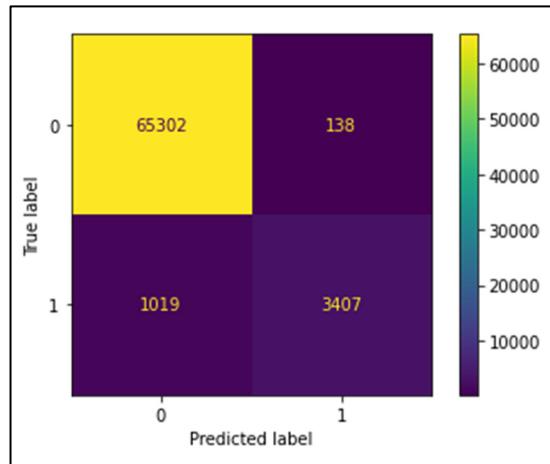


Figure 4.33 Matrice de confusion de l'SVM

#### 4.4 Analyse des résultats

Dans cette section nous allons comparer les résultats obtenus des deux capteurs.

- **Résultats de régression :**

La figure 4.34 nous illustre l'histogramme de variation du score R2 des deux roulements, on constate que les scores R2 obtenu par la bibliothèque des algorithmes avec les données du capteur Adxl 1002Z sont toujours mieux que ceux obtenus avec les données enregistrées par le capteur Adxl 335.

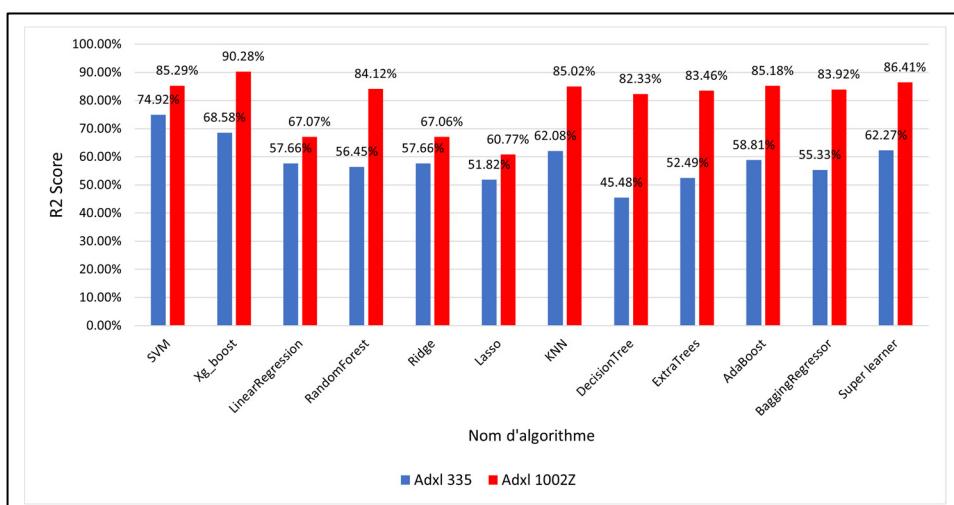


Figure 4.34 Variation du R2 score des deux roulements

Ainsi, la figure 4.35 nous montre la variation des erreurs racine moyenne carré et moyenne absolue carré qui sont proportionnels pour les deux capteurs, mais l'Adxl 1002Z a toujours des valeurs moins grandes que les erreurs du l'Adxl 335. Ce qui nous valide la supériorité de l'adxl 1002Z.

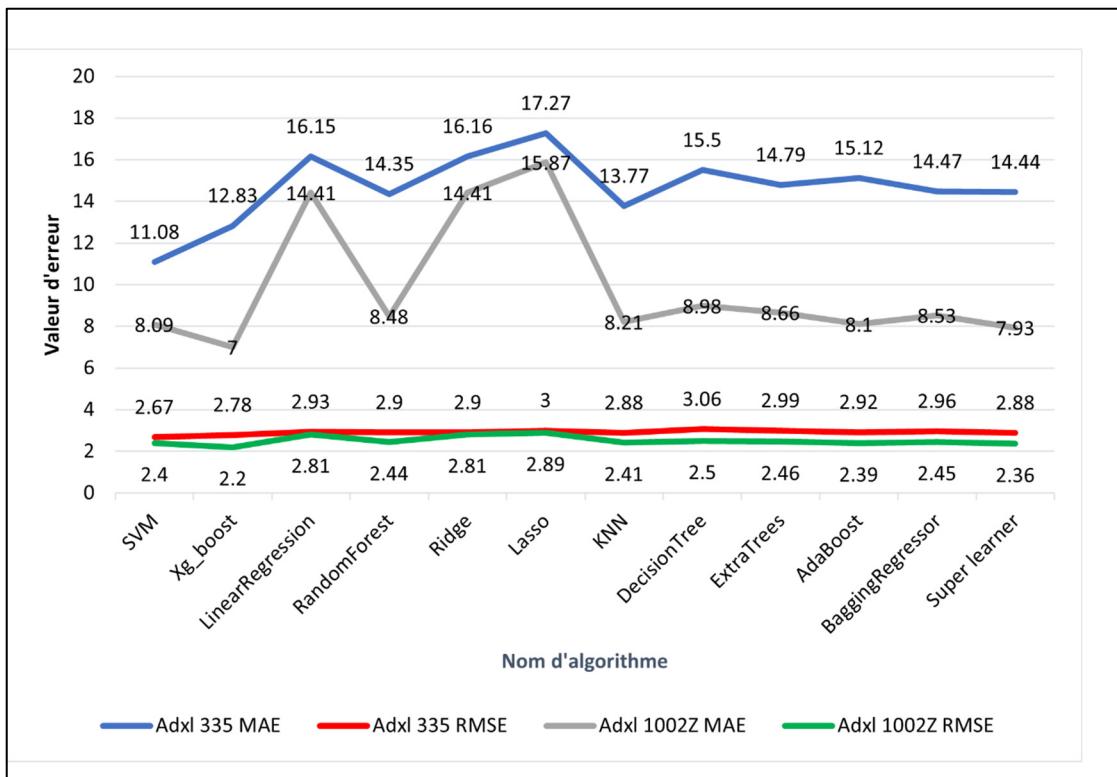


Figure 4.35 Variation des erreurs des algorithmes

- **Résultats de Classification :**

Les résultats de classification n'étaient pas bons pour les deux capteurs comme c'est affiché dans la figure 4.36, toutefois les F-score de classification obtenu avec les données obtenues par le capteur Adxl 1002Z sont un peu plus grand que celles avec le capteur Adxl 335 en atteindront un maximum de 38.76% par l'algorithme Logistic Regression, avec une supériorité pour l'adxl1002Z avec la plupart des modèles de classification.

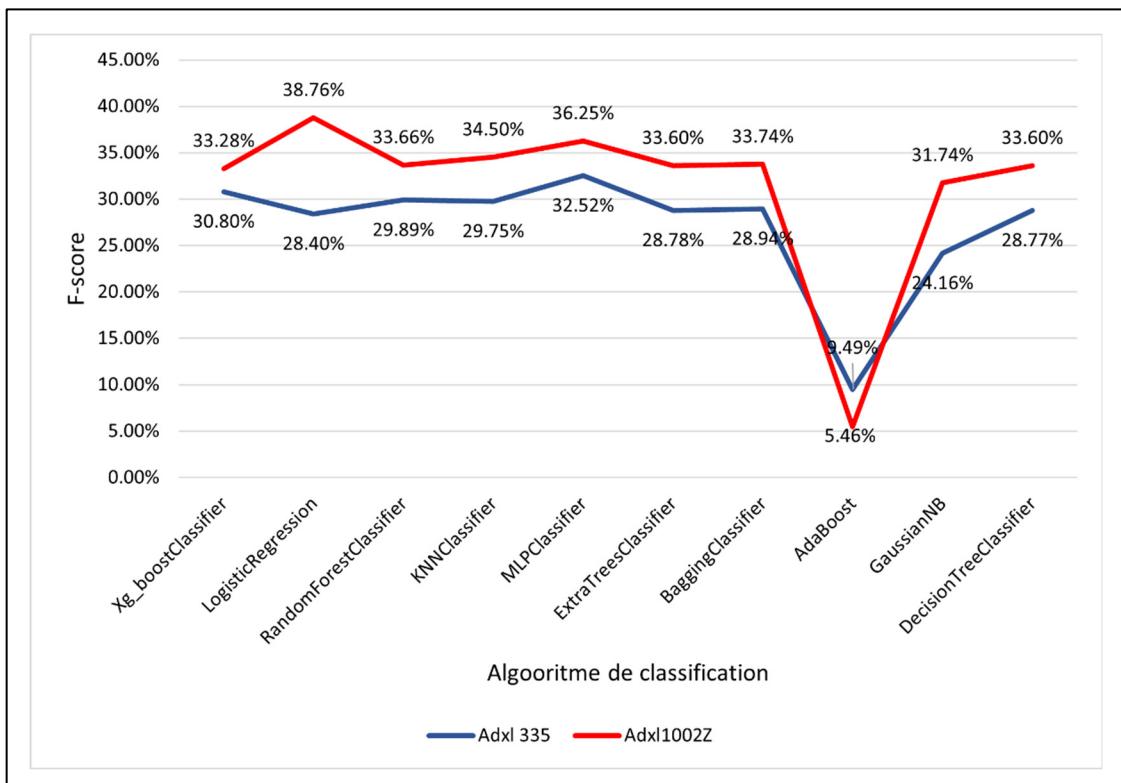


Figure 4.36 Variation du F-score de classification

- **Résultats de Clustering :**

La figure 4.37 nous illustre les résultats de clustering, la totalité des algorithmes ont eu un très bon score pour les deux capteurs. Les F-scores des deux capteurs varient entre 95.78 % et 100%, la plupart des algorithmes ont eu un score de 100% sauf le Logistic regressor, l'MLP Classifier et le Decision tree. Ces résultats nous montrent l'efficacité du clustering pour ce type de problème et que les deux capteurs ont donné des bons résultats, avec une supériorité pour l'adxl1002Z avec les deux modèles DecisionTreeClassifier et LogisticRegression.

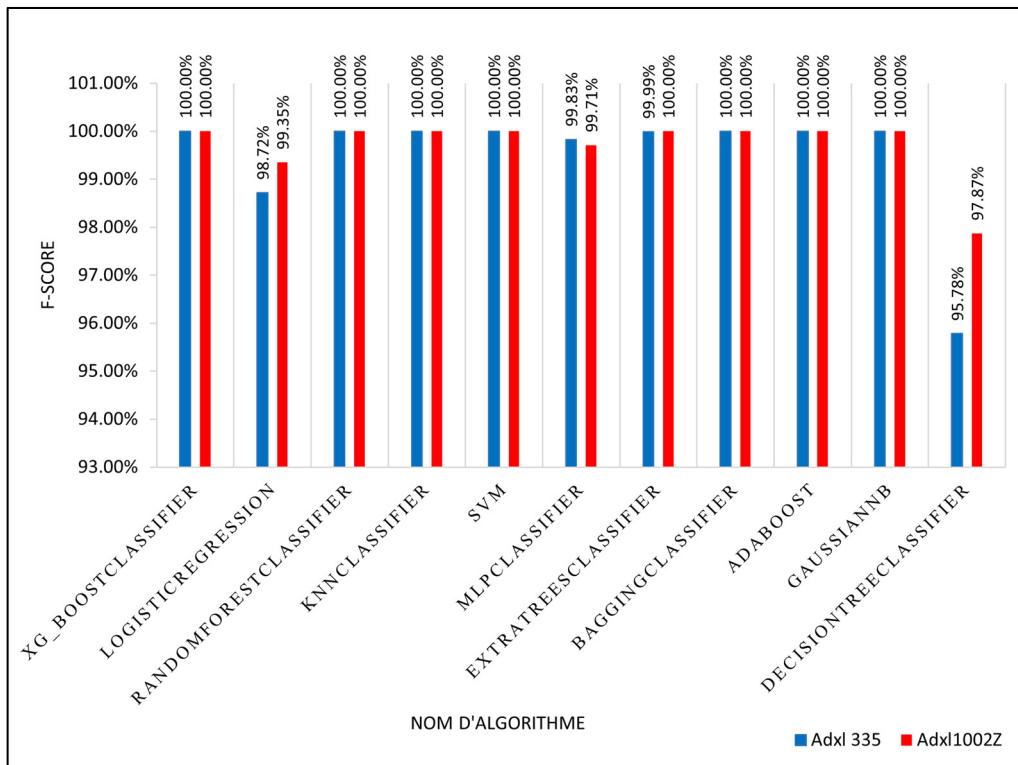


Figure 4.37 Variation du F-score de clusering

- **Résultats de détection et classification des anomalies :**

La détection des anomalies a été faite avec l'architecture d'AE qui est arrivé à détecter les anomalies pour les deux tests, par la suite la classification des données en échantillons normaux et échantillons d'anomalies est performé, les résultats sont illustrés dans la figure 4.38, les modèles sont arrivés à bien les classifier avec un F-score maximal de 98.26% par l'algorithme SVM.

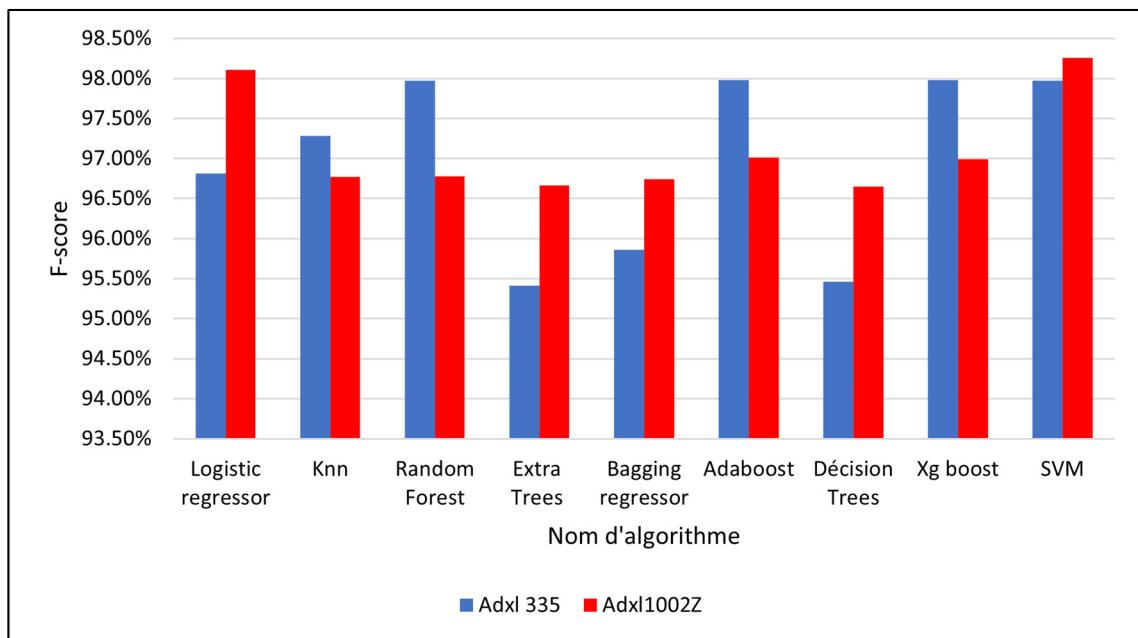


Figure 4.38 Variation du F-score de classification des anomalies

- Cette comparaison nous amène à conclure que la performance de l'Adxl 1002Z est supérieure à l'Adxl 335 pour l'utilisation dans un contexte de maintenance prédictive pour la régression, classification, clustering et aussi pour la détection et la classification des anomalies puisque l'adxl 1002Z détient des caractéristiques plus puissantes que l'adxl335 comme nous l'avons montré dans le paragraphe 4.2.1, néanmoins, l'Adxl 335 a des bonnes performances comparables à celles de l'Adxl1002Z, et qui peut être utilisé aussi pour son prix, étant le moins chère dans le marché.

## 4.5 Conclusion

Dans ce chapitre, nous avons fait la description du banc de test conçu pour faire les expériences dans notre laboratoire. Par la suite, la même démarche de la première partie a été adapté sur nos propres données de vibrations enregistrés par les accéléromètres Adxl 335 et Adxl 1002Z. Les résultats nous ont montrés que ces derniers sont capables à capter des données de vibrations qui peuvent être utilisés pour estimer l'état de santé des roulements, pas seulement ça, mais

aussi de les utiliser pour détecter les anomalies et faire leurs classifications selon leurs types, et ce avec une supériorité pour l'accéléromètre Adxl 1002Z.

## CONCLUSION

Cette recherche se concentre sur l'exploration de l'utilisation des MEMS accéléromètres pour effectuer la surveillance de l'état des roulements. Elle devrait permettre de réduire de manière significative les coûts de maintenance des machines industrielles. Les vibrations ont été choisies comme paramètre de mesure parce qu'elles se sont avérées sensibles et efficaces pour divers types de défauts de machines. En tant que composant important des machines tournantes, les roulements ont été largement utilisés dans le secteur industriel et représentent également un pourcentage élevé des défauts des machines. Pour cette raison, le roulement a été choisi comme cible de l'étude dans ce projet.

Selon l'enquête faite sur les MEMS accéléromètres disponibles dans le marché, l'Adxl 335 est un bon candidat pour la surveillance des conditions industrielles en raison de ses caractéristiques inhérentes telles que l'intervalle d'accélération capté, son faible coût, sa faible sensibilité et sa capacité à résister et fonctionner dans des conditions dures. Il convient parfaitement à la mesure des données dans une usine. De surcroit, un autre MEMS accéléromètre Adxl 1002Z a été sélectionné pour ses capacités supérieures à celle du premier et aussi pour qu'il a été conçu pour la maintenance prédictive.

Une première partie du projet a été consacrée pour faire une étude d'évaluation de la performance des modèles d'intelligence artificielles envers l'estimation de l'état de santé des roulements, la détection des anomalies et leur classification à travers une base de données open source. Par ailleurs, la même démarche est appliquée sur les données de vibrations collectées via les capteurs MEMS sélectionnés. Au cours de cette partie on est arrivé à estimer le RUL de roulements avec un R2 score de 72.85% par l'algorithme Super Learner, un score F1 de classification des données en 10 classe égale à 75.02% par l'algorithme RandomForest Classifier, un F-score de 98.89% pour le clustering des données en 3 classes et un F-score de 97.42% pour la classification des anomalies après leur détection. Ensuite une analyse par comparaison a été faite sur les résultats des deux parties, cette comparaison a montré que les capteurs testés en combinaison avec l'approche d'apprentissage machine sont capable

d'estimer le RUL des roulements avec un R2 score de 90.28%, et aussi de faire le clustering des données avec un F-score de 100%, notamment on est arrivé à détecter l'anomalie dès son déclenchement et puis la classifier avec un score F égale à 98.26%.

En définitive, les résultats expérimentaux montrent que les défauts de roulement simulés peuvent être diagnostiqués avec succès à partir des données enregistrées par les MEMS qu'on a investigué, ainsi que leurs états de santé avec une performance supérieure à l'accéléromètre Adxl1002Z.

## LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

Adli, B., & Rusmin, P. H. (2020). Vibration Measuring Tools for Rotary Pumping Machine with Accelerometer MEMS Sensor. *Proceeding - 1st FORTEI-International Conference on Electrical Engineering, FORTEI-ICEE 2020*, 69–74. doi: 10.1109/FORTEI-ICEE50915.2020.9249860

Ag, S. (2018). *IN*.

Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*, 203, 810–821. doi: 10.1016/j.jclepro.2018.08.207

Ali, M. Z., John's Newfoundland, S., & Canada, L. (2019). *Induction Motors Fault Diagnosis Using Machine Learning and Advanced Signal Processing Techniques*. October.

Arun Faisal, I., Waluyo Purboyo, T., & Siswo Raharjo Ansori, A. (2019). A Review of Accelerometer Sensor and Gyroscope Sensor in IMU Sensors on Motion Capture. *Journal of Engineering and Applied Sciences*, 15(3), 826–829. doi : 10.36478/jeasci.2020.826.829

Buchaiah, S., & Shakya, P. (2022). Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. *Measurement: Journal of the International Measurement Confederation*, 188(December 2021), 110506. doi: 10.1016/j.measure.2021.110506

Cachada, A., Barbosa, J., Leitão, P., Geraldes, C. A. S., Deusdado, L., Costa, J., Teixeira, C., Teixeira, J., Moreira, A. H. J., Moreira, P. M., & Romero, L. (2018). Maintenance 4.0: Intelligent and Predictive Maintenance System Architecture. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, 2018-September*, 139–146. doi: 10.1109/ETFA.2018.8502489

Cameron, A. C., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2), 329–342. doi: 10.1016/s0304-4076(96)01818-0

Campos, J., Sharma, P., Albano, M., Ferreira, L. L., & Larrañaga, M. (2020). An open source framework approach to support condition monitoring and maintenance. *Applied Sciences (Switzerland)*, 10(18). doi: 10.3390/APP10186360

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August*, 785–794. doi: 10.1145/2939672.2939785

Chen, X., Tian, G., Ding, S., & Wu, J. (2019). Investigation of skewness feature for evaluation of defects using eddy current pulsed thermography. *IEEE Sensors Journal*, 19(24), 12118–12125. doi: 10.1109/JSEN.2019.2936221

Djebili, O. (2013). *Contribution à la maintenance prédictive par analyse vibratoire des composants mécaniques tournants . Application aux butées à billes soumises à la fatigue de contact de roulement .* 156. <http://www.theses.fr/2013REIMS030.pdf>

Exploration, M., Based, R. B., Sensor, W., Original, N., Aliwan, C., Based, R. B., Network, W. S., Access, O., & Rights, M. (2016). *System for Rolling Bearing Based on.*

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. doi : 10.1006/jcss.1997.1504

Friedman, J. (2001). Greedy Function Approximation : A Gradient Boosting Machine Author ( s ) : Jerome H . Friedman Source : The Annals of Statistics , Vol . 29 , No . 5 ( Oct . , 2001 ), pp . 1189-1232 Published by : Institute of Mathematical Statistics Stable URL : <http://www.jstor.org/stable/2699986> <http://www.annalsofstatistics.org/2001/29/5/1189>

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. doi: 10.1007/s10994-006-6226-1

Han, T., Zhang, L., Yin, Z., & Tan, A. C. C. (2021). Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. *Measurement: Journal of the International Measurement Confederation*, 177(July 2020), 109022. doi : 10.1016/j.measurement.2021.109022

ISO 17757. (2019). INTERNATIONAL STANDARD iTeh STANDARD iTeh STANDARD PREVIEW. *International Organization for Standardization*, 10406-1:20, 3–6.

Izza, Y., Ignatiev, A., & Marques-Silva, J. (2020). *On Explaining Decision Trees.* <http://arxiv.org/abs/2010.11034>

Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12343 LNCS, 503–515. doi: 10.1007/978-3-030-62008-0\_35

Kardani, N., Zhou, A., Nazem, M., & Shen, S. L. (2020). Estimation of Bearing Capacity of Piles in Cohesionless Soil Using Optimised Machine Learning Approaches. *Geotechnical and Geological Engineering*, 38(2), 2271–2291. doi: 10.1007/s10706-019-01085-8

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585, 609–629. doi: 10.1016/j.ins.2021.11.036

Kavana, V., & Neethi, M. (2018). Fault Analysis and Predictive Maintenance of Induction Motor Using Machine Learning. *3rd International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques, ICEECCOT 2018, December*, 963–966. doi: 10.1109/ICEECCOT43722.2018.9001543

Kavitha, S., Joseph Daniel, R., & Sumangala, K. (2016). Design and Analysis of MEMS Comb Drive Capacitive Accelerometer for SHM and Seismic Applications. In *Measurement: Journal of the International Measurement Confederation* (Vol. 93). doi: 10.1016/j.measurement.2016.07.029

Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*. doi: 10.1109/ICCCNT.2013.6726842

Khiari, J., & Olaverri-Monreal, C. (2020). Boosting Algorithms for Delivery Time Prediction in Transportation Logistics. *IEEE International Conference on Data Mining Workshops, ICDMW, 2020-Novem*, 251–258. doi: 10.1109/ICDMW51313.2020.00043

Korkua, S., Jain, H., Lee, W. J., & Kwan, C. (2010). Wireless health monitoring system for vibration detection of induction motors. *Conference Record - Industrial and Commercial Power Systems Technical Conference*, 21–26. doi: 10.1109/ICPS.2010.5489899

Kumar, P., & Hati, A. S. (2021). Deep convolutional neural network based on adaptive gradient optimizer for fault detection in SCIM. *ISA Transactions*, 111, 350–359. doi: 10.1016/j.isatra.2020.10.052

Lamraoui, M., Thomas, M., El Badaoui, M., & Zaghbani, I. (2011). New Indicators Based on Cyclostationarity Approach for Machining Monitoring. *Surveillance 6, November 2014*, 1–27. [http://jean.fabri.perso.sfr.fr/sfm/papiers\\_surveillance/29\\_Lamraoui.pdf](http://jean.fabri.perso.sfr.fr/sfm/papiers_surveillance/29_Lamraoui.pdf)

Larranaga, M., Salokangas, R., Saarela, O., & Kaarmila, P. (2020). Low-Cost Solutions for Maintenance with a Raspberry Pi. *30th European Safety and Reliability Conference, ESREL 2020 and 15th Probabilistic Safety Assessment and Management Conference, PSAM 2020, June*, 3400–3406. doi: 10.3850/978-981-14-8593-0

Li, R., Ran, C., Luo, J., Feng, S., & Zhang, B. (2019). Rolling bearing fault diagnosis method based on dispersion entropy and SVM. *Proceedings - 2019 International Conference on Sensing, oDiagnostics, oPrognostics, oand oControl, oSDPCo2019*, 0596–600. doi: 10.1109/SDPC.2019.00112

Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451–461. doi: 10.1016/S0031-3203(02)00060-2

Lipton, Z. C., Elkan, C., & Narayanaswamy, B. (2014). *Thresholding Classifiers to Maximize F1 Score*. <http://arxiv.org/abs/1402.1892>

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 413–422. doi: 10.1109/ICDM.2008.17

Lu, Q., Shen, X., Wang, X., Li, M., Li, J., & Zhang, M. (2021). Fault Diagnosis of Rolling Bearing Based on Improved VMD and KNN. *Mathematical Problems in Engineering*, 2021. doi: 10.1155/2021/2530315

Lughofer, E., & Sayed-Mouchaweh, M. (2019). Prologue: Predictive maintenance in dynamic systems. In *Predictive Maintenance in Dynamic Systems: Advanced Methods, Decision Support Tools and Real-World Applications*. doi: 10.1007/978-3-030-05645-2\_1

Maleki, S., Maleki, S., & Jennings, N. R. (2021). Unsupervised anomaly detection with LSTM autoencoders using statistical data-filtering. *Applied Soft Computing*, 108, 107443. doi: 10.1016/j.asoc.2021.107443

Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(4), 140–147. doi: 10.38094/jasst1457

Meng, Z., Zhang, Y., Zhu, B., Pan, Z., Cui, L., Li, J., & Fan, F. (2022). Research on rolling bearing fault diagnosis method based on ARMA and optimized MOMEDA. *Measurement: Journal of the International Measurement Confederation*, 189(August 2021), 110465. doi: 10.1016/j.measurement.2021.110465

Murphy, C. (n.d.). *VISIT ANALOG.COM Why MEMS Accelerometers Are Becoming the Designer's Best Choice for CbM Applications*.

Murphy, C. (2020). Choosing the Most Suitable Predictive Maintenance Sensor Structural Health Monitoring. *Analog*, 1–10. <https://www.analog.com/media/en/technical-documentation/tech-articles/choosing-the-most-suitable-predictive-maintenance-sensor.pdf>

Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, 57(December 2019), 102282. doi: 10.1016/j.ijinfomgt.2020.102282

Nissoul, H., El Harrouti, T., Serrou, D., & Abouabdellah, A. (2020). Impact of maintenance 4.0 on the performance of the production function: Application to the automotive industry in Morocco. *Proceedings - 2020 5th International Conference on Logistics Operations Management, GOL 2020*, 0, 2–6. doi: 10.1109/GOL49479.2020.9314737

Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, 11(6), 583–605. doi.org/10.3233/ida-2007-11602

Polley, E. C., Hubbard, A. E., & Laan, M. J. Van Der. (2010). Super Learner In Prediction Eric. *The Berkeley Electronic Press*.

Pr, T. (n.d.). *Étude préliminaire*.

Prasad, N. R., Almanza-Garcia, S., & Lu, T. T. (2009). Anomaly detection. *Computers, Materials and Continua*, 14(1), 1–22. doi: 10.1145/1541880.1541882

PURWAHYUDI, B. (2018). Recurrent Neural Network (Rnn) Based Bearing Fault Classification of Induction Motor Employed in Home Water Pump System. *Journal of Electrical Engineering and Computer Sciences*, Vol 3 Number 1, June 2018, 3(1), 405–412. <https://jeecs.ubhara.ac.id/index.php/JeeCSV3N1/article/view/110>

Qiao, M., Yan, S., Tang, X., & Xu, C. (2020). Deep Convolutional and LSTM Recurrent Neural Networks for Rolling Bearing Fault Diagnosis under Strong Noises and Variable Loads. *IEEE Access*, 8, 66257–66269. doi: 10.1109/ACCESS.2020.2985617

Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). *A Survey of Predictive Maintenance: Systems, Purposes and Approaches*. XX(Xx), 1–36. <http://arxiv.org/abs/1912.07383>

Rehab, A., Ali, I., Gomaa, W., & Fors, M. N. (2021). Bearings Fault Detection Using Hidden Markov Models and Principal Component Analysis Enhanced Features. *PHM Society European Conference*, 06(1), 011–11. <http://papers.phmsociety.org/index.php/phme/article/view/2947>

Roy, S. S., Dey, S., & Chatterjee, S. (2020). Autocorrelation Aided Random Forest Classifier-Based Bearing Fault Detection Framework. *IEEE Sensors Journal*, 20(18), 10792–10800. doi: 10.1109/JSEN.2020.2995109

Sayed-Mouchaweh, M. (2018). Fault diagnosis of hybrid dynamic and complex systems. In *Fault Diagnosis of Hybrid Dynamic and Complex Systems*. doi: 10.1007/978-3-319-74014-0

Sharma, A., Jigyasu, R., Mathew, L., & Chatterji, S. (2018). Bearing Fault Diagnosis Using Weighted K-Nearest Neighbor. *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018, Icoei*, 1132–1137. doi: 10.1109/ICOEI.2018.8553800

Singh, M., Sharma, S., & Kaur, A. (2013). Performance Analysis of Decision Trees. *International Journal of Computer Applications*, 71(19), 975–8887.

Tiddens', 'W. (2018). “Setting sail towards predictive maintenance.” In *“Research.Utwente.Nl”*: Vol. nan.

Tran, M. Q., Elsisi, M., Mahmoud, K., Liu, M. K., Lehtonen, M., & Darwish, M. M. F. (2021). Experimental Setup for Online Fault Diagnosis of Induction Machines via Promising IoT and Machine Learning: Towards Industry 4.0 Empowerment. *IEEE Access*, 9, 115429–115441. doi: 10.1109/ACCESS.2021.3105297

Van Der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). *Statistical Applications in Genetics and Molecular Biology Super Learner Super Learner*. 6(1).

Vapnik, V., & Izmailov, R. (2021). Reinforced SVM method and memorization mechanisms. *Pattern Recognition*, 119, 108018. doi: 10.1016/j.patcog.2021.108018

Varanis, M., Silva, A., Mereles, A., & Pederiva, R. (2018). MEMS accelerometers for mechanical vibrations analysis: a comprehensive review with applications. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 40(11), 1–18. doi: 10.1007/s40430-018-1445-5

Villacorta, J. J., Mart, R. D., Balmori, J., Magdaleno, Á., Gamaliel, L., Izquierdo, A., & Basterra, L. (2021). *Low-Cost Structural Health Monitoring System*. 1–16.

Wang, J., Wang, D., Wang, S., Li, W., & Song, K. (2021). Fault Diagnosis of Bearings Based on Multi-Sensor Information Fusion and 2D Convolutional Neural Network. *IEEE Access*, 9, 23717–23725. doi: 10.1109/ACCESS.2021.3056767

Wen, X. (2011). *A hybrid intelligent technique for induction motor condition monitoring*. April, o127. [http://eprints.port.ac.uk/2644/5Cnfiles/4661/A\\_Hybrid\\_Intelligent\\_Technique\\_for\\_Induction\\_Motor\\_Condition\\_Monitoring\\_XIN\\_WEN.pdf](http://eprints.port.ac.uk/2644/5Cnfiles/4661/A_Hybrid_Intelligent_Technique_for_Induction_Motor_Condition_Monitoring_XIN_WEN.pdf)

Zhang, S., Ye, F., Wang, B., & Habetler, T. G. (2021). Semi-Supervised Bearing Fault Diagnosis and Classification Using Variational Autoencoder-Based Deep Generative Models. *IEEE Sensors Journal*, 21(5), 6476–6486. doi: 10.1109/JSEN.2020.3040696

Zou, X., Hu, Y., Tian, Z., & Shen, K. (2019). Logistic Regression Model Optimization and Case Analysis. *Proceedings of IEEE 7th International Conference on Computer Science and Network Technology, oICCSNT o2019*, o135–139.  
doi: 10.1109/ICCSNT47585.2019.8962457