

Stratégies intelligentes de maintenance en production manufacturière dans le contexte de l'économie circulaire

par

El hadji Oumar DIAO

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE
LA MAITRISE EN GÉNIE DE LA PRODUCTION AUTOMATISÉE
M. Sc. A.

MONTRÉAL, LE 05 SEPTEMBRE 2023

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



El hadji Oumar DIAO, 2023



Cette licence [Creative Commons](https://creativecommons.org/licenses/by-nc-nd/4.0/) 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. Jean-Pierre Kenné, directeur de mémoire
Département de Génie mécanique à l'École de technologie supérieure

M. Christian Belleau, président du jury
Département de Génie mécanique à l'École de technologie supérieure

M. Victor Songmene, membre du jury
Département de Génie mécanique à l'École de technologie supérieure

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

LE 03 AOÛT 2023

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens vraiment à manifester ma profonde gratitude et mes sincères remerciements à mon directeur de recherche, le professeur Jean-Pierre Kenné, pour son aide précieuse, son professionnalisme, son dévouement pour la recherche, ses nombreux conseils éclairés ainsi que son soutien financier. J'ai apprécié sa collaboration et ses efforts m'ont permis de me surpasser dans l'atteinte de mes objectifs académiques.

Je remercie également les membres de mon jury de soutenance en l'occurrence le professeur Christian Belleau de m'avoir fait l'honneur de présider le jury et le professeur Victor Songmene, membre du jury d'avoir accepté et participé à l'évaluation de ce mémoire.

J'exprime également mes remerciements à l'ensemble du personnel enseignant de l'ÉTS de Montréal qui par leur expertise et la qualité de leur formation m'a permis d'acquérir les connaissances générale et techniques dans mon domaine d'étude.

Je suis reconnaissant envers tous mes frères et sœurs, mes amis et proches parents qui m'ont soutenu durant tout mon parcours scolaire, je remercie également ma famille au Canada pour leur accueil, leur aide, leurs conseils, leur soutien émotionnel et financier.

Je remercie infiniment mon père Cheikh Bara Diao et ma mère Thiama Diagne pour l'éducation exemplaire qu'ils m'ont octroyée sans oublier une pensée profonde pour ma maman décédée, que la terre de Touba lui soit légère.

Je tiens également à exprimer ma gratitude envers ma bien-aimée, ma tendre épouse Mame Diarra Gueye qui a toujours cru en moi. Ses nombreux encouragements, sa confiance en moi m'a donné la force nécessaire pour relever les défis rencontrés dans mes recherches.

Je remercie toutes les personnes, qui m'ont aidé et je dédie ce modeste travail à ma Maman ainsi qu'à ma Femme.

Stratégies intelligentes de maintenance en production manufacturière dans le contexte de l'économie circulaire

El hadji Oumar DIAO

RÉSUMÉ

L'économie circulaire est devenue de nos jours une approche innovatrice pour la gestion des ressources naturelles. Cette approche est de plus en plus intégrée en milieu industriel car elle permet de satisfaire les exigences environnementales. Dans ce contexte, une maintenance intelligente pourra améliorer la performance des équipements de production. Ainsi dans ce mémoire notre but sera de proposer une stratégie de maintenance intelligente de type prédictive afin de déterminer l'état des composants dans les prochaines 24h. Le modèle prédictif obtenu fournira des probabilités de défaillances de ces composants. Plusieurs contextes manufacturiers seront considérés où nous aurons des composants disposés en série, en parallèle ou en série et parallèle. Pour chaque disposition nous chercherons à optimiser la production, l'inventaire et le plan de maintenance préventive tout en prenant en compte le recyclage lors d'une maintenance préventive. Les bénéfices réalisés sur chaque dispositif seront comparés entre eux afin de déterminer l'impact de la maintenance prédictive dans une entreprise manufacturière.

Nous avons donc mis en place deux modèles dont le premier est un modèle de prédiction de défaillance et le second est un modèle d'optimisation. En ce qui concerne le modèle de prédiction, il utilise un pipeline constitué de données prétraitées et un modèle de classifieur utilisant quatre algorithmes que sont la LR (*Logistic Regression*) ou régression logistique, la RF (*Random Forest*) ou forêt aléatoire, l'AdaBoost (*Adaptive Boosting*) et le KNN (*K Nearest Neighbors*) ou les k plus proches voisins. Une analyse de la performance est établie afin de déterminer le meilleur modèle capable de classifier la défaillance d'un type de composant. Les résultats de ce premier modèle ont fourni les probabilités de défaillance de chaque composant. Ces probabilités ont été exploitées dans la construction du modèle d'optimisation. Une MIQP (*Mixed-Integer Quadratic Programming*) ou programmation quadratique en nombres entiers mixtes a été développée dans le modèle d'optimisation.

La construction de ces modèles a nécessité plusieurs techniques et outils dont la collecte de données télémétriques, l'apprentissage automatique supervisé et l'utilisation du solveur de LINGO 19.0. Dans ce mémoire nous avons présenté en premier lieu le contexte de notre étude dans le chapitre 1, suivi d'une revue de la littérature traitée dans le chapitre 2. Les leviers de mise en place de la maintenance prédictive ont été abordés dans le chapitre 3. En ce qui concerne le modèle prédictif, il a été traité dans le chapitre 4, ensuite plusieurs configurations industrielles ont été définies et enfin un modèle d'optimisation a été présenté dans le chapitre 5 exploitant les résultats du modèle prédictif associé à l'économie circulaire.

Mots-clés : planification de maintenance, maintenance intelligente, maintenance prédictive, apprentissage automatique, MIQP, économie circulaire

Intelligent maintenance strategies in manufacturing production within the context of the circular economy

El hadji Oumar DIAO

ABSTRACT

Circular economy has become an innovative approach for natural resource management nowadays. This approach is increasingly integrated in industrial settings as it enables meeting environmental requirements. In this context, intelligent maintenance can enhance the performance of production equipment. Therefore, the aim of this dissertation is to propose a predictive intelligent maintenance strategy to determine the condition of components within the next 24 hours. The predictive model obtained will provide probabilities of component failures. Several manufacturing contexts will be considered, where components are arranged in series, parallel, or series-parallel configurations. For each arrangement, we will seek to optimize production, inventory, and preventive maintenance plans while considering recycling during preventive maintenance. The benefits achieved for each configuration will be compared to determine the impact of predictive maintenance in a manufacturing company.

We have therefore established two models, the first being a failure prediction model and the second an optimization model. Regarding the prediction model, it utilizes a pipeline consisting of preprocessed data and a classifier model using four algorithms: Logistic Regression (LR), Random Forest (RF), Adaptive Boosting (AdaBoost), and K Nearest Neighbors (KNN). An analysis of performance is conducted to determine the best model capable of classifying the failure of a component type. The results of this first model provide the failure probabilities for each component. These probabilities are then utilized in constructing the optimization model, which employs Mixed-Integer Quadratic Programming (MIQP).

The construction of these models required several techniques and tools, including telemetry data collection, supervised machine learning, and the utilization of the LINGO 19.0 solver. In this dissertation, we first present the context of our study in Chapter 1, followed by a literature review in Chapter 2. The implementation drivers of predictive maintenance are addressed in Chapter 3. As for the predictive model, it is discussed in Chapter 4, followed by the definition of various industrial configurations, and finally, an optimization model is presented in Chapter 5, leveraging the results of the predictive model in association with the circular economy.

Keywords: maintenance planning, intelligent maintenance, predictive maintenance, machine learning, MIQP, circular economy

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 CONTEXTUALISATION.....	3
1.1 Introduction.....	3
1.2 Mise en contexte	3
1.3 Problématique	5
1.4 Objectifs et contributions.....	8
1.5 Méthodologie	9
1.6 Organisation du mémoire.....	10
1.7 Conclusion	10
CHAPITRE 2 REVUE DE LA LITTERATURE	11
2.1 Introduction.....	11
2.2 Intelligence artificielle	11
2.2.1 Défaillance équipement	12
2.2.2 Défaillance potentielle	12
2.2.3 Notion de dégradation.....	13
2.2.4 Notion de diagnostic	14
2.2.5 Notion de pronostic.....	15
2.2.6 Définition durée de vie utile	15
2.3 Définition d'un plan de concept.....	16
2.3.1 Plan de concept global pour notre problématique.....	16
2.3.2 Plan de concept sous-problème 1.....	17
2.3.3 Plan de concept sous-problème 2.....	17
2.3.4 Plan de concept sous-problème 3.....	17
2.4 Maintenance 4.0.....	18
2.4.1 Approche basée sur les données	19
2.4.2 Stratégie de maintenance prédictive	20
2.5 Économie circulaire	22
2.6 Conclusion	23
CHAPITRE 3 LES LEVIERS DE MISE EN PLACE DE LA MAINTENANCE INTELLIGENTE	25
3.1 Introduction.....	25
3.2 Présentation des sources de données.....	25
3.2.1 Séries temporelles	26
3.2.2 Séries transversales	27
3.2.3 Séries longitudinales	27
3.2.4 Séries panel	27
3.3 Description des outils statistiques.....	28
3.3.1 Statistiques temporelles	28

3.3.2	Statistiques multivariées	28
3.3.3	Statistiques inférentielles	29
3.3.4	Statistiques descriptives	29
3.4	Analyse Exploratoire des données	32
3.5	Analyse des données de télémétrie	33
3.5.1	Statistique descriptive du jeu de donnée télémétrie	35
3.5.2	Distribution des variables numérique de télémétrie.....	37
3.6	Analyse des données sur les erreurs.....	40
3.7	Analyse des données de maintenance	43
3.8	Analyse des données sur les défaillances des composants	45
3.9	Analyse des métadonnées des machines	47
3.10	Conclusion	49
CHAPITRE 4	UTILITÉ DE LA MAINTENANCE INTELLIGENTE DANS L'INDUSTRIE.....	51
4.1	Introduction.....	51
4.2	Préparation des données.....	51
4.2.1	Apprentissage automatique.....	52
4.2.2	Regroupement des données.....	53
4.2.3	Ingénierie des caractéristiques	59
4.3	Sélection des caractéristiques	65
4.3.1	Sous-ensemble de variable dépendante	65
4.3.2	Sous-ensemble de variable indépendantes.....	65
4.3.3	Sous-ensemble de variable non informative	65
4.4	Construction du modèle	66
4.4.1	La régression logistique	68
4.4.2	Forêt aléatoire	70
4.4.3	AdaBoost.....	71
4.4.4	Algorithme k plus proche voisin.....	75
4.5	Présentation des résultats	77
4.5.1	Précision d'un modèle.....	77
4.5.2	Rappel	78
4.5.3	F-mesure	78
4.5.4	Exactitude	79
4.5.5	Moyenne macroscopique (Macro avg)	79
4.5.6	Moyenne pondérée (Weighted avg).....	79
4.5.7	Matrice de confusion.....	80
4.5.8	Résultats pour la régression logistique	81
4.5.9	Résultats pour la forêt aléatoire	84
4.5.10	Résultats pour l'AdaBoost	85
4.5.11	Résultats pour algorithme des k plus proche voisin.....	87
4.6	Validation du modèle.....	88
4.6.1	Vérification surapprentissage.....	88
4.6.2	Courbe ROC des algorithmes	89
4.7	Optimisation des hyperparamètres.....	93

4.7.1	Hyperparamètre pour la Régression Logistique	94
4.7.2	Réglage hyperparamètre pour la Forêt Aléatoire.....	96
4.7.3	Sélection des variables les plus importantes.....	97
4.8	Prédiction et Déploiements.....	101
4.9	Conclusion	104
CHAPITRE 5 IMPACT DE L'ÉCONOMIE CIRCULAIRE EN MILIEU		
	INDUSTRIEL	105
5.1	Introduction.....	105
5.2	Notion de Fiabilité	105
5.2.1	Relation entre fiabilité et défaillance	105
5.2.2	Fiabilité composant en série.....	108
5.2.3	Fiabilité composant en parallèle	109
5.3	Fiabilité appliquée à la maintenance prédictive.....	111
5.3.1	Disposition du sous cas 1	111
5.3.2	Disposition du sous cas 2.....	113
5.3.3	Disposition du sous cas 3	114
5.4	Modèle d'optimisation.....	123
5.4.1	Objectifs de la production.....	124
5.4.2	Ensembles et indices	126
5.4.3	Paramètres.....	126
5.4.4	Variable de décision.....	127
5.4.5	Fonction-objectif.....	127
5.4.6	Contraintes	128
5.5	Expérimentation.....	130
5.6	Discussion.....	134
5.7	conclusion	136
CONCLUSION.....		137
ANNEXE I	SYNTHESE REVUE DE LA LITTÉRATURE.....	139
ANNEXE II	ANALYSE EXPLORATOIRE DES DONNÉES	143
ANNEXE III	MODÈLE PRÉDICTIVE.....	145
ANNEXE IV	MODÈLE D'OPTIMISATION CODE LINGO.....	147
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES		149

LISTE DES TABLEAUX

	Page
Tableau 3.1	Les premières valeurs de télémétrie.....34
Tableau 3.2	Statistique descriptive du Jeu de données de Télémétrie.....35
Tableau 3.3	Les types de variables du jeu de données de télémétrie.....38
Tableau 3.4	Les cinq premières lignes du jeu de données des erreurs40
Tableau 3.5	Les types de variables du jeu de données des erreurs.....40
Tableau 3.6	Les cinq premières lignes du jeu de données de maintenance.....43
Tableau 3.7	Les types de variables du jeu de données de maintenance43
Tableau 3.8	Les cinq premières lignes du jeu de données de défaillance46
Tableau 3.9	Les types de variables du jeu de données de défaillance46
Tableau 3.10	Les métadonnées.....48
Tableau 3.11	Les types de variables sur les métadonnées.....48
Tableau 3.12	Statistique descriptive de la variable age48
Tableau 4.1	Encodage des erreurs54
Tableau 4.2	Encodage des erreurs et ajout de variables télémétriques.....55
Tableau 4.3	Encodage de la variable comp du Jeu de données des maintenances55
Tableau 4.4	Fusion de télémétrie, erreurs, maintenance et machine56
Tableau 4.5	Fusion de tous jeux de données57
Tableau 4.6	Caractéristique de décalage minimum pour fenêtre de 3h.....60
Tableau 4.7	Caractéristique de décalage maximum pour fenêtre de 3h60
Tableau 4.8	Caractéristique de décalage des moyennes pour fenêtre de 3h.....61
Tableau 4.9	Caractéristique de décalage des écart-type pour fenêtre de 3h.....61
Tableau 4.10	Caractéristique de décalage minimum pour fenêtre de 24h.....62

Tableau 4.11	Rapport de classification par la régression logistique.....	84
Tableau 4.12	Rapport de classification pour la forêt aléatoire	85
Tableau 4.13	Rapport de classification pour AdaBoost	87
Tableau 4.14	Rapport de classification KNN	88
Tableau 4.15	Vérification surapprentissage.....	89
Tableau 4.16	Rapport de classification par réglage hyperparamètre de la régression logistique	95
Tableau 4.17	Rapport de classification pour la Forêt aléatoire avec 1000 arbres	96
Tableau 4.18	Rapport de classification obtenu avec forêt aléatoire et seuil S1.....	99
Tableau 4.19	Rapport de classification avec nombre arbre, profondeur et S1	100
Tableau 4.20	Rapport de classification pour forêt aléatoire et deuxième réglage sous S1	100
Tableau 4.21	Probabilité de défaillance sur la machine 1	103
Tableau 5.1	Fiabilité pour le sous cas 1	113
Tableau 5.2	Fiabilité pour le sous cas 2.....	114
Tableau 5.3	Fiabilité sous cas 3-A.....	116
Tableau 5.4	Fiabilité sous cas 3-B	118
Tableau 5.5	Fiabilité sous cas 3-C	120
Tableau 5.6	Fiabilité sous cas 3-A'	121
Tableau 5.7	Fiabilité sous cas 3-B'	122
Tableau 5.8	Fiabilité Sous cas 3-C'.....	123
Tableau 5.9	Temps de traitement du produit sur une machine	125
Tableau 5.10	Ventes prévues	126
Tableau 5.11	Planning de maintenance préventive initial	126
Tableau 5.12	Plan de production optimale pour scénario 1.....	131

Tableau 5.13	Inventaire optimale pour scénario 1	131
Tableau 5.14	Plan de maintenance optimale pour scénario 1	131
Tableau 5.15	Fiabilité dispositif en série	135
Tableau 5.16	Plan de maintenance optimal dispositif en série	135
Tableau 5.17	Fiabilité dispositif en parallèle	136

LISTE DES FIGURES

	Page
Figure 1.1	Différentes type de maintenance tirée de Gordon, Burnak, Onel & Pistikopoulos (2020)5
Figure 1.2	Schématisation de l'économie linéaire tirée de RECYC-QUÉBEC (2023)6
Figure 1.3	Schématisation de l'économie circulaire tirée de RECYC-QUÉBEC (2023)8
Figure 2.1	Exemple de dégradation de système tirée de Zwingelstein (2019).....14
Figure 3.1	Intervalle de confiance tirée de Minitab (2021).....29
Figure 3.2	Diagramme en boîte tirée de Menet-Nédélec, Lampert, & Riou (2011).....32
Figure 3.3	Distribution de la tension suivant le mois de janvier34
Figure 3.4	Distribution du jeu de données de télémétrie en fonction des 3 machines36
Figure 3.5	Organigramme d'une variable38
Figure 3.6	Diagramme en boîte des variables numériques du jeu de donnée télémétrie.....39
Figure 3.7	Distribution des erreurs suivant les machines.....42
Figure 3.8	Fréquence de maintenance préventive sur les composants44
Figure 3.9	Fréquence mensuel de maintenance préventive sur les composants45
Figure 3.10	Distribution des composants remplacés lors d'une maintenance corrective sur chaque machine47
Figure 4.1	Variante d'un apprentissage automatique53
Figure 4.2	Matrice de corrélation des variables58
Figure 4.3	Matrice de corrélation après ingénierie des caractéristiques63
Figure 4.4	Schéma du modèle67

Figure 4.5	Différence entre régression linéaire et logistique tirée de Voxco (2023).....	69
Figure 4.6	Principe d'une régression logistique adaptée de Voxco (2023)	69
Figure 4.7	Architecture d'une forêt aléatoire tirée de Engels (s.d).....	71
Figure 4.8	Principe des k plus proche voisin.....	76
Figure 4.9	Principe de la matrice de confusion tirée de Rapid Miner (2023)	81
Figure 4.10	Ensemble de formation et de test	82
Figure 4.11	Matrice de confusion par la régression logistique	83
Figure 4.12	Matrice de confusion par la forêt aléatoire	84
Figure 4.13	Matrice de confusion initial pour AdaBoost.....	86
Figure 4.14	Matrice de confusion initiale pour KNN	87
Figure 4.15	Courbe ROC initiale par Régression Logistique	90
Figure 4.16	Courbe ROC initiale par la Forêt aléatoire	91
Figure 4.17	Courbe ROC pour AdaBoost	92
Figure 4.18	Courbe ROC pour KNN.....	92
Figure 4.19	Principe du réglage des hyperparamètres	93
Figure 4.20	Courbe ROC par réglage des hyperparamètres de la régression logistique.....	95
Figure 4.21	Courbe ROC pour forêt aléatoire à 1000 arbres	97
Figure 4.22	Première partie des variables les plus importantes	98
Figure 4.23	Deuxième partie des variables les plus importantes	98
Figure 4.24	Probabilité défaillance de la classe none	101
Figure 4.25	Probabilité de défaillance du composant 2	102
Figure 4.26	Probabilité de défaillance du composant 4	102
Figure 5.1	Évolution taux de défaillance en fonction du temps tirée de openclassrooms (2023)	107

Figure 5.2	Évolution du taux de défaillance.....	108
Figure 5.3	Disposition en série.....	109
Figure 5.4	Disposition en parallèle.....	111
Figure 5.5	Dispositif sous cas 1.....	112
Figure 5.6	Dispositif sous cas 2.....	114
Figure 5.7	Dispositif Sous cas 3 - A.....	116
Figure 5.8	Dispositif Sous cas 3 B.....	118
Figure 5.9	Dispositif sous cas 3-C.....	120
Figure 5.10	Disposition sous cas 3-A'.....	121
Figure 5.11	Dispositif sous cas 3- B'.....	122
Figure 5.12	Disposition sous cas C'.....	123
Figure 5.13	Système manufacturier.....	124
Figure 5.14	Évolution du bénéfice en fonction des scénarii	134

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

IA	Intelligence artificielle
IoT	<i>Internet of Things</i>
RF	<i>Random Forest</i>
LR	<i>Linear Regression</i>
AdaBoost	<i>adaptive boosting</i>
KNN	<i>K-Nearest Neighbors</i>
ML	<i>Machine Learning</i>
CBM	<i>Condition-based maintenance</i>
RUL	<i>Remaining Useful Life</i>
SCADA	<i>Supervisory Control And Data Acquisition</i>
PIG	<i>Pipeline Inspection Gauge</i>
CSV	<i>Comma-separated Values</i>
ROC	<i>Receiving Operator Characteristic</i>
AUC	<i>Area Under Curve</i>
MLP	<i>Multilayer Perceptron</i>
MILP	<i>Mixed-Integer Linear Programming</i>
MIQP	<i>Mixed-Integer Quadratic Programming</i>
DBSCAN	<i>Density-based Spatial Clustering of Applications with Noise</i>
ARIMA	<i>AutoRegressive Integrated Moving Average</i>
ACP	Analyse en composantes principales
λ	Taux de panne

INTRODUCTION

La maintenance industrielle constitue un levier ou un outil d'aide à la décision. Elle permet par son efficacité d'atteindre des objectifs ambitieux à savoir l'optimisation des termes qualité, coûts et délai. Depuis le développement du numérique, la maintenance a évolué vers des formes dites intelligentes. Pour cela, de nouvelles technologies sont utilisées. Nous pouvons en citer l'IoT (*Internet of Things*) ou internet des objets et l'IA (intelligence artificielle) qui permettront d'optimiser la disponibilité des ressources techniques et matérielles. Cependant, la mise en œuvre d'une stratégie de maintenance intelligente repose en grande partie sur l'utilisation et l'exploitation de données dans le but de prévenir de futures défaillances afin d'améliorer la disponibilité des ressources de production tout en minimisant les coûts de maintenance.

Par ailleurs, avec le réchauffement climatique et la diminution des ressources naturelles, l'économie circulaire se présente comme une alternative qui respecterait la plupart des contraintes associées au développement durable. Pour cela, l'économie circulaire pourrait être définie comme étant un concept révolutionnaire qui vise à maximiser l'utilisation des équipements tout en minimisant les déchets et l'ensemble des pertes (McDonald, Normandin, & Sauv , 2016). C'est ainsi que l'économie circulaire associ e   une strat gie de maintenance intelligente permettrait d'augmenter la dur e de vie des machines. Pour cela, les composants d fectueux pourront  tre diagnostiqu s et remplac s avant m me que la panne ne survienne. D'autre part, la production sera maximis e par l'utilisation des donn es collect es qui am lioreront la fiabilit  et la durabilit  des  quipements. Dans ce travail, nous chercherons   r pondre   un certain nombre de questions qui sont :

- Quand est-ce qu'une d faillance d' quipement pourrait survenir ?
- Si une d faillance d' quipement est pr dite, quelle sera sa probabilit  associ e ?
- En utilisant ces  quipements soumis   des d faillances, quelle en sera l'impact en termes de productivit  ?
- Est-ce que l' conomie circulaire peut am liorer la comp titivit  ?

Ces questions montrent que la défaillance doit être prise en compte afin de satisfaire les exigences économiques, sociales et environnementales. Dans ce sens le recours à la maintenance prédictive pourrait être un moyen permettant d'une part de prédire l'état d'un équipement et d'autre part de minimiser l'apparition des pannes.

Ce type de maintenance a été traité dans plusieurs recherches dont nous pouvons citer les travaux de Jemmali (2022), qui étudie une maintenance prédictive permettant de détecter l'état d'une turbine de puissance en utilisant plusieurs classifieurs. Ainsi le mode panne et fonctionnement sont prédits évitant ainsi les arrêts non programmés. De la même façon la maintenance prédictive a été étudiée dans les travaux de Jemli (2023), l'auteur traite la défaillance de composants qui pourrait apparaître dans les moteurs électriques. Ainsi son modèle prédictif élaboré permet d'estimer l'état de santé de ces équipements en utilisant un modèle de classification basé sur les SVM (*Support Vector Machine*). Par ailleurs, l'économie circulaire est devenu un moyen incontournable dans l'optimisation et la gestion des ressources, c'est dans ce sens que Delpla (2021) a étudié un système manufacturier en boucle fermée, où le recyclage de produits est pris en compte dans son modèle Lingo.

Dans notre mémoire, nous aurons donc plusieurs objectifs à savoir la mise en place d'une maintenance prédictive et la mise en place d'un modèle d'optimisation prenant en compte le recyclage. De ce fait, nous nous organiserons en cinq chapitres dont le premier traitera le contexte général de notre recherche. Le deuxième chapitre sera réservé aux concepts fondamentaux et à la revue de la littérature, où nous présenterons les études réalisées dans la maintenance intelligente ainsi que celles traitée dans l'économie circulaire. Dans le chapitre trois les outils de mise en œuvre de maintenance prédictive seront détaillés en présentant l'analyse exploratoire des données. Dans le chapitre quatre un modèle prédictif sera construit assurant la stratégie de maintenance intelligente, nous présenterons également dans ce chapitre l'ensemble des étapes nécessaires pour la création d'un modèle de prédiction. Enfin le chapitre cinq étudiera un modèle d'optimisation de type MIQP qui exploitera les probabilités de défaillances dans des configurations, où le recyclage de composants sera pris en compte.

CHAPITRE 1

CONTEXTUALISATION

1.1 Introduction

La maintenance industrielle permet d'assurer une bonne utilisation des équipements favorisant ainsi un bon fonctionnement des outils qui les composent. Ainsi la maintenance permet de corriger ou de prévenir des défaillances d'une ressource technique. Pour cela, elle s'appuie sur des inspections fréquentes, des réparations, des remplacements de pièces défectueuses, des tâches de lubrification de systèmes mécaniques ou d'autres types d'entretien jugés pertinents par les besoins de l'entreprise. L'économie circulaire intéresse de plus en plus les industriels de ce fait, nous l'associons avec la maintenance pour en faire notre travail de recherche, ainsi ce chapitre permettra aux lecteurs de mieux comprendre l'utilité de nos recherches.

1.2 Mise en contexte

Bien que la maintenance industrielle propose une multitude de tâches, elle peut toutefois être décomposée en plusieurs branches que sont la maintenance préventive, la maintenance corrective, la maintenance prédictive et même la maintenance prescriptive. En ce qui concerne la maintenance préventive, son rôle consiste à mettre en place, effectuer et suivre des tâches planifiées dans le but d'empêcher l'apparition des pannes. Quant à la maintenance corrective, elle s'utilise pour traiter et rectifier une défaillance déjà présente sur un équipement; cela revient à intervenir pour réparer une panne d'un composant, machine, outil ou système.

En revanche la maintenance prédictive se distingue de la prévention et de la correction dans le sens, où elle fait appel à des stratégies de surveillance, d'analyse et de contrôle pour prédire l'apparition d'une panne future ainsi la maintenance prédictive pourra mettre en place des plans de maintenance préventive. Concrètement, cela revient à dire que la maintenance prédictive permet de fournir un plan de maintenance préventive.

Cependant, la maintenance prédictive regroupe et utilise des techniques sophistiquées comme le *Machine Learning* (ML) ou apprentissage automatisée, les techniques de collecte de données par l'intermédiaire des IoT, également dans certaines situations la réalité augmentée (RA) et la réalité virtuelle (RV) pourront être utilisées.

Un autre type de maintenance qui serait intéressant d'expliquer serait la maintenance conditionnelle ou *condition-based maintenance* (CBM) qui peut être définie comme étant un exemple de maintenance préventive avec la particularité que la CBM utilisera des données de surveillance en temps réel dans le but de détecter les signes précoces de défaillances avant qu'une panne ne survienne.

Dans cet optique, la CBM se différencie de la maintenance basée sur les intervalles ou (MIB) qui se définit comme étant une stratégie de maintenance préventive se basant sur une planification de la maintenance à des moments et intervalles de temps prédéfinis comme mettre en place un planning de maintenance à fréquence mensuel, hebdomadaire, mensuel, bimensuel, etc. dans le but de respecter certaines exigences liées aux ressources techniques et/ou aux recommandations des fabricants.

Parmi les tâches effectuées pendant ces intervalles de temps nous pouvons citer l'inspection, le nettoyage, la lubrification et diverses réparations qui pourraient favoriser une meilleure condition d'utilisation. Ce type de maintenance est fréquent en milieu industriel et peut être utilisé dans l'industrie automobile, l'aéronautique, les équipements de production et les systèmes mécatroniques.

En ce qui concerne la maintenance prescriptive, elle se différencie de la maintenance prédictive dans le sens où elle utilise les résultats de la maintenance prédictive afin de proposer des décisions d'exploitation, de production et de maintenance (Gordon, Burnak, Onel, & Pistikopoulos, 2020).

En un mot, nous dirons que la maintenance industrielle est un garant qui assure l'excellence opérationnelle en favorisant les niveaux de production élevés et les respects des contraintes liées aux ressources administratives et humaines. La figure 1.1 résume les différents types de maintenance. Avec une multitude de stratégies de maintenance, il convient donc de déterminer la meilleure approche qui pourrait être bénéfique pour un système manufacturier donné.

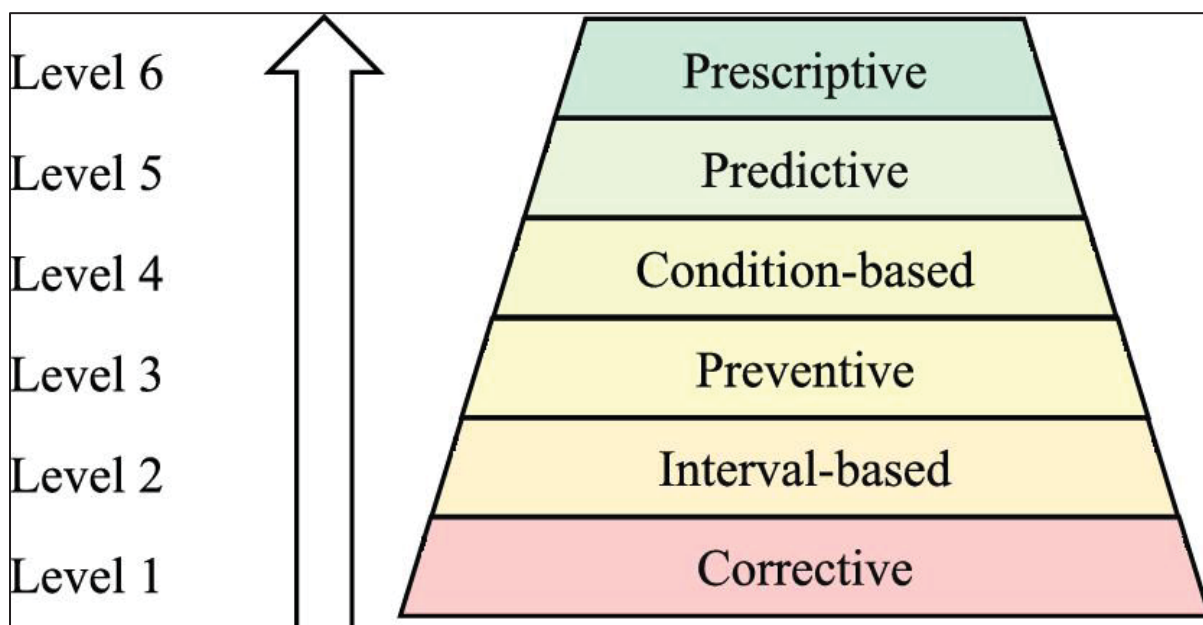


Figure 1.1 Différentes type de maintenance tirée de Gordon, Burnak, Onel & Pistikopoulos (2020)

La mise en œuvre d'un type de maintenance nécessite de respecter certains défis et contraintes que nous tenterons d'expliquer dans le paragraphe suivant en mettant l'accent sur notre problématique de recherche.

1.3 Problématique

La conjoncture actuelle de l'économie mondiale est marquée par la rareté voire la disparition de certaines ressources. Pour éviter ces problèmes, l'économie circulaire constitue une approche qui pourrait se substituer à l'économie linéaire traditionnelle.

Le concept de l'économie linéaire traditionnel est représenté par la figure 1.2

L'économie linéaire qui consiste à extraire, produire, consommer et jeter se différencie de l'économie circulaire dans la mesure, où elle prend en compte la gestion des ressources en intégrant la manière de produire et de consommer. Les ressources étant limitées donc il convient de minimiser sa consommation afin de préserver l'écosystème et l'environnement. Les contraintes liées à l'écosystème seront prises en compte par une approche écoconception qui consiste à intégrer l'impact environnemental des produits au cours de leur cycle de vie en commençant par la phase de conception jusqu'à la phase d'utilisation.

Ainsi l'écoconception utilisant l'impact environnemental ne doit pas être confondu avec l'analyse du cycle de vie (Aït-El-Hadj, 2013).

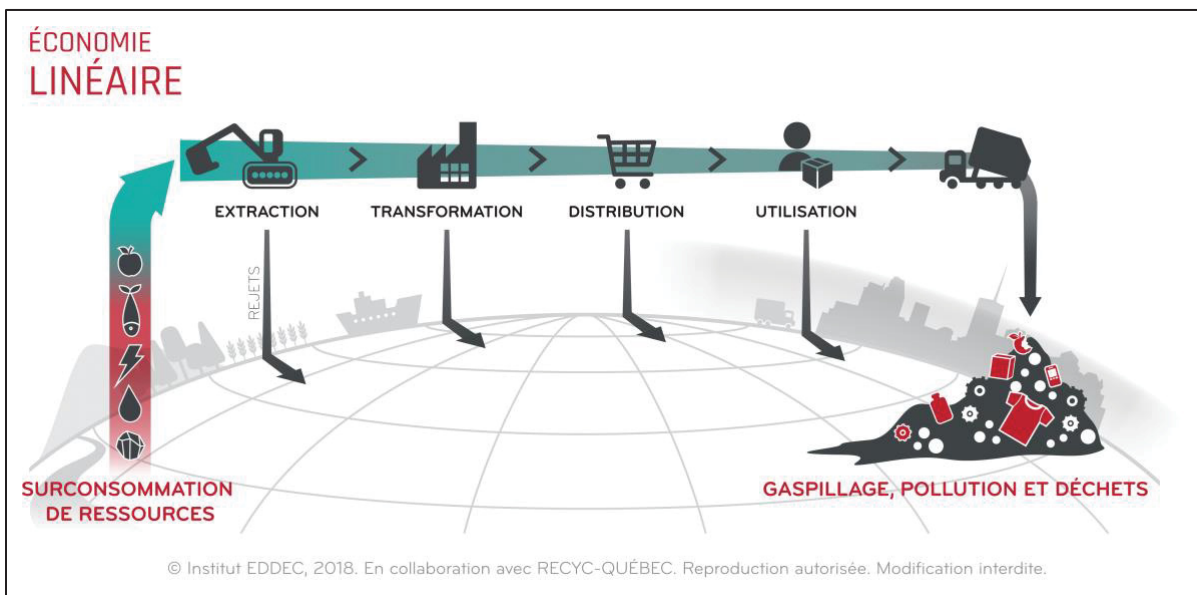


Figure 1.2 Schématisation de l'économie linéaire tirée de RECYC-QUÉBEC (2023)

De ce fait, Le recours de matériaux de recyclage permet de mettre en place l'approche d'écoconception. L'économie circulaire permet également d'optimiser l'utilisation des ressources pour cela trois axes seront considérés à savoir :

- La fréquence d'utilisation des produits
- La maximisation de la durée de vie des équipements ainsi que les composants
- La réutilisation de produits en fin de vie

La fréquence d'utilisation des produits peut être intégrée dans une approche d'économie circulaire en utilisant l'économie collaborative et la location court terme. En ce qui concerne l'économie collaborative elle est basée « sur un partage et un échange de bien et de services entre particuliers » (Saint-Esprit, 2023). L'économie collaborative souvent appelée économie de partage peut être un gage de réussite pour la mise en place de l'économie circulaire car elle permet un partage de ressources non utilisées ou peu utilisées entraînant ainsi une faible émission de gaz à effet de serre.

La durée de vie des produits doit être allongée afin de favoriser la réussite d'une stratégie d'économie circulaire pour cela, les équipements doivent être entretenus régulièrement afin de réduire leur niveau d'usure.

L'économie fonctionnelle permet également de maximiser la durée de vie des équipements car elle consiste à offrir des services plutôt que de vendre les produits (ADEME Expertises, 2022).

La réutilisation des produits en fin de vie est incontournable dans une démarche d'économie circulaire et peut être appliquée par la mise en place du recyclage qui se définit comme étant un moyen de transformation et ou d'utilisation de déchets comme matière première (EC, 2019).

La figure 1.3 résume les principaux axes de l'économie circulaire.

Ainsi cette approche d'économie circulaire permettra de réduire l'utilisation des ressources naturelles, de contrôler l'accès aux ressources protégées, de minimiser les déchets prolongeant ainsi la durée de vie des machines ou ressources techniques (McDonald et al., 2016).

Ainsi nous chercherons à savoir comment assurer l'efficacité et la réussite d'une approche d'économie circulaire ?

Une réponse probable à cette question serait sans doute qu'une maintenance efficace permettrait de faciliter la mise en place d'une économie circulaire.

Pour cela, la maintenance associée à l'économie circulaire permettra de mieux maîtriser le cycle de vie des produits réduisant ainsi l'utilisation des ressources naturelles et la production de déchets industriels. Face à ces défis de taille, notre problématique consistera à étudier les stratégies de maintenance intelligente dans le contexte d'économie circulaire. Cependant, les objectifs et contribution seront fournis dans le paragraphe suivant afin de mieux comprendre l'utilité et le choix de notre sujet de recherche.

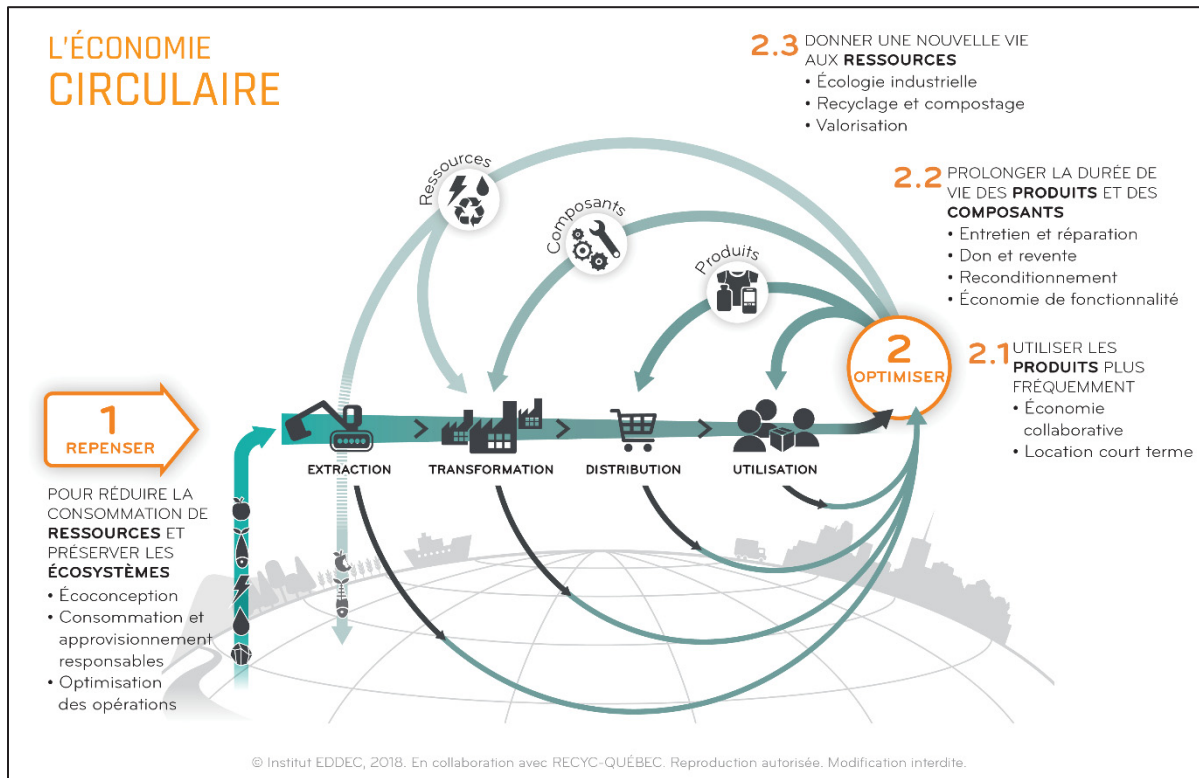


Figure 1.3 Schématisation de l'économie circulaire tirée de RECYC-QUÉBEC (2023)

1.4 Objectifs et contributions

Notre projet de recherche est de mettre en place des stratégies de maintenance intelligente dans le contexte de l'économie circulaire. Nous proposerons une approche basée sur la durabilité afin de garantir une maximisation de la durée de vie des équipements ainsi nous réduirons la consommation de ressources en essayant de minimiser les déchets.

Ainsi notre approche permet à une entreprise manufacturière de relever un certain nombre de défis que sont :

- Prolonger la durée de vie de ces équipements à travers une approche dynamique et efficiente de maintenance.
- Mettre et déployer une approche d'économie circulaire valorisante.

- Réduire les coûts d'exploitation de façon globale en mettant l'accent sur la prévention des défaillances qui engendrerait des coûts de réparation exorbitantes et des dépassements de budget alloué à l'entretien des équipements.

Notre principale contribution sera de proposer une stratégie efficace dans laquelle nous intégrerons une touche d'économie circulaire aux stratégies existantes. Afin de mieux comprendre notre projet et comment nous y procéderons, nous expliquerons en détails dans la prochaine section, la méthodologie adoptée pour la résolution de notre problématique de recherche.

1.5 Méthodologie

Notre problématique consiste à étudier une stratégie de maintenance intelligente dans un contexte d'économie circulaire. À l'issue de nos recherches bibliographiques préliminaires nous avons constaté que l'association de la maintenance intelligente et de l'économie circulaire n'est pas traitée dans la littérature. Nous en avons déduit que notre sujet est unique et mérite d'être traité par une méthodologie claire et efficace.

Pour cela, nous avons divisé notre problématique en trois sous-problème, où chaque sous-problème constituera une contribution de recherche. La résolution de l'ensemble de nos sous-problèmes permettra de traiter notre problématique dans son intégralité. Les trois sous-problèmes jugés pertinents à notre niveau sont :

- Le sous-problème 1 traitera les leviers de mise en place d'une maintenance intelligente.
- Le sous-problème 2 s'intéressera à l'utilité de la maintenance intelligente dans l'industrie.
- Le sous-problème 3 étudiera l'impact de l'économie circulaire en milieu industriel.

Le plan de résolution des sous-problèmes est exposé dans la section suivante

1.6 Organisation du mémoire

La rédaction de notre mémoire s'articule autour de la résolution des sous-problèmes. Pour cela, nous scindons notre travail en quatre chapitres excluant ce premier chapitre qui porte sur le contexte et le cadre théorique de nos travaux.

Le chapitre 2, présente les notions essentielles à la compréhension de notre sujet que sont les éléments liés à l'IA suivies d'une revue de la littérature.

Le chapitre 3, traite le sous-problème 1 portant sur des leviers de mise en place de la maintenance intelligente.

Le chapitre 4 nous traite le sous-problème 2 portant sur l'utilité de la maintenance intelligente dans l'industrie basée sur l'utilisation de l'apprentissage automatique dans le but de mettre en place une stratégie de maintenance prédictive.

Le chapitre 5 étudie l'impact de l'économie circulaire en milieu industriel en intégrant le concept de recyclage dans le modèle d'optimisation résolu dans ce mémoire par le solveur LINGO.

1.7 Conclusion

Dans ce chapitre nous avons montré que notre sujet mérite un effort de recherche à cause de sa complexité mais aussi à cause de l'importance qu'il peut procurer en termes d'avancée scientifique et technologique. Les hypothèses ont été définies dans la résolution des sous problèmes. La revue des travaux sera présenter dans le prochain chapitre afin de trouver les travaux existants dans les thèmes de maintenance et d'économie circulaire.

CHAPITRE 2

REVUE DE LA LITTERATURE

2.1 Introduction

Dans le présent chapitre, nous mettrons en lumière la notion d'IA qui permettra de mieux appréhender la maintenance intelligente. Ensuite, une présentation des différentes notions de défaillances sera faite. Afin de comprendre la méthodologie de recherche appliquée nous parlerons des plans de concept qui nous ont permis de vérifier les études déjà existantes dans le domaine de la maintenance 4.0 d'une part et de regarder les recherches scientifiques dans le secteur de l'économie circulaire d'autre part. Enfin ce chapitre sera clôturé par un tableau récapitulatif placé en annexe qui résumera l'ensemble de travaux connexe à notre problématique de recherche.

2.2 Intelligence artificielle

L'intelligence artificielle ou (IA) se définit comme étant une partie de l'informatique spécialisée dans la mise en œuvre de machines dites intelligentes qui auront les aptitudes cognitives très avancés qui leur permettrais d'étudier et de comprendre comme les humains. Pour cela, elle se base sur des algorithmes et des modèles mathématiques dans le but d'analyser, d'interpréter, de résoudre des problèmes complexes ou même de prendre des décisions qui idéalement nécessiterait l'intervention humaine.

L'IA est utilisé dans plusieurs secteurs d'activités nous pouvons en citer par exemple le secteur de la reconnaissance de forme des objets, l'apprentissage automatisée, la classification des objets, la conversation, la maintenance industrielle et bien d'autre domaine.

Cette vision est partagée par de nombreux chercheurs dont un qui soutient qu' « Aujourd'hui, nous disposons d'un grand nombre de systèmes d'IA qui font partie intégrante des processus métier critiques »(Benda, 1993).

L'IA appliquée à la maintenance permettra de prédire la défaillance de composants. Pour cela dans les lignes suivantes la notion de défaillance y sera présentée.

2.2.1 Défaillance équipement

Une défaillance sur une ressource technique apparaît lorsque l'équipement en question ne fonctionne plus ou pas de la meilleure façon prévue pour son utilisation. L'apparition de défaillance peut être due par plusieurs facteurs comme les erreurs de conception, les erreurs de fabrication, un niveau d'usure très élevé, des erreurs de manipulations et d'entretiens.

Par ailleurs des scientifiques affirment que la défaillance peut être défini selon la norme NF EN 13306 comme étant « une perte d'aptitude d'un bien à accomplir une fonction requise »(Zwingelstein, 2019). De la même façon la norme ISO 16079-1 qui est réservé pour le diagnostic et la surveillance des éoliennes définit à son tour la défaillance comme étant « la cessation de l'aptitude d'un composant ou d'une machine à effectuer la fonction requise » (Zwingelstein, 2019). En revanche la norme CEI 60050 la définit en ces termes « une perte de l'aptitude d'une entité à fonctionner tel que requis» (Zwingelstein, 2019). Étant donné que nous travaillons sur des stratégies de maintenance, nous jugeons pertinent de présenter les défaillances potentielles dans le paragraphe suivant.

2.2.2 Défaillance potentielle

La défaillance potentielle indique une situation où un équipement pourrait probablement tomber en panne ou lorsque l'équipement tomberais dans un état de dysfonctionnement. La défaillance potentielle se résume à dire qu'il n'y'a pas de panne à l'instant t mais elle pourrait se produire dans un future proche. La plupart des services de maintenance procède à des inspections régulières ou de la maintenance préventive afin de déceler des défaillances potentielles, de ce fait des mesures pourront être prises afin d'éviter l'apparition de la panne.

Par ailleurs, nous soulignons que la non prise en compte d'une défaillance potentielle pourrait entrainer des pannes réelles survenant à des moments non prévus ce qui engendreraient des maintenances correctives et des augmentations de couts alloués aux entretiens des équipements.

La norme ISO 16079-1 se différencie de celle de JA 1012 dans le sens, où elle affirme qu'une défaillance potentielle indique le moment où un défaut peut être détecté. La dégradation qui est souvent rencontré dans les problématiques de maintenance sera présentée ci-dessous.

2.2.3 Notion de dégradation

Une dégradation peut être présentée comme un instant présent à partir duquel l'état de santé d'un équipement s'est détérioré par rapport à l'instant précédent. Nous pouvons la définir autrement en disant qu'une dégradation n'est rien d'autre qu'une régression de l'état de santé d'un équipement ou de ses performances. Plusieurs facteurs peuvent entraîner la dégradation d'une ressource technique, nous pouvons citer par exemple un niveau d'usure très élevé, l'apparition de défauts non traités ou non détecté ou négligé de même que l'absence d'un plan de maintenance efficace et rigoureux.

L'éradication d'une dégradation peut se faire par l'intermédiaire d'une maintenance préventive qui rendra le système dans un état de fonctionnement optimal. Il existe plusieurs méthodes pour détecter une dégradation, nous pouvons en citer la surveillance des équipements, les inspections visuelles, l'analyse vibratoire et l'analyse fréquentielle.

Cependant une dégradation d'équipement doit être pris en compte le plus rapidement possible car elle pourrait entraîner l'utilisation de la maintenance corrective. En effet nous avons détaillés ces concepts car ils sont fondamentaux pour une bonne compréhension des stratégies de maintenance.

Pour cela, la figure 2.1 ci-dessous permet de mieux visualiser l'effet de la dégradation sur la santé d'un équipement. Comme nous l'avons vu précédemment la dégradation peut être défini selon différentes normes, par exemple la CEI 60050 reconnaît une dégradation sur le fait qu'elle désigne une modification préjudiciable de l'aptitude à satisfaire des exigences tandis que la norme NF EN 13306 indique que la dégradation représente un changement néfaste de l'état physique.

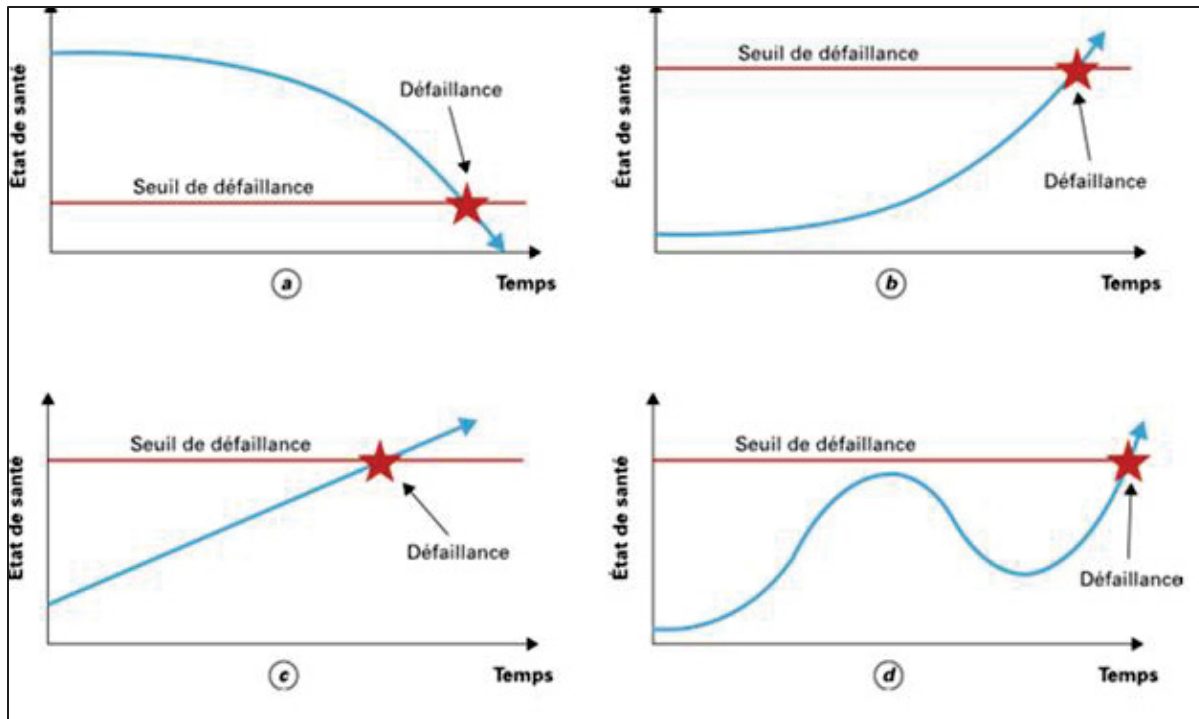


Figure 2.1 Exemple de dégradation de système tirée de Zwingelstein (2019)

Étant donné que les dégradations peuvent aboutir sur des défaillances donc il convient de trouver un moyen de les détecter d'où la mise en place d'un diagnostic. Ce dernier sera expliqué ci-dessous.

2.2.4 Notion de diagnostic

Le diagnostic se définit comme étant un outil qui permet de déceler, d'identifier, d'étudier et d'analyser les dysfonctionnements d'un système. Dans le cadre d'une maintenance intelligente la mise en place du diagnostic peut se faire par l'intermédiaire de capteurs connectés et des logiciels chargés d'effectuer l'analyse des données. Ainsi les données sont étudiées à temps réel et permettront de vérifier leur écart par rapport aux valeurs souhaités. Si toutefois des écarts sont constatés des mesures pourront être prises pour réparer les défaillances ou pour rendre l'équipement dans un état de fonctionnement optimal.

Le diagnostic s'avère utile en milieu industriel car il facilitera le traitement d'une défaillance empêchant ainsi l'arrêt du système. Le diagnostic bien qu'efficace ne garantit pas toujours le succès d'une stratégie de maintenance pour cela nous tenterons d'apporter un supplément appelé le pronostic qui sera détaillé ci-après.

2.2.5 Notion de pronostic

Dans la littérature nous trouvons que le mot pronostic est issu de la concaténation de deux termes grecs que sont « Pro » et « Gnosis » qui veulent dire respectivement événement avant et capacité d'acquérir des connaissances. Concrètement, cela signifie la capacité à acquérir des connaissances sur des événements avant qu'elles ne surviennent. Ainsi dans le cadre d'une maintenance intelligente le pronostic désigne la prédiction de probabilité de défaillance voir de panne d'une ressource technique dans un horizon fixé à l'avance.

Afin de mettre en place l'approche d'un pronostic, les équipes de maintenance s'appuieront sur des modèles statistiques, des algorithmes de ML en utilisant les données dans le but de fournir dans certaines conditions une estimation de la durée de vie restante et dans d'autres conditions d'évaluer le temps où une prochaine panne pourrait survenir.

Idéalement le pronostic permet d'établir une stratégie de maintenance prédictive ce qui pourrait être un atout considérable en matière d'optimisation des temps de maintenance, des coûts de maintenance et de planification de la production. L'étude de la durée de vie utile restante d'un équipement ou RUL (*Remaining Useful life*) peut être considéré comme un socle de la maintenance prédictive. Il existe de nombreux travaux qui traitent la notion de la RUL. Dans le paragraphe suivant la RUL sera expliquée brièvement.

2.2.6 Définition durée de vie utile

La durée de vie utile est un terme fréquemment employé dans une approche de maintenance prédictive. Son rôle est de fournir des informations afin de dire si un équipement pourrait tomber en panne.

La mise en place du RUL se fait à travers l'utilisation et l'exploitation de données historiques sur l'état d'un équipement ou machine. Pour cela, elle a été l'objet de plusieurs études par exemple elle a été utilisée pour la prédiction de la durée de vie utile sur des batteries.

Dans le cadre de leur étude portant sur les batteries, les auteurs la définissent comme étant un nombre de cycles nécessaire pour faire décroître la capacité maximale jusqu'au niveau de référence (Zhao, Tian, Cheng, Zhang, & Zhu, 2022).

La présentation du concept de RUL nous permet de mieux appréhender un exemple de maintenance intelligente. Dans le paragraphe suivant, nous proposerons les plans de concepts qui expliquera notre méthodologie pour la recherche bibliographique.

2.3 Définition d'un plan de concept

Notre motivation pour le plan de concept réside dans le fait qu'il est considéré comme élément indispensable pour une bonne recherche bibliographique. Pour cela, nous avons récupéré une liste de travaux pertinents pour la résolution de notre problématique de recherche. Le plan de concept global est fourni ci-dessous.

2.3.1 Plan de concept global pour notre problématique

Le plan de concept global consiste à trouver des termes liés à notre problématique pour cela les mots clés suivants ont été retenus :

- Stratégies
- Intelligentes
- Maintenances
- Production manufacturière
- Contexte
- Économie circulaire

Pour chaque mot-clé, nous lui trouverons des synonymes ainsi nous chercherons des travaux en rapport avec ces mots-clés.

2.3.2 Plan de concept sous-problème 1

Le sous-problème 1 est défini par les leviers de mise en place de la maintenance intelligente, pour cela les mots-clés qui lui sont associés pourraient être :

- Leviers
- Maintenance
- Intelligente

2.3.3 Plan de concept sous-problème 2

Le sous problème 2 définit par l'utilité de la maintenance intelligente dans l'industrie. De la même manière, nous fournissons ci-dessous les mots-clés de référence :

- Maintenance
- Intelligence
- Industrie

2.3.4 Plan de concept sous-problème 3

Le sous problème 3 définit par l'impact de l'économie circulaire en milieu industriel a été analysé et nous permet de citer les mots-clés suivants :

- Impact
- Économie circulaire
- Milieu industriel

2.4 Maintenance 4.0

La maintenance 4.0 est considéré comme une version améliorée des stratégies de maintenance existantes. Pour cela, elle fait appel à l'IoT, à l'intelligence artificielle, à l'exploitation de données. La maintenance 4.0 permet d'améliorer et d'optimiser la planification de la maintenance. Elle regroupe plusieurs stratégies de maintenance comme la maintenance prédictive, la maintenance prescriptive, la maintenance basée sur les conditions. La maintenance 4.0 est de plus en plus rencontrée en milieu manufacturier, en revanche son application reste limitée car de nombreux industriel éprouve des difficultés pour sa mise en application (Ribeiro, Mendes, Forcellini, & Frazzon, 2022). Ainsi Ribeiro et ses collègues ont établi une revue de la littérature portant sur la maintenance 4.0 en utilisant l'analyse SWOT (*Strengths Weaknesses Opportunities Threats*) ou analyse (Forces Faiblesses Opportunités Menaces).

Pour respecter certaines exigences environnementales, la maintenance 4.0 peut être appliquée dans un contexte de développement durable (El kihel, El kihel, & Bouyahrouzi, 2022). Les auteurs ont démontré dans leurs travaux que la maintenance 4.0 associé au développement durable permet d'une part d'augmenter la fiabilité et la disponibilité des équipements et d'autre part de réduire l'impact environnemental.

D'autres travaux ont traité également l'association de la maintenance à l'économie circulaire. Nous citons pour cela les recherches effectuées par Burggräf et ses collègues portant sur l'analyse du cycle de vie pour le reconditionnement adaptatif pour cela les auteurs ont proposé une approche holistique pour mettre en place une stratégie de maintenance dans le but de maximiser la durée du cycle de vie des équipements. Ainsi leurs résultats seront comparés à ceux obtenu par des types de maintenances traditionnelles (Burggräf et al., 2022).

D'autres techniques de maintenance 4.0 ont été développées, nous pouvons citer la réalité augmentée qui peut être défini comme étant « une technologie et un outil qui affiche et crée une vue améliorée de la réalité physique en immergeant des éléments numériques tels que des images, du texte, des sons et des vidéos » (Ruel, 2021). Ainsi la réalité augmentée permet de superposer plusieurs informations. L'utilisation de la réalité augmentée par l'intermédiaire de

lunettes intelligentes permet d'optimiser les tâches de maintenances (Bouabdellah & Aissani, 2022). Leurs travaux ont facilité l'accès aux informations techniques et aux historiques d'intervention. Plusieurs approches peuvent être pris en compte dans la maintenance 4.0, nous présentons celle basée sur les données dans la section suivante.

2.4.1 Approche basée sur les données

L'approche basée sur les données ont été étudiées dans la littérature nous pouvons citer pour cela les travaux portant sur un outil de maintenance intelligente appliqué à un four. Dans cet article, l'accent est mis sur l'automatisation d'un système manufacturier afin de faciliter l'exploitation des ressources qu'elle soit humaine ou machine. Pour cela, les auteurs utilisent la surveillance des données pour déterminer les pannes qui pourraient survenir(Fumagalli, Macchi, Colace, Rondi, & Alfieri, 2016).

Dans le second article, les travaux sont réalisés sur un système manufacturier de type minier. Pour cela, les auteurs utilisent l'IA par l'intermédiaire de l'utilisation, de l'exploitation et de la combinaison des différentes données fournis par des capteurs. Ainsi en utilisant les données historiques collectées sur une période de quatre ans, les auteurs ont mis en place un système de classification des événements de la machine en fonction des caractéristiques de la machine telle que sa puissance(Lever & King, 1993).

L'utilisation des données s'avère important dans une stratégie de maintenance intelligente c'est la raison pour laquelle des chercheurs ont étudié comment stocker des données volumineuses? Pour cela, ils ont mis en place un système de stockage de plusieurs données en se basant sur l'oléoduc et la gazoduc clouding d'un SCADA (*Supervisory Control And Data Acquisition*) (Liu, Shimin, & Mancang, 2018).

Nous pouvons rencontrer le SCADA dans les industries pétrolières, gazières et hydrauliques. Les entreprises manufacturières disposant de pipeline nécessitent une surveillance et un plan de maintenance efficace. C'est la raison pour laquelle des études ont été réalisés pour mettre en place des inspections afin de réduire la probabilité de défaillance. Les données utilisées par les auteurs sont obtenues par l'utilisation d'un PIG (*Pipeline Inspection Gauge*) servant à inspecter et évaluer l'état des pipelines(Sozen, Yurdakul, & Ic, 2022).

Ainsi les données utilisées via l'exploitation du PIG permettront de déterminer les probabilités de défaillances grâce au recours de la simulation Monte Carlo. Pour cela, Le PIG doit être introduit à l'intérieur des pipelines et permettra de détecter divers défauts en l'occurrence les fuites et d'autres anomalies. La littérature étudiée nous renseigne sur l'utilité de disposer de données de qualité afin de mettre en place une bonne stratégie de maintenance. Dans la prochaine section, nous présenterons des études faites sur la maintenance prédictive dans l'industrie.

2.4.2 Stratégie de maintenance prédictive

La maintenance prédictive est largement représentée dans la littérature. Nous présentons un article qui parle de maintenance prédictive. Les auteurs ont mis en place le ML qui permettra de détecter les sons anormaux.

Ils ont utilisé un ensemble d'entraînement constitué de sons normaux puis de tester leur modèle sur différents types de sons normaux et anormaux. Ainsi les résultats obtenus permettront de définir un planning de maintenance (Mrazovac, Ilian, & Hulea, 2021).

Un autre exemple de maintenance prédictive a été développé sur des réseaux de neurones, pour cela nous citons les travaux de Fordal et ses collègues. Leur méthodologie consistait à utiliser les données collectées en entrée afin d'alimenter un réseau de neurone artificiel ANN (Fordal et al., 2023).

L'utilisation du clustering est également traitée dans la littérature. De ce fait, nous citons les travaux réalisés par Calzavara et ses collaborateurs. Ils ont mis en place un système de maintenance prédictif qui permettra de prédire l'évolution temporelle de l'état de santé d'une machine. Ils procèdent par une segmentation temporelle suivie de l'utilisation d'un DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*).

Ainsi le DBSCAN permettra de regrouper les points segmentés. La motivation principale du DBSCAN résulte du fait qu'il pourra être appliqué lorsque les données collectées ne sont pas suffisantes (Calzavara, Oliosi, & Ferrari, 2021).

En dehors du clustering, nous avons rencontré des travaux basés sur la classification. Il s'agit d'études réalisés sur la surveillance de boîte de vitesse planétaire (PGB) utilisés dans l'industrie automobile (Sharma & Tiwari, 2022). Les auteurs ont utilisé le concept de l'extraction de caractéristiques.

Ainsi, ils ont employé plusieurs classifieurs dont on peut citer :

- La machine à vecteur de support (SVM).
- La forêt aléatoire (RF).
- Les K plus proches voisins (KNN).
- Le perceptron multicouche (MLP).

Cependant dans d'autres travaux, nous remarquons qu'une stratégie de maintenance prédictive est adoptée dans le but de détecter les pannes futures. Leur méthodologie repose sur la mise en place d'une ML basée sur le modèle ARIMA (*AutoRegressive Integrated Moving Average*) (Mohan, Roselyn, Uthra, Devaraj, & Umachandran, 2021). En ce qui concerne le modèle ARIMA, il sert à prédire une évolution ou une tendance future d'une variable (Lelong, s.d.)

La planification de la maintenance est un domaine qui intéresse de plus en plus les chercheurs. Nous pouvons citer les études réalisées sur une planification optimale de la production et de la maintenance (Antao, Reis, & Goncalves, 2018). Les auteurs ont mis en place trois modules que sont :

- Un module de maintenance prédictive en utilisant plusieurs classifieurs.
- Un module d'optimisation qui utilise un algorithme génétique.
- Un module de simulation basée sur des données réelles afin de vérifier l'optimisation conjointe de la production et de la maintenance.

De la même façon, la planification de la maintenance a été associée avec un modèle d'optimisation multi-objectifs respectant les contraintes de sécurité c'est le cas de travaux portant sur une maintenance prescriptive. Les auteurs ont utilisé plusieurs classifieurs afin de déterminer la défaillance future des composants et l'ont adapté à un modèle d'optimisation en utilisant GAMS (*General Algebraic Modeling System*) (Gordon et al., 2020).

Étant donné que les méthodes exactes ne permettent pas de trouver facilement les solutions optimales, d'autres chercheurs ont tenter les méthodes approchées via les méta-heuristiques pour résoudre des problèmes de planification de maintenance. Ainsi nous citons les travaux réalisés dans le cadre de plusieurs machines parallèles non identique et prenant en compte la maintenance.

De ce fait les auteurs ont comparé les résultats obtenus par deux heuristiques différents qui vont minimiser le retard associé à l'exécution de la maintenance.

Ensuite un modèle MILP (*Mixed Integer Linear Program*) ou programmation linéaire mixte en nombre entiers sera utilisé afin d'optimiser les résultats obtenus (Beldar, Moghtader, Giret, & Ansariipoor, 2022).

2.5 Économie circulaire

L'économie circulaire prend en compte l'aspect de la réutilisation des produits en fin de vie. Pour cela, nous citons les travaux effectués sur l'impact de l'économie circulaire dans l'industrie papetière. Les auteurs ont mis l'accent sur l'utilisation des déchets de papiers afin de fabriquer des produits secondaires (Sakthivelmurugan, Senthilkumar, & Karthick, 2022).

D'autres travaux se sont penchés sur l'impact de l'industrie 4.0 sur l'économie circulaire. Ainsi les auteurs ont démontré à travers leurs études que les différentes techniques liées à l'industrie 4.0 peuvent influencer la mise en œuvre et la réussite de l'économie circulaire (Lei, Cai, Cui, Wu, & Liu, 2022).

Parmi les techniques de l'industrie 4.0 utilisés par Lei et ses collègues, nous citons l'IoT, la fabrication additive, le big data et l'IA.

Schögl et de ses collaborateurs traitent l'économie circulaire en utilisant les technologies numériques. Ils s'appuient sur l'analyse dite descriptive hiérarchiques de type grappe et non paramétriques en se basant sur des test de Kruskal-Wallis et la corrélation de rang de spearman (Schögl, Rusch, Stumpf, & Baumgartner, 2023).

L'économie circulaire attire de plus en plus les scientifiques c'est la raison pour laquelle les études de Chen et de ses confrères ont porté sur la réduction des émissions et la neutralité de carbone. Pour cela, ils ont essayé de quantifier les revenus obtenus par la mise en place de

bornes de recharges électrique. En effet leur méthodologie consistait à utiliser un modèle d'optimisation multi-objectif afin de maximiser les bénéfices obtenus (Chen Xianda et al., 2021).

L'économie circulaire a été traitée dans la littérature accompagnée de la recherche opérationnelle c'est le cas de travaux effectués par Rajput et ses collaborateurs, ils ont traité l'économie circulaire associée à la logistique inverse dans le contexte de l'industrie 4.0. Un modèle MILP a été développé, par ailleurs ils utilisent le déploiement des capteurs IoT dans leur modèle. Ainsi leur modèle d'optimisation permettra d'une part de minimiser les coûts et d'autre part de maximiser la durée de vie des produits (Rajput & Singh, 2022).

La chaîne industrielle écologique a été étudiée dans des travaux, Des chercheurs ont étudié l'impact de la théorie de fiabilité dans l'économie circulaire. Ils ont conclu que les résultats changent en fonction de la disposition des structures donc plus la structure est disposée en série plus la stabilité est faible en revanche plus la structure est disposée en parallèle plus la stabilité est forte.

Ainsi la planification de la disposition est résolue par l'utilisation d'une programmation non linéaire (Wang et al., 2016).

Le Tableau-A I-1 présent dans l'annexe résume l'ensemble des travaux pertinents rencontrés dans la littérature.

2.6 Conclusion

La revue de la littérature est un étape clé dans la résolution d'une problématique de recherche car elle permet de montrer l'ensemble des travaux et études existantes dans le domaine. Ainsi les recherches antérieures ont facilité notre compréhension dans les thèmes de la maintenance intelligente associé à l'économie circulaire. La revue de la littérature nous a fourni différentes méthodologies qui nous ont inspirer à traiter notre sujet de recherche.

Malgré qu'il y'ait autant de travaux scientifiques sur notre sujet, nous avons constaté que la plupart des travaux étudiés n'ont pas pris en compte la planification de la maintenance préventive, Or la maintenance prédictive permet de minimiser la maintenance corrective au

profil de la maintenance préventive. Dans la suite de notre travail, nous prenons en compte la planification de la maintenance préventive.

À l'issue de la revue de la littérature, nous avons déduit que le traitement et l'exploitation des données pourrait être une hypothèse pertinente pour la résolution du sous-problème 1. De la même façon une stratégie de maintenance intelligente pourrait être la mise en place d'une maintenance prédictive afin de prédire les pannes d'équipement.

Tandis que la résolution de notre sous-problème 3 passera par la mise en place d'un modèle d'optimisation de type MIQP (*Mixed-Integer Quadratic Programming*) ou programmation quadratique à variables Mixtes.

Les différentes hypothèses obtenues par l'étude de la revue de la littérature seront traitées dans les chapitres suivantes.

CHAPITRE 3

LES LEVIERS DE MISE EN PLACE DE LA MAINTENANCE INTELLIGENTE

3.1 Introduction

Dans ce chapitre, nous abordons les outils que nous disposons pour proposer une stratégie de maintenance intelligente. Ainsi nous considérons que l'exploitation des données nous permettrais de résoudre ce sous-problème. Il convient de noter que la collecte des données représente l'étape initiale pour tout projet de ML. Elle permet de regrouper un ensemble d'information dont le but est de résoudre une problématique bien spécifique. Elle implique la prise en compte de diverses sources d'information provenant de lieu et d'échelle de temps différents. Ainsi les informations collectées serviront ou aideront à prendre des décisions sur le court, moyen et long terme. Notre méthodologie dans ce chapitre sera résumée en trois grands axes que sont :

- Une présentation des sources de données.
- Une description des outils statistiques.
- Une analyse exploratoire des données.

Nous traiterons successivement chaque grand axe de notre méthodologie afin de couvrir les différentes spécificités de ce sous-problème.

3.2 Présentation des sources de données

Les données peuvent être obtenus de deux façons différentes que sont la méthode qualitative et la méthode quantitative. En ce qui concerne la méthode qualitative, elle nécessite de passer par une étape d'analyse et de compréhension des données et est défini dans la littérature par ces termes « la technique de collecte qualitative se base principalement sur l'analyse de plusieurs facteurs afin de fournir une compréhension approfondie des données brutes »(Lelong, 2023). Elle se différencie de la méthode quantitative, qui est considéré comme étant très pratique pour obtenir des données.

Ainsi nous dirons que la méthode quantitative permet de fournir des informations sous forme numérique donc plus manipulable en termes de modélisation. C'est dans ce sens que cette méthode quantitative a été utilisée pour l'acquisition des données dans notre mémoire. L'une des techniques les plus utilisés en milieu industriel pour la collecte de donnée est la télémétrie. En effet la télémétrie est considérée comme une méthode servant à collecter, à mesurer et à transmettre différentes sources de données. La particularité de la télémétrie résulte du fait que les données peuvent provenir d'équipement où la communication se fait à distance. Nous retrouvons cette méthode dans plusieurs secteurs industriels comme :

- L'industrie aéronautique.
- L'industrie automobile.
- L'industrie de l'informatique et du numérique.
- L'industrie des télécommunications.

La télémétrie utilise des capteurs qui collectent des données. Ces données peuvent être des grandeurs physiques différentes par exemple la tension, la vibration ou autres. Cependant l'internet est souvent utilisé pour transmettre à distance les données. Ainsi le débit de transfert ou la fréquence peut se faire de façon continu ou discrète ou à intervalle régulier. Les données obtenues peuvent être regroupées et organisées sous différentes séries.

Nous définirons une série comme étant un regroupement organisé de plusieurs sources de données. Ainsi selon la nature de l'étude nous retrouvons plusieurs types de séries qui seront expliqué brièvement ci-dessous.

3.2.1 Séries temporelles

Une série temporelle se définit comme étant un regroupement de données suivant un ordre chronologique. Chaque observation ou mesure est effectué à un instant t . Les séries temporelles sont fréquemment utilisés en industrie car elles permettent de suivre l'évolution et la tendance des données en fonction du temps. Une série temporelle peut regrouper plusieurs concepts que sont la tendance, la fluctuation saisonnière et la fluctuation cyclique (Parent, s.d.).

3.2.2 Séries transversales

Une série transversale est appelée encore étude transversale. Elle fournit des informations sur des individus à un instant t et ne présentent pas d'expérience. Dans une étude transversale, les individus traités ont les mêmes caractéristiques ce qui fait que la comparaison ne se basera pas sur une évolution. Une étude transversale s'applique sur un échantillon de grand taille (Voxco, 2023b). En revanche l'étude transversale ne permet pas de prouver qu'il y'a une relation de cause à effet, de ce fait une étude longitudinale pourrait être privilégié dans certaines situations.

3.2.3 Séries longitudinales

Les séries ou étude longitudinale permettent de faire des mesures répétitives sur les mêmes individus dans le but de suivre leur évolution dans le temps (Thomas, 2020). Elle ne requiert pas de durée spécifique du moment que les observations sont effectuées à plusieurs reprises. Ce type d'étude est fréquemment rencontrés en milieu médical mais aussi dans le secteur de l'économie de même que dans l'épidémiologie. L'étude longitudinale a pour principale avantage de suivre l'évolution des individus dans le temps mais également d'établir des relations de cause à effet. Cette étude ne convient pas à tous les cas de figure, car sa mise en œuvre est souvent coûteuse.

3.2.4 Séries panel

Une série est dite panel lorsqu'elle fait appel simultanément à des éléments de séries temporelles et transversales. De ce fait, elles utilisent des données collectées suivant un ordre chronologique et en même temps ces données seront issues de sources différentes. Les données collectées et regroupés en série seront étudiés à travers des outils statistiques que nous décrivons dans la prochaine partie.

3.3 Description des outils statistiques

Les statistiques peuvent être considérés d'une part comme des supports qui nous aide à comprendre les données et d'autre part nous fournit une analyse cohérente des informations collectées. Il existe plusieurs types de statistiques utilisées dans l'étude des données nous en citons les statistiques non paramétriques, qui sont utilisés pour analyser des données lorsque les hypothèses ne sont pas prises en compte. D'autres statistiques seront décrites ci-dessous.

3.3.1 Statistiques temporelles

Les statistiques temporelles permettent d'analyser les séries temporelles. Parmi les techniques qui y sont utilisés, nous pouvons citer les méthodes de lissage et les modèles ARIMA. Le lissage peut être utilisé car il pourra traiter les imperfections et les écarts de mesures dans les observations. Le modèle ARIMA est également utilisé car il est considéré comme un outil d'analyse statistique applicable dans les séries temporelles pour faire des prédictions. Le terme autorégressif signifie l'utilisation de données antérieures pour prédire les données futures.

3.3.2 Statistiques multivariées

Les statistiques multivariées permettent de traiter simultanément des données disposant de plusieurs variables. Parmi les techniques qui y sont utilisé, nous pouvons citer l'analyse en composantes principales (ACP) et l'analyse de corrélation. À propos de l'analyse en composantes principales (ACP), elle s'utilise pour réduire la dimensionnalité dans un ensemble de données. De ce fait, l'ACP va créer de nouvelles variables non corrélées dites composantes principales issue de l'ensemble de donnée initiales (Jolliffe & Cadima, 2016). Par ailleurs les nouvelles variables corrélées devront conserver au maximum la variabilité encore appelée information statistique. En revanche l'analyse de corrélation permet de déterminer la relation qui existe entre variables.

3.3.3 Statistiques inférentielles

Une statistique inférentielle permet de déduire des conclusions. L'utilisation des statistiques inférentielles permet d'estimer ou d'évaluer les paramètres, de connaître les intervalles dits de confiance et de faire des tests d'hypothèses. Un intervalle de confiance se définit comme un ensemble de valeurs susceptible de contenir la valeur d'un individu inconnu (Minitab, 2021). La figure ci-dessous illustre le principe d'un intervalle de confiance

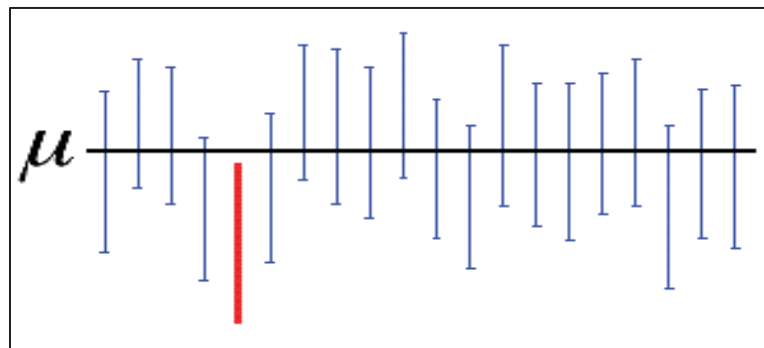


Figure 3.1 Intervalle de confiance tirée de Minitab (2021)

3.3.4 Statistiques descriptives

Ce sont des types de statistiques fréquemment utilisés en analyse de données. Elles permettent de fournir une présentation intuitive des données en utilisant des mesures comme la moyenne, la médiane, l'écart type ainsi que les quantiles. Ces mesures constituent la base de l'analyse des données.

Pour cela la moyenne désigne la valeur centrale de données étudiés. Cependant la présence de valeurs aberrantes entraîne une variation de la moyenne contrairement à la médiane qui n'est pas impactée par les valeurs aberrantes ou extrêmes. Elle est donnée par la formule suivante :

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (3.1)$$

Avec :

μ = la valeur de la moyenne à calculée.

x_i = la valeur pour chaque observation.

n = le nombre total d'observation.

En ce qui concerne la médiane, elle désigne une valeur qui permet de diviser un ensemble de données en deux parties égales. Ainsi la moitié des données se trouvera en dessous de la médiane et l'autre moitié se trouvera au-dessus de la médiane.

L'écart type est souvent utilisé et permet donc de mesurer la dispersion des valeurs dans un ensemble de données. Il montre comment les observations individuelles s'écartent de la moyenne. Il pourra être calculé par la formule suivante.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad (3.2)$$

Avec :

σ = écart type à calculer

μ = valeur de la moyenne à calculée.

x_i = valeur pour chaque observation.

n = nombre total d'observation.

Afin de vérifier comment nos données sont distribuées, nous utilisons la division des données. Ainsi nous parlerons de quartile.

Le premier quartile permet de diviser un ensemble de données en quatre parties égales. Ainsi 25% des données observées seront en dessous du premier quartile et les 75% restantes se situeront au-dessus du premier quartile. Dans la littérature nous rencontrons le premier quartile sous l'appellation Q1.

Cependant il faut noter que la division des données nécessite de les trier d'abord suivant l'ordre croissant ainsi nous pourrions déterminer la médiane, le premier quartile et le troisième quartile. Le troisième quartile est nommé Q3 et représente le complément de Q1, c'est-à-dire que 75%

des données se trouverons en-dessous de Q3 et seulement 25% des données restantes seront au-dessus de Q3.

Ainsi nous définirons l'intervalle interquartile (IQR) et s'exprime par la différence de Q3 et Q1. En analyse de données l'intervalle interquartile est fréquemment utilisé pour mesurer la dispersion et la variabilité des données et se calcule donc comme suit :

$$IQR = Q3 - Q1 \quad (3.3)$$

Toutes ces notions statistiques pourront être représenter sous formes graphique en utilisant le diagramme de la boite à moustache ou diagramme en boite ou *boxplot*.

Le diagramme en boite représente un graphique qui permet de résumer à lui seul la statistique descriptive. Il permet de visualiser une distribution de données de même que ses caractéristiques. Parmi les mesures statistiques affichés dans un diagramme en boite, nous pouvons en citer :

- Le minimum.
- Le premier quartile.
- La médiane.
- Le troisième quartile.
- Le maximum.

Ainsi les lignes sont appelées des moustaches et les valeurs observées au-delà de IQR sont appelés des valeurs extrêmes ou aberrantes.

La figure suivante montre le principe et la représentation d'un diagramme en boite.

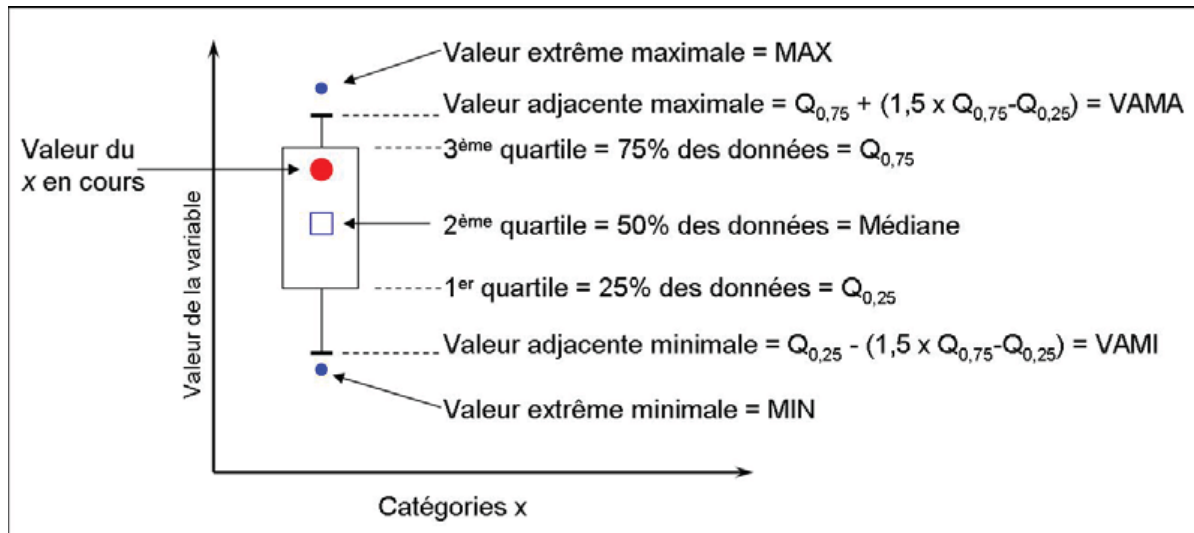


Figure 3.2 Diagramme en boîte tirée de Menet-Nédélec, Lampert, & Riou (2011)

3.4 Analyse Exploratoire des données

L'analyse des données a été effectuée sur Google Colab avec le langage python. Par ailleurs, l'ensemble des données utilisés dans ce mémoire peuvent être trouver dans les travaux de recherche effectués par l'équipe d'apprentissage automatique de Microsoft Azure (Fidan Boylu Ouz, 2016).

Il s'agit de cinq sources de données présentées sous forme de fichiers csv. Nous définirons un fichier csv ou *Comma-separated Values* comme étant un type de fichier qui permet de stocker et d'échanger des informations sous forme tabulaire.

Dans un tel fichier, les valeurs sont séparées par des virgules mais elles peuvent également être séparés par des délimiteurs comme des points-virgules et des tabulations.

Un fichier csv peut être présenté sous forme matricielle ou tableau disposant d'un nombre de lignes et de colonnes. Une ligne du fichier représente une observation ou point de mesure ou enregistrement tandis que chaque colonne désigne une variable ou caractéristique. L'ensemble du fichier csv est appelé un jeu de donnée. Il existe plusieurs formats de fichier utilisés dans la collecte de données nous pouvons en citer les fichiers JSON (*Javascript Object Notation*), les fichiers Excel (XLSX) etc.

Nos fichiers utilisés dans nos travaux sont décrits de la manière suivante:

- Données de série chronologique de télémétrie présentées dans **PdM_telemetry.csv**
- Données sur les erreurs collectées dans les machines présentées dans **PdM_errors.csv**
- Données sur le nombre de maintenance présentées dans **PdM_maint.csv**
- Données sur les défaillances des composants présentées dans **PdM_failures.csv**
- Métadonnées des machines présentées dans **PdM_Machines.csv**

L'ensemble de ces données fournissent des informations sur un total de cent machines. Dans notre étude seulement les données des trois premières machines seront considérées et non la totalité des machines. Par ailleurs, une analyse approfondie sera appliquée dans la partie suivante afin de voir la nature et l'évolution de ces données dans le temps.

3.5 Analyse des données de télémétrie

Les données présentées dans le jeu de donnée de la télémétrie concernent la tension, la rotation, la vibration et la pression des machines. Ils sont moyennés par heure et déployés sur trois machines pendant l'année 2015. Les données observées sont de type série chronologique car nous observons par exemple sur la figure 3.2 l'évolution des données de tension suivant le mois de janvier.

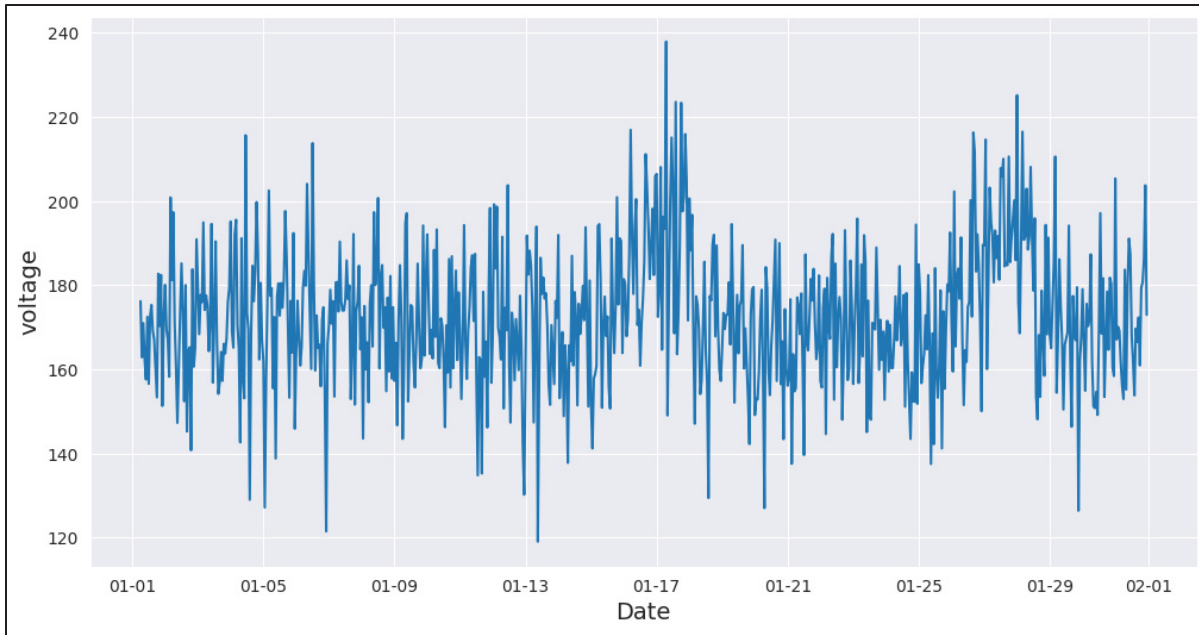


Figure 3.3 Distribution de la tension suivant le mois de janvier

Nous pouvons observer les trois premières valeurs du jeu de donnée de télémétrie suivant le tableau ci-dessous.

Tableau 3.1 Les premières valeurs de télémétrie

Numéro d'observation	datetime	machineID	Volt	rotate	pressure	vibration
0	2015-01-01 06:00	1	176,217853	418,504078	113,077936	45,0876858
1	2015-01-01 07:00	1	162,879223	402,74749	95,4605254	43,4139727
2	2015-01-01 08:00	1	170,989902	527,349826	75,2379049	34,1788471

3.5.1 Statistique descriptive du jeu de donnée télémétrie

La statistique descriptive du jeu de donnée de télémétrie est fournie par le tableau 3.2 et nous renseigne que notre jeu de donnée contient 26283 points de mesure ou observations. Pour chaque variable observée nous pouvons vérifier la distribution et la répartition de ces données. Ainsi l'éloignement des valeurs minimales et maximales permet de supposer qu'il y aurait peut-être des valeurs extrêmes.

La présence de valeurs aberrantes dans certains cas peut entraîner un biais dans l'analyse et l'interprétation des résultats. Une visualisation par le diagramme en boîte sera tracée afin de vérifier les variables télémétriques. La comparaison des variables permet de mieux comprendre les distributions et de définir les variables qui semblent être corrélés entre elles. Ainsi la distribution peut être visualisé sur la figure 3.3.

Tableau 3.2 Statistique descriptive du Jeu de données de Télémétrie

	machineID	volt	rotate	pressure	vibration
Nombre d'observation	26283	26283	26283	26283	26283
Moyenne	2	170,7512048	446,4362181	100,6204499	40,45440787
Écart-type	0,816512114	15,4105105	52,72674422	10,84246802	5,370604325
Valeur minimale	1	112,4866135	174,1269683	54,91401347	22,07808667
Premier quartile	1	160,4017388	412,7303997	93,34837758	36,82786457
Médiane	2	170,5485512	447,6694962	100,3035329	40,3244207
Troisième quartile	3	180,8834301	481,5691482	107,2822148	43,83843836
Valeur maximale	3	237,938518	636,364569	155,2575272	67,6334348

Les données pour chacune des trois machines sont suivies et une visualisation des caractéristiques est donnée ci-dessous. L'analyse du graphique permet de voir que les données sont groupées en cluster. Cependant la visualisation ne permet pas de dire qu'il y aurait une corrélation entre les variables.

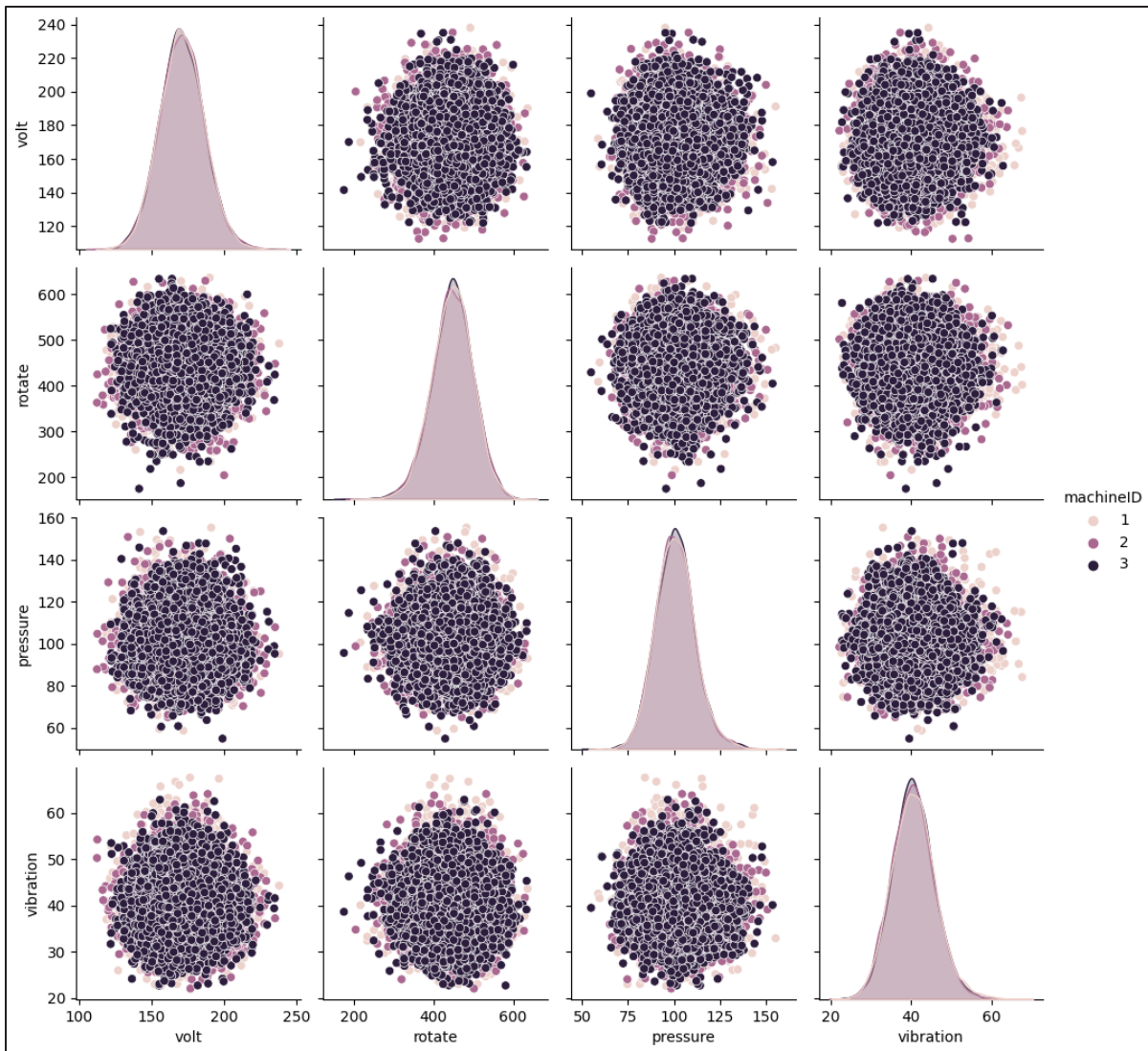


Figure 3.4 Distribution du jeu de données de télémétrie en fonction des 3 machines

3.5.2 Distribution des variables numérique de télémétrie

Un jeu de donnée dispose de plusieurs variables ou caractéristiques qui peuvent être divisée généralement en deux grandes familles que sont :

- Les variables qualitatives ou catégorielles.
- Les variables quantitatives ou numériques.

Les variables qualitatives sont appelées encore variables catégoriques et se décompose à leur tour en deux sous-parties que sont les variables nominales et les variables ordinales. Au niveau des variables nominales il n'y a pas d'ordre particulier tandis que les variables ordinales sont des catégories avec un ordre ou une hiérarchie bien définie.

Les variables quantitatives sont les plus rencontrés dans les travaux de recherche scientifique car elles sont fournies par les mesures, les calculs et les estimations. Ainsi les variables quantitatives peuvent être divisées en quatre parties sous forme de variables continues, variables discrètes, variables binaires et les variables temporelles.

Les variables continues regroupent des données qui peuvent prendre n'importe quelle valeur dans un intervalle ou plage de donnée. Elles se différencient des variables discrètes où ces valeurs sont entières et non fractionnaires. En ce qui concerne les variables binaires elles regroupent les données ne pouvant prendre que deux valeurs.

Cependant, les variables temporelles sont très utilisées dans les séries chronologiques car ses données sont recueillies et enregistrées par date et heure. Ainsi pour tous les jeux de données utilisées dans nos recherches, les variables seront traitées en fonction de leur appartenance c'est-à-dire qu'elle soit qualitative ou quantitative. L'organigramme est présenté à la figure 3.4 pour mieux comprendre les différences entre variable qualitative et quantitative.

Notre jeu de donnée de télémétrie contient six variables où la nature de chaque variable est résumé dans le tableau ci-dessous. Étant donné que la variable datetime doit être de type temporel et non un objet. Le type object en science de donnée fait référence à un type de données générique d'où sa transformation montrée par la colonne variable finale.

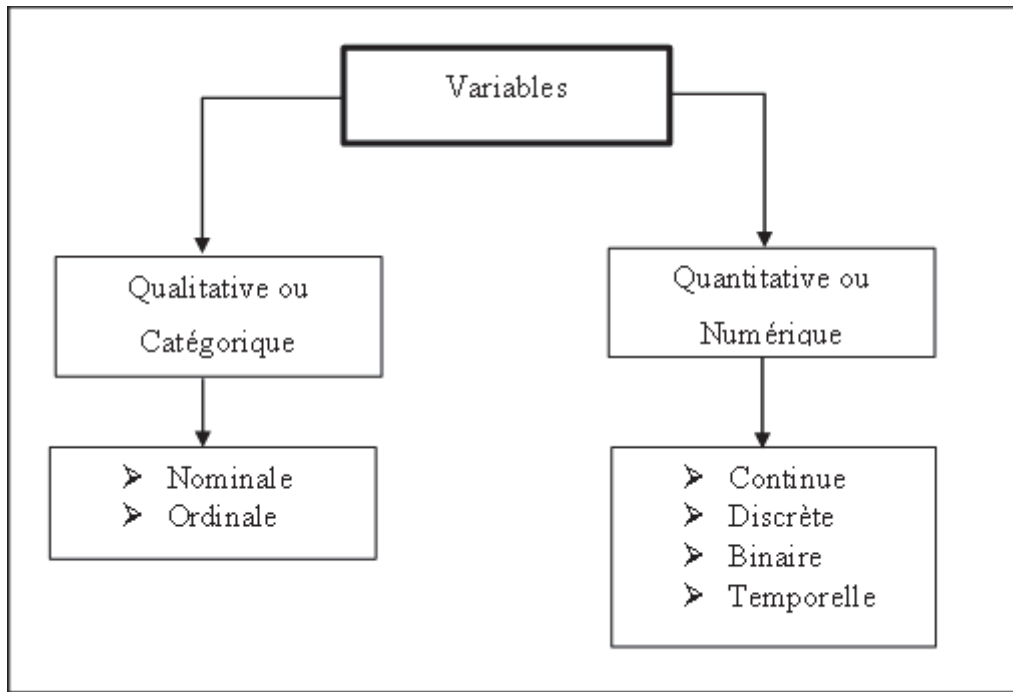


Figure 3.5 Organigramme d'une variable

Tableau 3.3 Les types de variables du jeu de données de télémétrie

Variables	Type de variable initiale	Type de variable finale
datetime	object	Temporelle
machineID	Entière	Entière
volt	Réelle	Réelle
rotate	Réelle	Réelle
pressure	Réelle	Réelle
vibration	Réelle	Réelle

Nous pouvons remarquer que le jeu de donnée de télémétrie dispose de quatre variables numériques continues que sont les variables de la tension, la rotation, la pression et la vibration. Par ailleurs la variable machineID est type numérique discrète. La détermination du type variable constitue l'une des premières étapes dans l'analyse des données. Ainsi nos variables numériques seront tracées à l'aide du diagramme en boîte.

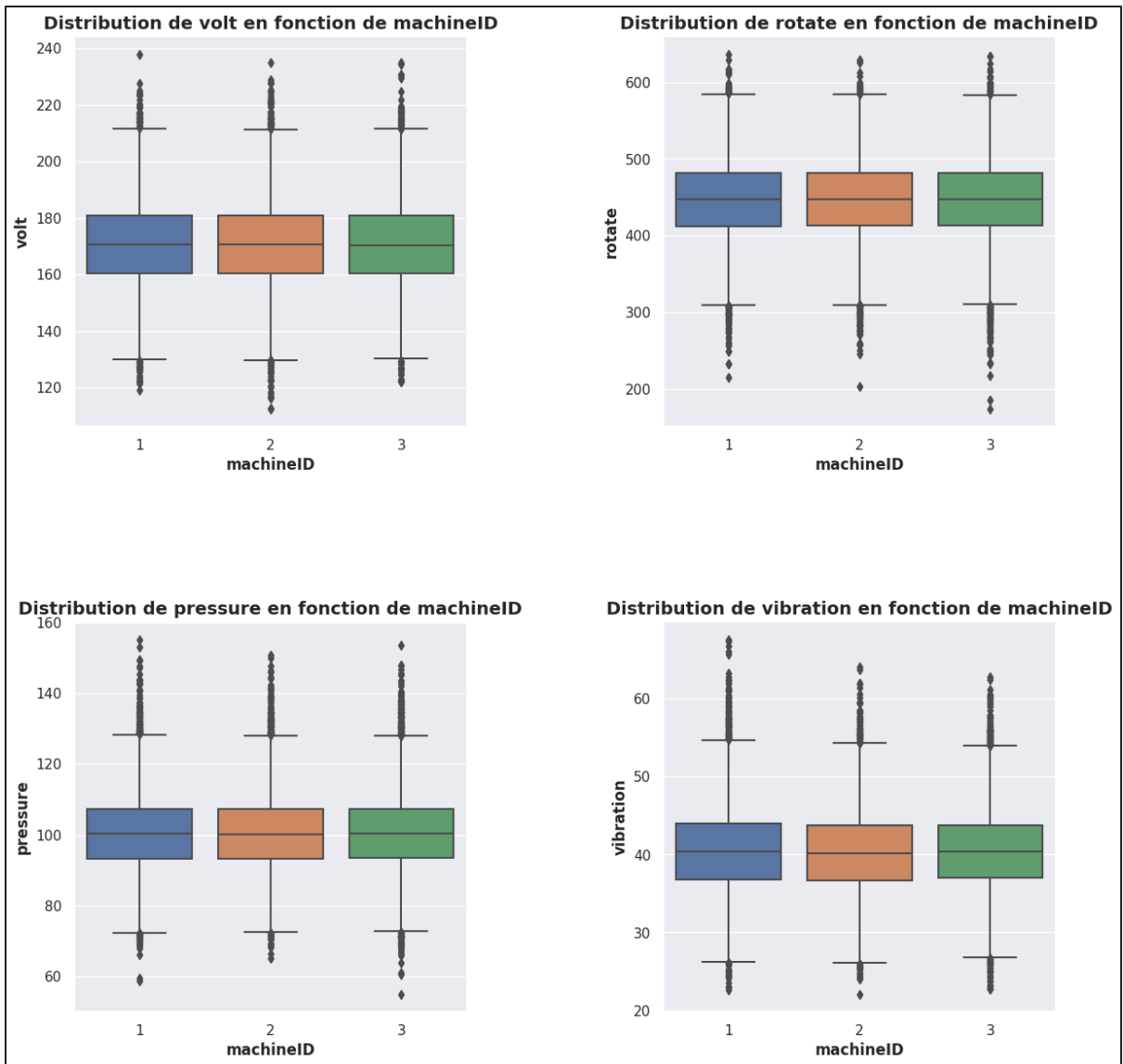


Figure 3.6 Diagramme en boîte des variables numériques du jeu de donnée télémétrie

L'analyse du diagramme en boîte confirme bien que les variables numériques présentent de nombreuses valeurs extrêmes. Ces valeurs seront maintenues car notre recherche se fait sur des séries chronologiques. Un changement de ces valeurs pourrait entraîner un biais sur les résultats.

3.6 Analyse des données sur les erreurs

Ce jeu de données regroupe les erreurs survenues sur les trois machines en fonctionnement. Ces erreurs n'empêchent pas les machines de continuer leur fonctionnement. Ils ne doivent pas être considérés comme des pannes. Ces erreurs sont collectées à des fréquences variables. Il y a cinq types d'erreurs numérotés error1 jusqu'à error5. Le tableau ci-dessous présente les premières lignes du jeu de donnée des erreurs.

Tableau 3.4 Les cinq premières lignes du jeu de données des erreurs

Numéro observation	datetime	machineID	errorID
0	2015-01-03 07:00:00	1	error1
1	2015-01-03 20:00:00	1	error3
2	2015-01-04 06:00:00	1	error5
3	2015-01-10 15:00:00	1	error4
4	2015-01-22 10:00:00	1	error4

Nous remarquons que la variable errorID est une variable de type qualitative ou catégorique. L'analyse des types de variables du jeu de données des erreurs est fourni ci-dessous. Ainsi la variable datetime a été transformée en type temporelle.

Tableau 3.5 Les types de variables du jeu de données des erreurs

Variables	Type de variable initiale	Type de variable finale
datetime	Object	Temporelle
machineID	Entière	Entière
errorID	Object	Object

Dans ce jeu de donnée les statistiques descriptives ne pourraient pas être appliquées car la variable errorID n'est pas quantitative numérique. Ainsi une visualisation de la variable errorID permettrait de connaître la répartition des erreurs sur les trois machines.

En milieu industriel les erreurs rencontrées peuvent être de différentes sortes. Ainsi nous citons cinq types d'erreur qui pourraient être rencontrés sur les équipements manufacturiers. Elles peuvent être :

- Des erreurs de type mécanique.
- Des erreurs de type informatique.
- Des erreurs de type électrique.
- Des erreurs de type automatique.
- Des erreurs d'entretien.

Parmi les erreurs de type mécanique, nous pouvons citer un désalignement de convoyeur, une lubrification insuffisante entraînant une usure prématurée des composants. Les erreurs de type informatique serait par exemple un dysfonctionnement de logiciel entraînant une programmation non conforme sur l'exécution des tâches automatisées. Les erreurs de type électrique peuvent être rencontrés sous forme de panne de courant et de courts-circuits. Les erreurs de type automatique peuvent concerner les capteurs connectés qui présenteraient des problèmes de calibration. Enfin les erreurs d'entretien peuvent être une maintenance incomplète entraînant un ralentissement de la cadence de production.

Dans nos travaux la nature des erreurs n'est pas précisée donc nous allons les considérer comme des erreurs classiques et adaptable selon n'importe quel contexte manufacturier. La fréquence des erreurs est fournie par l'histogramme des erreurs et nous montrent que plus le numéro d'erreur augmente plus sa fréquence baisse. Ainsi les erreurs de type 1 et 2 sont les plus fréquentes cependant la fréquence des erreurs de type 5 est presque la moitié de celle du type 1 et 2.

Étant donné que les erreurs ne sont pas des pannes, nous pouvons supposer que l'augmentation des erreurs pourrait engendrer une dégradation accélérée des machines. Pour cela, une réduction des erreurs de type 1 et 2 serait bénéfique au planning de maintenance de même que sur la réduction du temps et du coût de la maintenance.

L'analyse des diagrammes circulaires nous montrent les taux en pourcentage de type d'erreur rencontrés ainsi nous dirons qu'au niveau de la machine 1, l'erreur de type 1 est majoritaire. Les erreurs rencontrées de type 4 et 5 sont presque identiques. Dans cette machine, les erreurs 1 et 3 sont les plus observés.

En ce qui concerne la machine 2, les erreurs de type 2 sont très fréquents mais nous remarquons que le nombre d'erreur de type 1 est égal à celui de type 3. Dans cette machine les erreurs de type 5 sont minoritaire. Enfin pour la machine 3, l'erreur 1 est le plus rencontré suivi de l'erreur de type 2. Cependant les nombre d'erreurs 3 et 4 sont identiques. Pareillement aux autres machines l'erreur de type 5 est très peu observé.

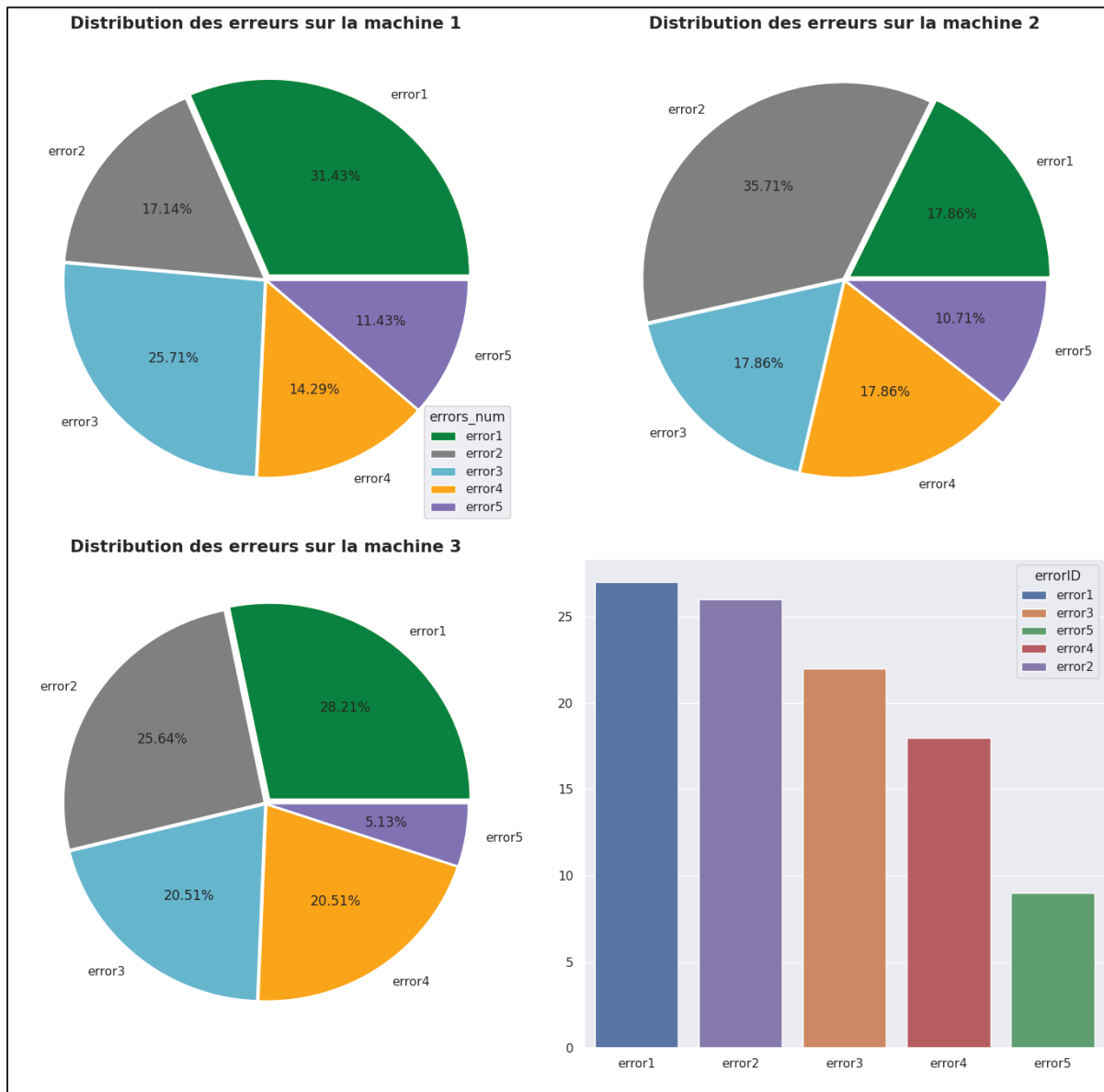


Figure 3.7 Distribution des erreurs suivant les machines

3.7 Analyse des données de maintenance

Ce jeu de données regroupe les enregistrements de remplacement des composants lors d'une maintenance préventive. Il convient de noter qu'un composant peut être remplacé dans deux situations différentes que sont lors d'une maintenance préventive ou corrective.

Dans le cas d'une maintenance corrective, les composants remplacés sont capturés et placés dans le fichier nommés échecs ou pannes ou Failure. Ces données de maintenance ont été collectés par heure durant l'année 2014 - 2015.

La maintenance est effectuée sur les composants à des fréquences variables. Ainsi nous disposons de quatre composants à maintenir nommé successivement comp1 jusqu'à comp4. Le tableau ci-dessous présente les premières lignes du jeu de donnée de maintenance.

Tableau 3.6 Les cinq premières lignes du jeu de données de maintenance

Numéro observation	datetime	machineID	comp
0	2014-06-01 06:00:00	1	Comp2
1	2014-07-16 06:00:00	1	Comp4
2	2014-07-31 06:00:00	1	Comp3
3	2014-12-13 06:00:00	1	Comp1
4	2015-01-05 06:00:00	1	Comp4

Nous remarquons que la variable comp est une variable de type qualitative ou catégorique. L'analyse des types de variables du jeu de données de maintenance est fournie ci-dessous. Ainsi la variable datetime a été transformée en type temporelle.

Tableau 3.7 Les types de variables du jeu de données de maintenance

Variabes	Type de variable initiale	Type de variable finale
datetime	Object	Temporelle
machineID	Entière	Entière
comp	Object	Object

Dans ce jeu de données les statistiques descriptives ne pourraient pas être appliquées car la variable comp n'est pas quantitative numérique. Ainsi nous ferons une visualisation comme sur le jeu de données des erreurs afin de connaître la répartition des composants maintenues sur les trois machines.

La fréquence des composants maintenus sur l'ensemble des trois machines peut être visualisée par la figure 3.8. Elle permet de constater que la maintenance préventive du composant 2 est plus fréquente tandis que le composant 3 est le moins maintenu. Selon la figure 3.8 nous remarquons qu'au niveau mensuel le nombre de maintenance pour tous les composants était plus faible en 2014 alors qu'à partir de 2015 nous observons que le nombre de dossiers de maintenance a augmenté. Cependant le nombre de maintenance préventive effectué sur le composant 1 au deuxième trimestre de 2015 a baissé comparé à celui du composant 2. Ainsi ce jeu de données de la maintenance préventive permettra de supposer que le taux d'utilisation des machines en 2015 était plus élevé que celui de 2014 car plus une machine est opérationnelle plus vite l'état de ses composants se dégraderont dans le temps.

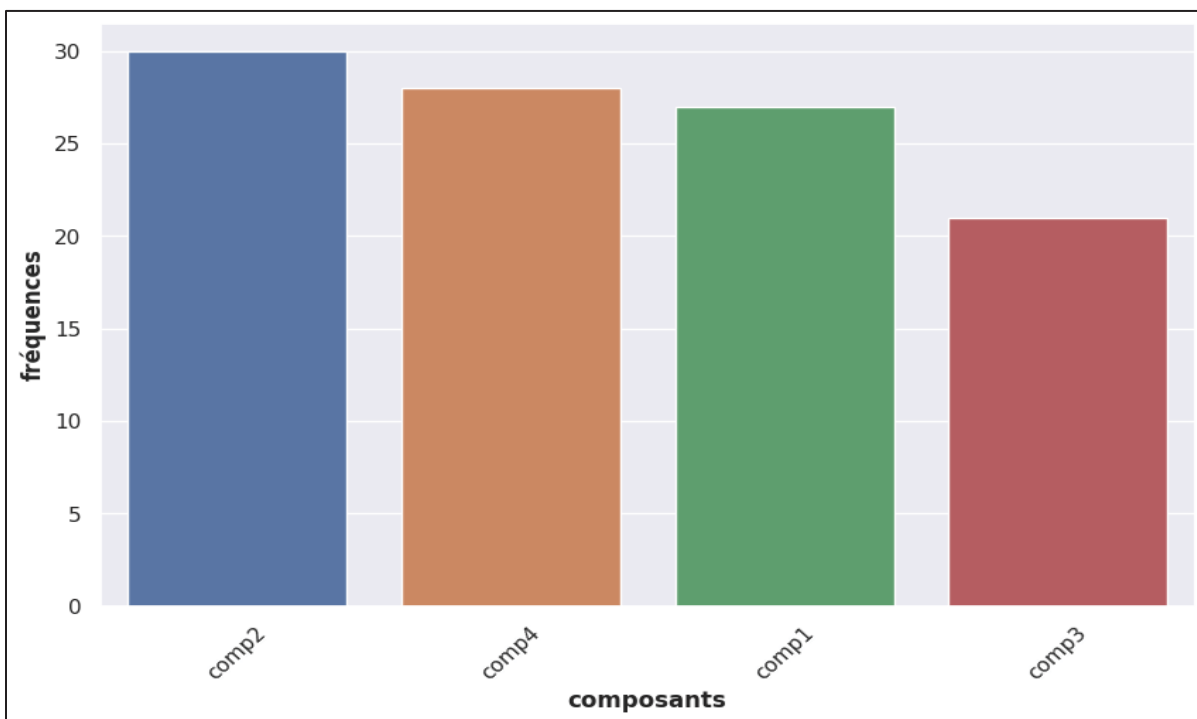


Figure 3.8 Fréquence de maintenance préventive sur les composants

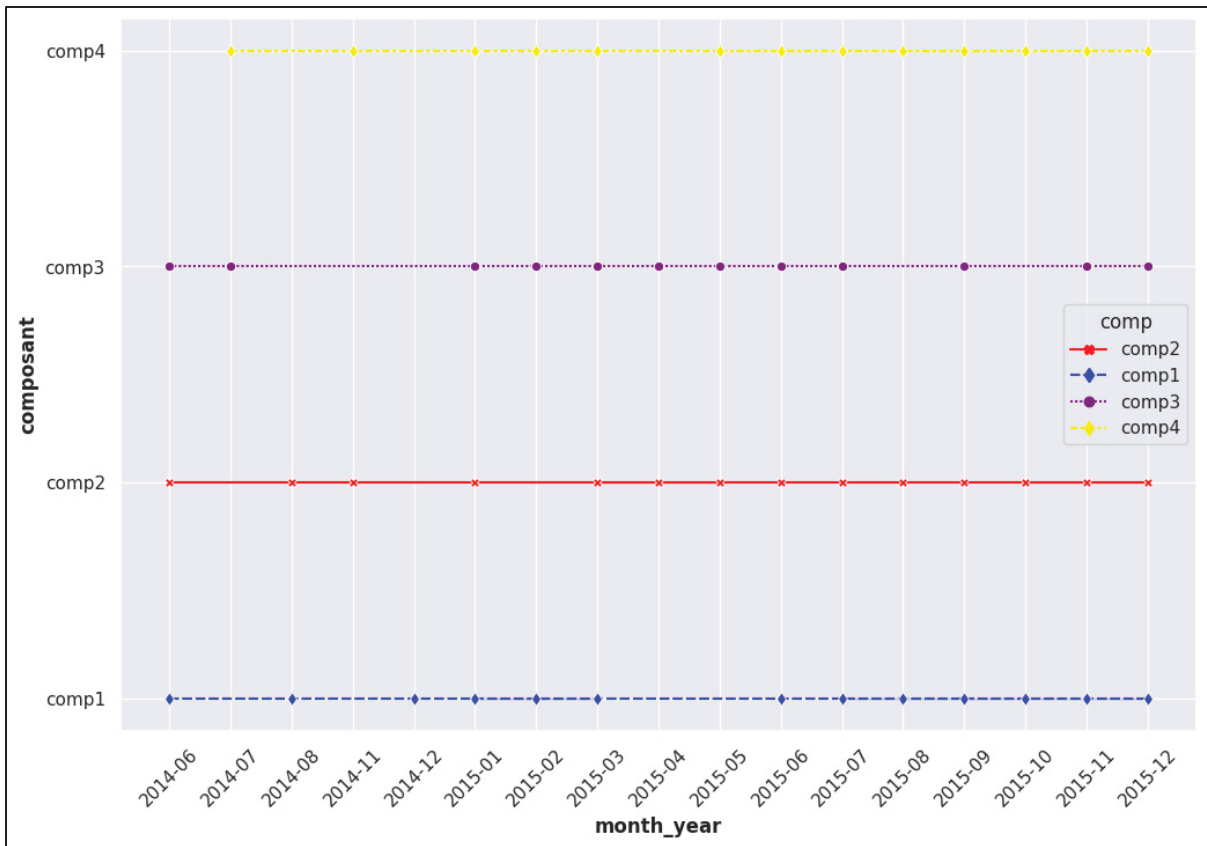


Figure 3.9 Fréquence mensuel de maintenance préventive sur les composants

3.8 Analyse des données sur les défaillances des composants

Ce jeu de données constitue l'une des parties les plus importantes des données utilisées dans notre recherche. Il contient le remplacement des composants tombés en panne. De la même façon dans le jeu de données de maintenance préventive les données de la maintenance corrective sont collectées à fréquence variable. Les composants remplacés dans le cadre d'une maintenance corrective sont les composants 1, 2 et 4.

La variable *failure* signifie le type de composant remplacé lors d'une maintenance corrective. L'analyse visuelle du tableau 3.8 permet de constater que cette variable *failure* est de type qualitative ou catégorique. Afin de confirmer notre déduction, le tableau 3.9 ci-dessous résume la nature des variables. Cependant la variable *datetime* a été transformée en type temporelle car *datetime* doit être une variable de type temporelle et non catégorique.

Tableau 3.8 Les cinq premières lignes du jeu de données de défaillance

Numéro observation	datetime	machineID	failure
0	2015-01-05 06:00:00	1	Comp4
1	2015-03-06 06:00:00	1	Comp1
2	2015-04-20 06:00:00	1	Comp2
3	2015-06-19 06:00:00	1	Comp4
4	2015-09-02 06:00:00	1	Comp4

Tableau 3.9 Les types de variables du jeu de données de défaillance

Variabes	Type de variable initiale	Type de variable finale
datetime	Object	Temporelle
machineID	Entière	Entière
failure	Object	Object

La présence de variable catégorique dans ce jeu de données ne permet pas d'appliquer les statistiques descriptives car notre variable failure n'est quantitative ou numérique. Nous utiliserons la même démarche que celle effectuée sur les jeux de données d'erreur et de maintenance préventive. Ainsi la répartition des composants remplacés lors d'une maintenance corrective sera donnée ci-dessous.

De façon plus détaillé la figure 3.10 nous montre que pour la machine 1 sur le premier trimestre de 2015 seulement deux types de composants en l'occurrence le composant 4 et 1 sont tombés en panne respectivement en janvier et mars. Au cours du deuxième trimestre tous les composants (1, 2 et 4) sont tombés en panne à des mois différents. Tandis que les deux derniers trimestres seulement les composants 2 et 4 ont été remplacé lors d'une maintenance corrective. En ce qui concerne la machine 2, nous remarquons qu'aucune maintenance corrective n'a été effectuée sur le premier trimestre de 2015. En revanche à partir du début du deuxième trimestre correspondant à mars 2015, nous observons deux maintenances correctives sur deux

composants différents à savoir le composant 1 et 2. Le reste de l'année un seul type de composant est tombé en panne en l'occurrence le composant 2. Au niveau de la machine 3 au maximum deux types de composants sont remplacés à savoir le composant 1 et 2. Pendant les trois premiers mois, les composants 1 et 2 sont traités en maintenance corrective. Au niveau du deuxième trimestre aucune maintenance corrective n'est effectuée. Cependant à partir de juin et pour chaque trimestre restant, uniquement le composant 2 est remplacé.

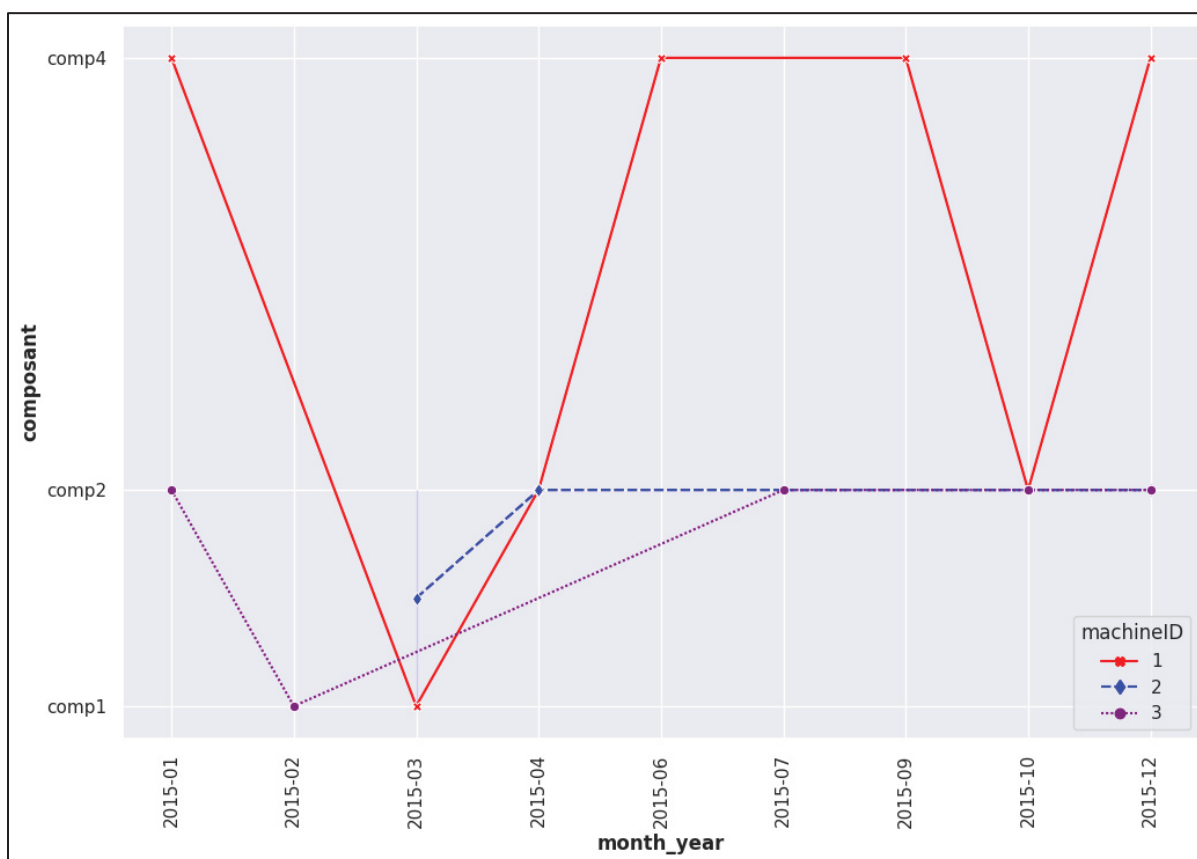


Figure 3.10 Distribution des composants remplacés lors d'une maintenance corrective sur chaque machine

3.9 Analyse des métadonnées des machines

Dans ce jeu de données, on trouve les informations concernant le type de model et l'âge des machines utilisés. Les premières lignes sont fournies par le tableau 3.10 ci-dessous et permet de constater que seule la variable model n'est pas numérique.

Tableau 3.10 Les métadonnées

Numéro observation	machineID	model	age
0	1	Model3	18
1	2	Model4	7
2	3	Model3	8

L'analyse des types de variables sera donnée ci-dessous.

Tableau 3.11 Les types de variables sur les métadonnées

Variables	Type de variable initiale	Type de variable finale
machineID	Entière	Entière
model	Object	Object
age	Entière	Entière

Nous remarquons qu'aucune variable n'est transformée car la variable model est de type catégorique tandis que l'âge et la machine sont de type quantitative ou numérique. De ce fait, la statistique descriptive ne pourra qu'être appliquée sur la variable age et machineID et non sur model. Bien que nous disposions que de trois valeurs pour ces métadonnées, nous ne jugeons pas utile d'étudier les statistiques descriptives de la variable age.

Tableau 3.12 Statistique descriptive de la variable age

	Valeur		Valeur
Nombre d'observation	3	Premier quartile	7,5
Moyenne	11	Médiane	8
Écart-type	6.082763	Troisième quartile	13
Valeur minimale	7	Valeur maximale	18

3.10 Conclusion

Ce chapitre nous a permis d'avoir une bonne compréhension de notre premier sous-problème à savoir les outils de mise en œuvre d'une stratégie de maintenance intelligente. La collecte, l'étude et l'analyse des données ont prouvé qu'elles sont des étapes incontournables dans une approche de maintenance intelligente. Pour cela, nous avons procédé en trois phases dont la première était d'indiquer les sources de données ainsi les différentes séries rencontrées en analyse de donnée. La description des outils statistiques a été traité dans la deuxième partie où nous avons montré l'impact des statistiques descriptives et enfin dans la dernière partie de ce chapitre l'ensemble des jeux de données utilisés dans nos travaux ont été détaillés en fournissant rigoureusement une méthodologie d'analyse des données. Ces données concernent la télémétrie, les erreurs, les dossiers de maintenance préventive, la défaillance des composants via la maintenance corrective et les métadonnées des machines. Afin de faciliter la compréhension, notre étude s'est limitée sur les trois premières machines, nous remarquons également que le prétraitement des données n'a été abordé dans ce chapitre car il n'y'avait pas de données manquantes. Dans le prochain chapitre ces données seront utilisées pour mettre en place une maintenance prédictive.

CHAPITRE 4

UTILITÉ DE LA MAINTENANCE INTELLIGENTE DANS L'INDUSTRIE

4.1 Introduction

Dans ce chapitre, nous allons aborder une maintenance intelligente en l'occurrence la maintenance prédictive. La mise en œuvre d'une telle stratégie de maintenance est facilitée par l'analyse des données. De ce fait l'IA par l'intermédiaire du ML nous permettra de proposer un modèle de prédiction des défaillances de composants. Ainsi nous passerons par différentes étapes pour montrer notre démarche.

Ainsi notre méthodologie sera résumée suivant trois approches qui seront :

- Une préparation des données pour faire la modélisation.
- Une présentation des résultats obtenus.
- Une prédiction de résultat.

Par ailleurs ces étapes suivies nous aideront à montrer le potentiel et l'avantage d'une maintenance intelligente de type prédictive.

4.2 Préparation des données

La préparation des données est une étape importante pour la réussite d'une démarche d'apprentissage automatisée. Cette phase de préparation rend les données mieux adapter pour la modélisation. Cependant un ensemble de sous-étapes peuvent être traités dans l'étape de la préparation des données c'est le cas du nettoyage des données, de la normalisation et de la sélection des caractéristiques. En ce qui concerne le nettoyage nous pouvons citer la gestion de valeurs manquante ou des valeurs extrêmes dans certains cas.

Par ailleurs, les données manquantes peuvent être traités de différentes façons. Dans certains cas ces valeurs sont remplacées par la moyenne des valeurs environnantes ou la valeur qui se répète le plus souvent.

Dans nos jeux de données nous n'avons pas détecter de valeurs manquantes. La normalisation est utilisée lorsque des données ne sont pas toutes à la même échelle.

Cependant la sélection des caractéristiques peut être utile dans les cas où les données présentent des variables redondantes ou non significatif.

Nous rappelons que l'ensemble de notre recherche a été réalisée par l'intermédiaire du langage Python sous Google Colab mais d'autres outils existent et pourrait être utilisé. Étant donné que notre support de base est le ML, nous la présentons ci-dessous.

4.2.1 Apprentissage automatique

L'apprentissage automatique ou ML est divisé en deux grands groupes que sont l'apprentissage supervisé et l'apprentissage non supervisé. L'apprentissage supervisé utilise des données identifiées c'est-à-dire qu'il va faire des prédictions sur des données dont les variables sont étiquetées. Ainsi des données sont utilisées en entrée du modèle et le résultat attendu sera connu (Jemmali, 2021). L'apprentissage supervisé s'utilise dans deux situations que sont la classification et la régression. En ce qui concerne la classification, elle est rencontrée dans des situations où la variable à prédire est de type qualitatif donc catégorielle tandis que la régression est utilisée pour faire des prédictions sur une variable quantitative donc numérique. En revanche l'apprentissage non supervisée utilise des données non étiquetées donc elle permet de chercher ou de trouver des motifs cachés dans les données sans étiquette. Nous dirons également que le but d'un algorithme d'apprentissage non supervisé est de trouver lui-même des structures (ou relations) sous-jacentes dans les données non étiquetées.

L'apprentissage automatique non supervisée utilise des techniques dites de segmentation ou clustering dont son rôle est de regrouper les observations c'est-à-dire que le modèle va identifier les caractéristiques similaires afin de séparer les objets. Une autre technique fréquemment utilisée est la réduction de dimension. Dans notre cas, notre objectif est de prédire la défaillance d'un composant or nous savons que les défaillances sont représentées par la variable failure qui est une variable qualitative donc notre problème sera basé sur un modèle d'apprentissage automatique supervisé de type classification.

Nous retrouverons également l'apprentissage automatique sous le nom d'apprentissage machine. Ainsi la figure ci-dessous résume le principe d'un apprentissage automatique.

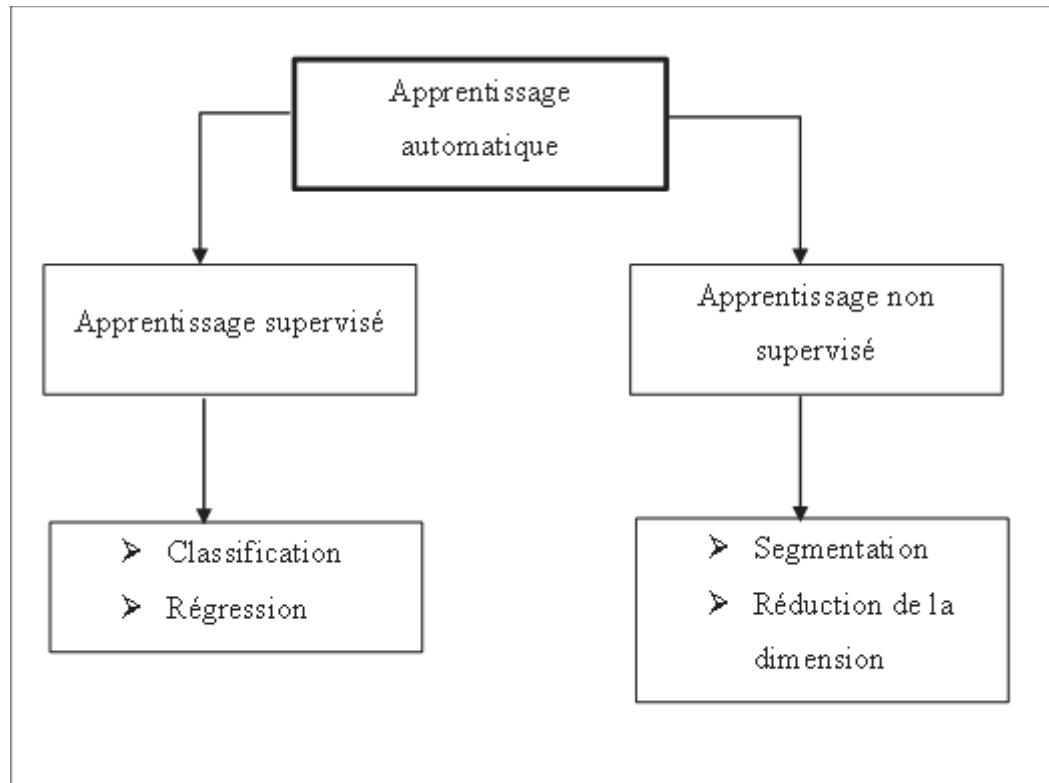


Figure 4.1 Variante d'un apprentissage automatique

4.2.2 Regroupement des données

L'utilisation de modèle ou d'algorithme d'apprentissage automatique nécessite des données quantitatives pour effectuer une prédiction. Cependant la plupart des jeux de données contiennent à la fois des variables qualitative et quantitative. Pour cela, une technique est utilisée pour convertir les variables qualitative ou catégorique en variable numérique.

Cette technique se nomme encodage. Elle peut être appliqué de différentes façon. Nous pouvons citer par exemple l'encodage par *one-hot* qui utilise un vecteur binaire représentant la catégorie. Ce vecteur sera égal à 1 si la catégorie est présente et 0 sinon. Dans la suite de notre mémoire ce type d'encodage sera utilisé.

Après avoir analyser les jeux de données dans le chapitre précédent nous les assemblerons afin d'obtenir un jeu de données global. Pour ce faire nous avons chargé dans notre environnement de travail plusieurs bibliothèque ou librairie afin de manipuler ces jeux de données. Nous pouvons citer la bibliothèque pandas, numpy, seaborn et matplotlib.

Nous avons commencé par le formatage de la variable datetime pour chaque jeu de donnée car cette variable n'était pas sous format temporel. Ensuite les variables de type object ont été convertis en type catégorique.

En ce qui concerne le jeu de données de télémétrie les variables volt, rotate, pressure et vibration ne seront pas transformées car elles sont numériques. En revanche la variable catégorique errorID du jeu de données des erreurs a été encodé par la façon *one-hot encoder*. Ainsi le tableau ci-dessous nous montre d'une part l'avantage de l'encodage et d'autre part nous montre un type d'erreur pour chaque machine à chaque heure.

Tableau 4.1 Encodage des erreurs

Numéro observation	machineID	datetime	error1	error2	error3	error4	error5
0	1	2015-01-03 07 :00 :00	1	0	0	0	0
1	1	2015-01-03 20 :00 :00	0	0	1	0	0
2	1	2015-01-04 06 :00 :00	0	0	0	0	1
3	1	2015-01-10 10 :00 :00	0	0	0	1	0

Étant donné que nous cherchons à assembler les jeux de données donc notre stratégie est d'ajouter les variable datetime et machineID du jeu de données de télémétrie. Ainsi tous les points horaires différent de ceux du tableau 4.1 auront des entrées nulles 0.0. Le tableau ci-dessous permettra de mieux comprendre l'ajout des variables datetime et machineID.

Tableau 4.2 Encodage des erreurs et ajout de variables télémétriques

Numéro observation	machineID	datetime	error1	error2	error3	error4	error5
0	1	2015-01-01 06 :00 :00	0.0	0.0	0.0	0.0	0.0
1	1	2015-01-01 07 :00 :00	0.0	0.0	0.0	0.0	0.0
2	1	2015-01-01 08 :00 :00	0.0	0.0	0.0	0.0	0.0
3	1	2015-01-01 09 :00 :00	0.0	0.0	0.0	0.0	0.0

Au niveau du jeu de données de maintenance la variable comp sera encodé de façon particulière. Nous calculerons ainsi pour chaque composant, le temps écoulé depuis la dernière maintenance. Le tableau ci-dessous permet de constater par exemple au 1^{er} janvier 2015 que 19 jours se sont écoulés depuis le dernier remplacement du composant 1 tandis que 214 jours se sont écoulés depuis le remplacement du composant 2.

Tableau 4.3 Encodage de la variable comp du Jeu de données des maintenances

Numéro observation	machineID	datetime	comp1	comp2	comp3	comp4
0	1	2015-01-01 06 :00 :00	19	214	154	169
1	1	2015-01-01 07 :00 :00	19,04	214,04	154,04	169,04
2	1	2015-01-01 08 :00 :00	19,08	214,08	154,08	169,08
3	1	2015-01-01 09 :00 :00	19,13	214,13	154,13	169,13

Au niveau du jeu de données de machine, nous n'encodons pas la variable model car sur les trois machines nous n'avons que trois observations contrairement aux autres jeux de données. De ce fait, les fonctionnalités ou variables de ce jeu de données seront conservés sans modification.

À présent nous assemblerons les jeux de données de télémétrie, des erreurs, et des maintenances. Nous commençons par la fusion des jeux de données de télémétrie et des erreurs en utilisant la méthode *left join*, les clés de fusion sont les variables datetime et machineID ainsi toutes les lignes de télémétrie seront maintenus tandis que les valeurs des erreurs seront ajoutés si et seulement si les valeurs des clés de fusion sont les mêmes. Ainsi la fusion de télémétrie avec les erreurs sera fusionnée à son tour avec le jeu de donnée de maintenance en utilisant *left join*. Cette fusion ainsi obtenue sera assemblé avec le jeu de données de machine en utilisant toujours la méthode left join pour cela la clé de fusion sera la variable machineID.

Tableau 4.4 Fusion de télémétrie, erreurs, maintenance et machine

Numéro d'observation	datetime	machineID	Volt	rotate	pressure	vibration	
0	2015-01-01 06:00	1	176,22	418,50	113,08	45,09	
1	2015-01-01 07:00	1	162,88	402,75	95,46	43,41	
Numéro observation	datetime	machineID	error1	error2	error3	error4	error5
0	2015-01-01 06:00	1	0.0	0.0	0.0	0.0	0.0
1	2015-01-01 07:00	1	0.0	0.0	0.0	0.0	0.0
Numéro observation	datetime	machineID	comp1	comp2	comp3	comp4	
0	2015-01-01 06:00	1	19	214	154	169	
1	2015-01-01 07:00	1	19,04	214,04	154,04	169,04	
Numéro observation	model	age					
0	model3	18					
1	model3	18					

Le tableau 4.4 illustre le principe de la fusion des données donc dans le tableau fusionné les variables sont ordonnées de la façon suivante datetime, machineID, volt, rotate, pressure, vibration, error1, error2, error3, error4, error5, comp1, comp2, comp3, comp4, model et age. Cependant dans notre cas la méthode left join a été utilisé mais d'autres type de fusion existent, traitant des spécificités variées(OpenClassrooms, 2023).

Enfin nous fusionnons tous les ensembles de données avec le jeu de donnée des pannes par la méthode de fusion *left-join*, ainsi les valeurs manquantes seront remplies par la méthode *bfill* c'est à dire un remplissage par les sept lignes suivantes. En revanche dans la variable failure les valeurs manquantes seront remplacées par none qui signifie aucune panne. Ainsi le tableau suivant permet de mieux visualiser la fusion totale de tous les ensembles de données.

Tableau 4.5 Fusion de tous jeux de données

Numéro d'observation	datetime	machineID	Volt	rotate	pressure	vibration	
0	2015-01-01 06:00	1	176,22	418,50	113,08	45,09	
1	2015-01-01 07:00	1	162,88	402,75	95,46	43,41	
Numéro observation	datetime	machineID	error1	error2	error3	error4	error5
0	2015-01-01 06:00	1	0.0	0.0	0.0	0.0	0.0
1	2015-01-01 07:00	1	0.0	0.0	0.0	0.0	0.0
Numéro observation	datetime	machineID	comp1	comp2	comp3	comp4	
0	2015-01-01 06:00	1	19	214	154	169	
1	2015-01-01 07:00	1	19,04	214,04	154,04	169,04	
Numéro observation	model	age	failure				
0	model3	18	none				
1	model3	18	none				

Le tableau 4.5 illustre le principe de la fusion des données donc dans le tableau fusionné les variables seront ordonnées de la façon suivante datetime, machineID, volt, rotate, pressure,

vibration, error1, error2, error3, error4, error5, comp1, comp2, comp3, comp4, model, age et failure. Ainsi nous obtenons 18 variables dans ce jeu de données finale.

Afin de procéder à la modélisation, nous devons vérifier les relations entre les variables. Une relation de corrélation entre les variables permettra de mieux faire des prédictions. Ainsi la figure ci-dessous donne la corrélation entre les variables numérique.

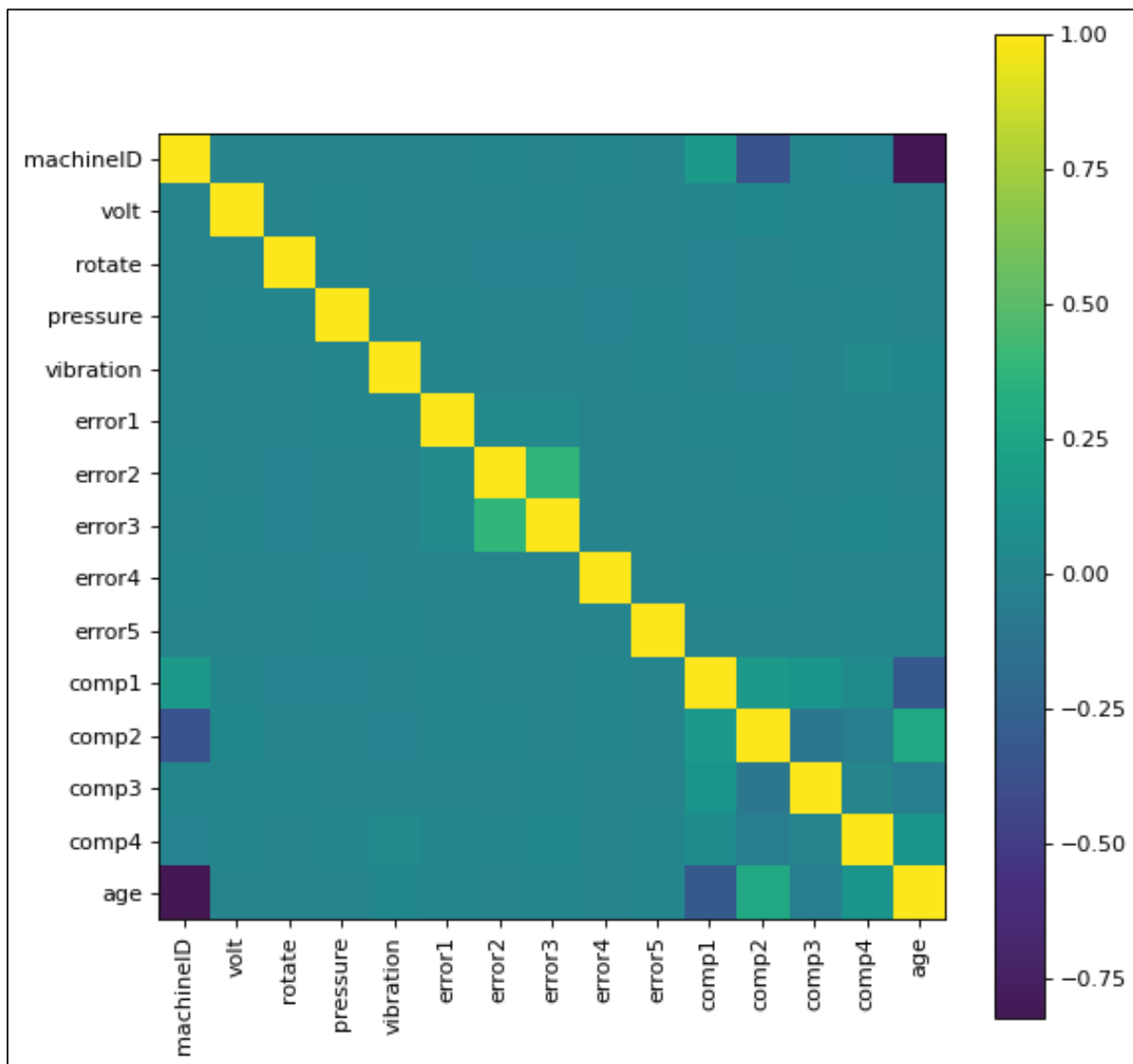


Figure 4.2 Matrice de corrélation des variables

L'analyse de la figure 4.2 nous renseigne sur le niveau de corrélation des variables. Cependant nous notons que très peu de variables sont corrélés entre elles car plus la couleur est proche du jaune et plus la corrélation est forte. De ce fait nous appliquons l'ingénierie des caractéristiques afin de vérifier si la matrice de corrélation sera mieux.

4.2.3 Ingénierie des caractéristiques

L'ingénierie des caractéristiques fait référence à la création de nouvelles caractéristiques ou variables à partir des données brutes. Ainsi le but de l'ingénierie des caractéristiques est de trouver de nouvelles caractéristiques qui pourront améliorer les performance d'un modèle d'apprentissage automatique (Data Science Team, 2021).

Dans notre cas, nous utiliserons des caractéristiques de décalages comme technique d'ingénierie de caractéristique. Une caractéristique de décalage peut être défini comme un moyen de transformation de problème de série chronologique en problème d'apprentissage automatique.

De ce fait nous l'appliquons sur le jeu de données de télémétrie afin de représenter l'historique à court terme de ces données. Pour cela une fenêtre de 3h sera définie et pour chaque fenêtre nous calculerons des mesures globales glissantes en utilisant la fonction de rééchantillonnage disponible dans la bibliothèque de pandas, ces mesures sont :

- Le minimum
- Le maximum
- La moyenne
- L'écart type

Ces mesures calculées deviennent les caractéristiques de décalage qui seront fusionnés avec les autres jeux de données. Ainsi nous pourrons vérifier la corrélation entre les variables et commencer notre construction du modèle.

La caractéristique de décalage sur le minimum est donnée par le tableau ci-dessous et permet de constater que les points horaires sont remplacés par des fenêtres de 3h. Ainsi nous obtenons un jeu de donnée contenant que des valeurs minimales pris sur un intervalle de 3h.

Tableau 4.6 Caractéristique de décalage minimum pour fenêtre de 3h

machine	datetime	volt_min_ 3h	rotate_min_ _3h	pressure_min_ _3h	vibration_min_ _3h
1	2015-01-01 09:00	162,879223	402,74749	75,2379049	34,1788471
1	2015-01-01 12:00	157,610021	346,149335	95,9270417	25,990511
1	2015-01-01 15:00	156,556031	398,648781	101,001083	35,4820087
1	2015-01-01 18:00	160,263954	382,483543	96,4809757	38,5436809
1	2015-01-01 21:00	153,353492	402,461187	86,0124403	39,7398827

De la même façon sur la fenêtre de 3h, les caractéristiques de décalages des maximums sont créées. Une valeur est obtenue dans un intervalle ou fenêtre de 3h en prenant l'intervalle fermé à gauche par exemple la valeur maximale à 9h sur la fenêtre de 3h est donnée par l'intervalle [06 :00 ; 09 :00[.

Ainsi dans cette intervalle la valeur maximale pour la tension (le volt) est de 176,217853 donc la valeur à 9h sera de 176,217853 ainsi de suite pour toutes les variables et pour tous les intervalles.

Le tableau ci-dessous résume les caractéristiques de décalage des maximums.

Tableau 4.7 Caractéristique de décalage maximum pour fenêtre de 3h

machine	datetime	volt_max_ 3h	rotate_max_ 3h	pressure_max_ 3h	vibration_max_ 3h
1	2015-01-01 09:00	176,217853	527,349826	113,0779355	45,08768576
1	2015-01-01 12:00	172,504839	435,376873	111,8866482	41,12214409
1	2015-01-01 15:00	175,324524	499,071623	111,7556843	45,48228685
1	2015-01-01 18:00	169,218423	460,85067	104,84823	42,67579991
1	2015-01-01 21:00	182,739113	490,672921	93,48495406	44,10855437

Le même principe expliqué plus haut est appliqué pour trouver les caractéristiques de décalage des moyennes et des écarts type et peut être visualisé par les tableaux ci-dessous.

Tableau 4.8 Caractéristique de décalage des moyennes pour fenêtre de 3h

datetime	Volt_mean_3h	Rotate_mean_3h	Pressure_mean_3h	vibration_mean_3h
2015-01-01 09:00	170,028993	449,533798	94,5921219	40,8935019
2015-01-01 12:00	164,192565	403,949857	105,687417	34,2558908
2015-01-01 15:00	168,134445	435,781707	107,793709	41,2394051
2015-01-01 18:00	165,514453	430,472823	101,703289	40,3737388
2015-01-01 21:00	168,809348	437,11112	90,9110605	41,7385423

Tableau 4.9 Caractéristique de décalage des écart-type pour fenêtre de 3h

datetime	volt_sd_3h	rotate_sd_3h	pressure_sd_3h	vibration_sd_3h
2015-01-01 09:00	6,7210322	67,8495995	18,9349564	5,87496979
2015-01-01 12:00	7,59657019	50,1204521	8,55503192	7,66222854
2015-01-01 15:00	10,1245841	55,0847339	5,90972113	5,16930429
2015-01-01 18:00	4,673269	42,0472779	4,55404715	2,1061083
2015-01-01 21:00	14,7521321	47,0486087	4,24415846	2,20788363

Sur les deux tableaux nous n'avons pas affiché la variable machine ID pour éviter d'encombrer les tableaux. Ainsi toutes les caractéristiques de décalage sur la fenêtre de 3h constituent l'historique à court terme. Cependant l'historique à moyen terme pourrait entraîner une meilleure corrélation entre les variables. De ce fait, nous calculons les caractéristiques de

décalage pour une fenêtre de 24h. Par ailleurs, ces caractéristiques seront les minimums, les maximums, les moyennes et les écarts-types.

Le tableau ci-dessous montre les caractéristiques de décalage des minimums.

Tableau 4.10 Caractéristique de décalage minimum pour fenêtre de 24h

datetime	volt_min_24h	rotate_min_24h	pressure_min_24h	vibration_min_24h
2015-01-02 06:00	151,335682	346,149335	75,2379049	25,990511
2015-01-02 09:00	151,335682	346,149335	75,2379049	25,990511
2015-01-02 12:00	147,300678	382,483543	78,8807796	25,990511
2015-01-02 15:00	147,300678	382,483543	78,8807796	29,5276645
2015-01-02 18:00	147,300678	382,483543	78,8807796	29,5276645

La particularité avec l'historique à moyen terme est d'utiliser une fenêtre de 24h puis effectuer un rééchantillonnage tous les 3h. Par exemple dans le tableau 4.10 les valeurs ont été obtenues par la manière suivante : Pour chaque point horaire la valeur minimale sur les 24 dernières heures est calculée ensuite nous appliquons un rééchantillonnage sur une fenêtre de 3h pour connaître la valeur minimale de la caractéristique de décalage sur la fenêtre de 24h.

Autrement dit le 02 janvier à 06h:00 la valeur minimale dans les dernières 24h est de 151,33, celui de 07h:00 à 151.33 et celui de 08h :00 à 151.33 donc si nous faisons un rééchantillonnage sur une fenêtre de 3h nous aurons l'intervalle [06 :00; 09 :00[, alors la valeur minimale dans cet intervalle sera de 151,33 d'où la valeur observée sur le tableau.

La méthode est appliquée pour obtenir les caractéristiques de décalage des maximums, des moyennes et des écarts-type.

Enfin l'ensemble des jeux de données de ces caractéristiques sera fusionné avec les erreurs, les maintenances, les métadonnées des machines et les pannes. Ainsi nous obtenons un jeu de données final constitué de 46 variables. Afin de vérifier la corrélation entre les variables nous traçons la matrice dans la figure ci-dessous.

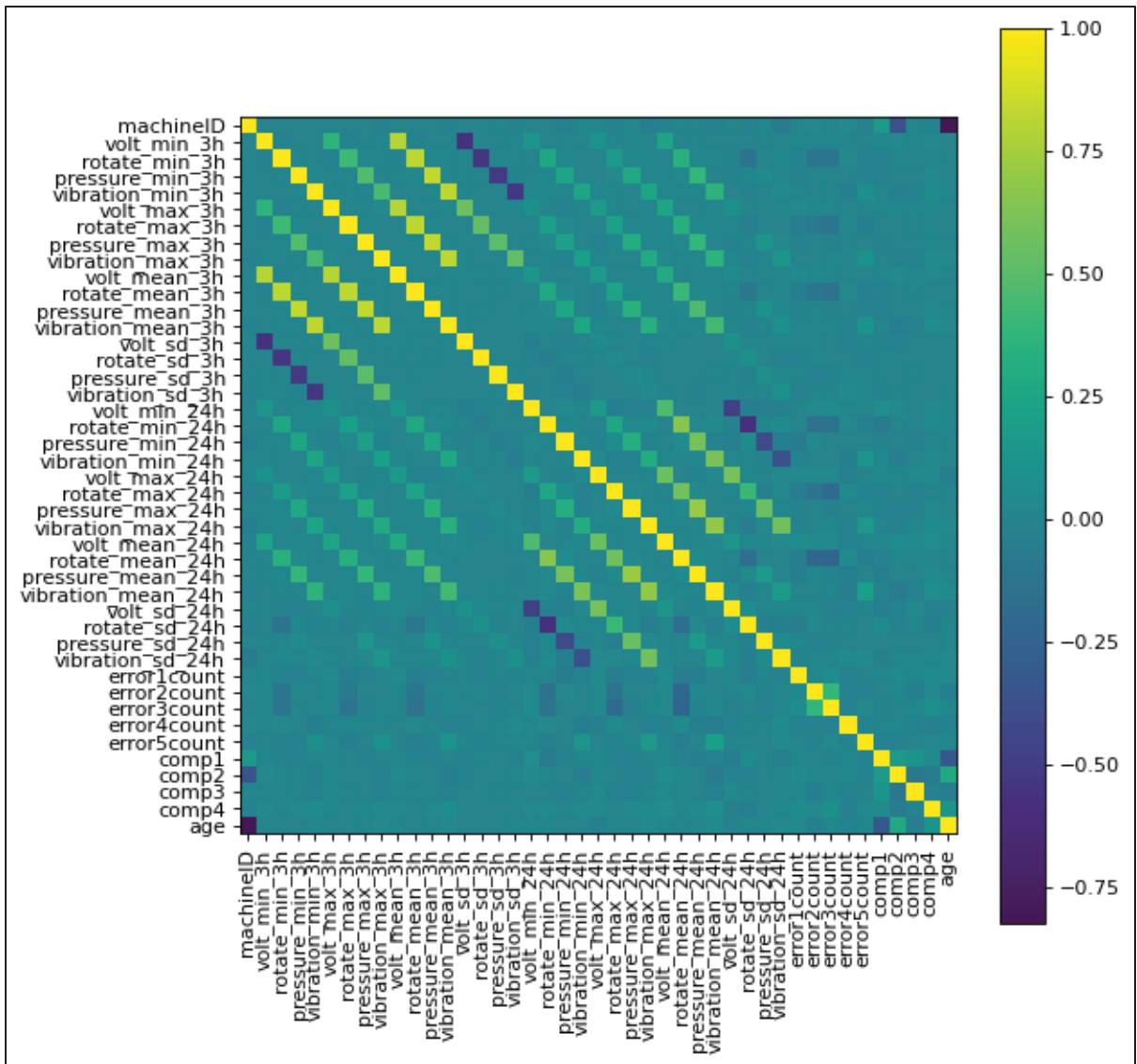


Figure 4.3 Matrice de corrélation après ingénierie des caractéristiques

La matrice de corrélation obtenue après l'étape de l'ingénierie des caractéristiques a montré que le nombre de variable corrélés a augmenté car nous observons que la couleur jaune est plus présente.

Ainsi nous notons que certaines variables sont corrélées positivement entre elles et d'autres négativement.

Parmi les variables corrélées positivement nous pouvons citer par exemple les couples de variables suivantes :

- (volt_mean_3h, volt_min_3h)
- (vibration_mean_3h, vibration_min_3h)
- (pressure_mean_3h, pressure_min_3h)

Les variables ci-dessus sont corrélées positivement entre elles car plus la couleur est proche du jaune et plus les variables sont corrélées.

Ainsi la corrélation positive permet de dire que ces couples de variables peuvent varier ensemble de manière cohérente cela va dire si l'une augmente l'autre aura tendance à augmenter. Cependant une corrélation n'implique pas forcément une relation de cause à effet.

En ce qui concerne les variables corrélées négativement nous pouvons nommer par exemple les couples suivants :

- (age, machineID)
- (comp2« nombre de jours écoulé depuis la dernière maintenance du comp2 », machineID)
- (volt_sd_3h, volt_min_3h)

Ces variables sont considérées comme corrélées négativement car la couleur est similaire au bleu foncé représentant un coefficient de corrélation proche de -1.

Ainsi la corrélation négative signifie ces couples de variable auront tendance à évoluer de manière inverse c'est-à-dire si l'une augmente l'autre aura tendance à diminuer. Par exemple dans notre cas, il y a trois machines étudiées nous pouvons observés que plus le numéro de machine augmente et plus l'âge à tendance à diminuer. De la même façon la relation de cause à effet n'est pas prouvée car l'âge de la machine 2 est inférieur à celle de la machine 3. S'il y'a une causalité l'âge de la machine 2 serait supérieur à celle de la machine 3.

Ainsi nous déterminons les caractéristiques retenues pour la modélisation dans la partie suivante.

4.3 Sélection des caractéristiques

Dans cette partie, nous exploitons les résultats obtenus après l'ingénierie des caractéristiques. Le jeu de données final obtenue présente des variables corrélées entre elles donc nous dirons que ce jeu de données final sera utilisé pour mettre en place un modèle prédictif. Étant donné que nous cherchons à savoir la défaillance d'un composant dans un avenir proche pour cela notre jeu de données final sera divisé en trois sous-ensembles qui seront traités ci-dessous.

4.3.1 Sous-ensemble de variable dépendante

La variable dépendante encore appelée variable cible dans notre cas est la variable qui est prédite c'est-à-dire le modèle construit va déterminer la défaillance de composant 1,2 et 4, car sur l'ensemble des trois machines les composants défectueux ou remplacés lors d'une maintenance corrective sont les composants 1,2 et 4. Ainsi notre variable dépendante est la variable failure.

4.3.2 Sous-ensemble de variable indépendantes

Les variables indépendantes ou explicatives sont les variables utilisées pour la construction du modèle. Ces variables sont :

- Toutes les caractéristiques de décalage dans la fenêtre de 3h
- Toutes les caractéristiques de décalage calculée dans un horizon moyen terme (24h)
- Toutes les caractéristiques ou variables des erreurs encodées
- Toutes les caractéristiques ou variables de maintenance encodée
- L'âge et le model des machines

4.3.3 Sous-ensemble de variable non informative

Nous considérons ces variables comme étant non importants pour la construction de notre modèle. Ces variables non informatives sont machineID et datetime.

Avec ces trois sous-ensembles de variable nous pourrons construire le modèle dans la partie suivante.

4.4 Construction du modèle

Le modèle devra prédire si le composant 1, 2 ou 4 tombe en panne dans les prochaines 24h. Pour cela nous divisons nos variables indépendantes en deux parties que sont :

- Variable quantitative numérique
- Variable qualitative ou catégorielle

La variable qualitative est la variable « model » et le reste des variables indépendantes sont les variables numériques. À partir de cette étape, nous utilisons un Column transformer qui va effectuer deux transformations sur les variables dont une transformation numérique appliqués sur les variables numérique et une transformation catégorielle appliquée sur la variable catégorielle.

Un transformer quantitative va effectuer la transformation numérique en faisant une standardisation des variables numérique alors qu'un transformer qualitative va effectuer une transformation catégorielle en faisant un encodage *one-hot encoder* sur la variable qualitative. L'ensemble de ces transformations seront mis dans un pipeline pour entrainer le modèle.

À titre d'information, nous définissons un pipeline comme un outil ou objet qui regroupe l'ensemble des étapes de préparation initiale jusqu'à l'obtention du modèle de prédiction finale. Ainsi notre pipeline effectuera deux taches successives que sont un prétraitement de données et une classification de défaillance en utilisant les classifieurs ou algorithmes suivants :

- Une Régression logistique ou LR (*Linear Regression*)
- Une forêt aléatoire ou RF (*Random Forest*)
- Un AdaBoost (*Adaptive Boosting*)
- Un algorithme des plus proches voisins ou KNN (*K-Nearest Neighbors*)

La création du modèle peut être résumé dans la figure ci-dessous.

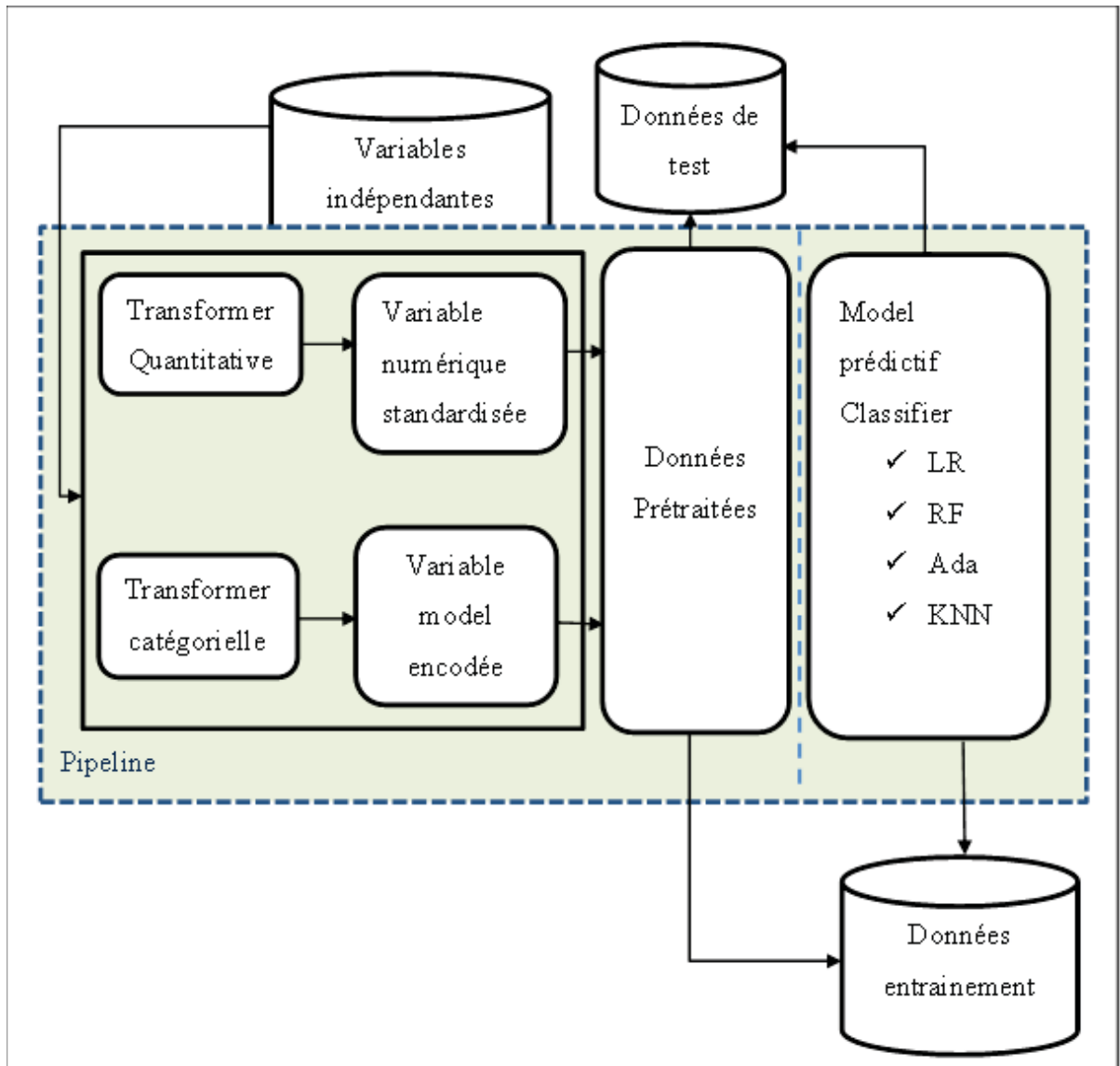


Figure 4.4 Schéma du modèle

Le schéma du modèle montre l'utilisation d'un pipeline pour la mise en place d'un modèle prédictif. Après le prétraitement les données sont divisées en donnée d'entraînement et donnée de test. Nous utilisons les neufs premiers mois de l'année 2015 comme données d'entraînements qui serviront à former le modèle et les trois derniers mois restants seront les données de test ou le modèle sera évalué.

Cependant les variables numériques sont standardisées ce qui signifie que ces variables seront mises à l'échelle on l'appelle également *feature scaling* ou même la normalisation. Le rôle de la standardisation est de transformer les données de sorte qu'elles auront une moyenne de zéro et un écart-type de 1 ainsi la normalisation pourrait réduire l'écart entre les variables d'une part et d'autre part de mettre toutes les variables presque à la même échelle.

La standardisation d'une variable numérique X est donnée par la formule suivante :

$$X_{standardisé} = \frac{(X - \mu)}{\sigma} \quad (4.1)$$

Avec

μ = Moyenne de X

σ = Écart-type de X

L'avantage du pipeline est de regrouper plusieurs algorithmes dans le classifieur ce qui facilite l'entraînement du modèle. Les algorithmes contenus dans le classifieur seront étudié ci-dessous.

4.4.1 La régression logistique

La régression logistique est souvent utilisée en classification malgré son nom. Elle se définit comme un algorithme d'apprentissage supervisé qui peut effectuer une classification binaire (une classe) ou une multi classe. Son rôle est de trouver la relation qui relie un ensemble de variables indépendantes et une variable cible. Ainsi la probabilité pour qu'un échantillon de l'ensemble d'entraînement appartient à une classe bien précise est évalué par l'algorithme. De ce fait, une classification sera établie sous la base d'un seuil de probabilité prédéfini. Il existe une multitude de techniques employés par la régression logistique lorsqu'elle est utilisée pour prédire une classe alors nous parlons de binaire et si pour plusieurs classes nous parlons de multinomial. Ces différentes techniques sont expliqués en détail dans le site de Jedha (Krajnc, 2023).

Il convient de noter qu'une régression linéaire est différente de la régression logistique en raison de la nature de la variable dépendante. Nous pouvons utiliser la régression linéaire dans le cas où la variable à prédire ou dépendante est de type continue par exemple une variable

dépendante pourrait être le RUL sur un composant. En revanche la régression logistique peut être appliquée dans le cas où la variable dépendante est de nature catégorielle (Voxco, 2023a). La figure ci-dessous montre le principe d'une régression linéaire et d'une régression logistique.

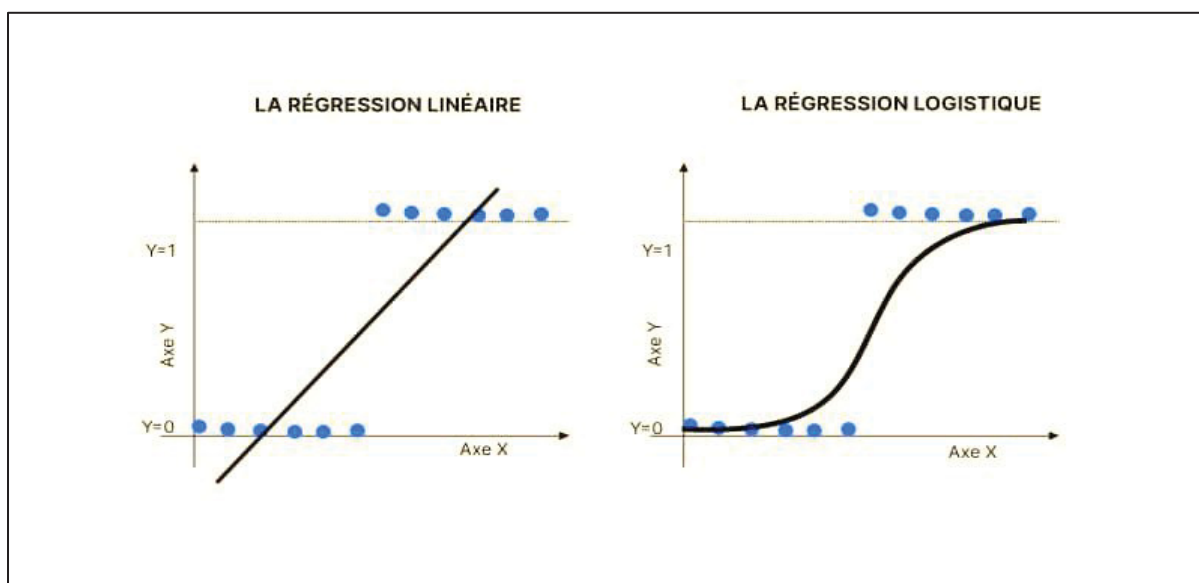


Figure 4.5 Différence entre régression linéaire et logistique tirée de Voxco (2023)

Cependant le principe général d'une régression logistique est matérialisé ci-dessous.

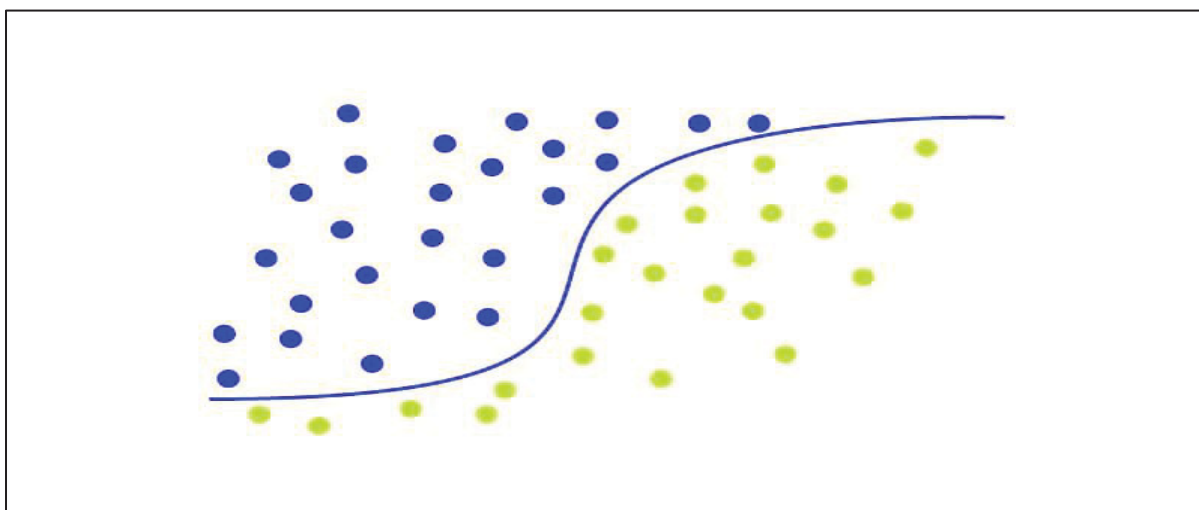


Figure 4.6 Principe d'une régression logistique adaptée de Voxco (2023)

4.4.2 Forêt aléatoire

La forêt aléatoire ou RF est un algorithme qui s'utilise aussi bien en classification qu'en régression donc elle est utilisée en apprentissage automatique supervisée. Son fonctionnement est relativement simple, et est basé sur des décisions prédictives fournies par un ensemble d'arbres de décision individuel.

La mise en œuvre de l'algorithme de RF nécessite des données d'entraînement de bonne qualité et prétraitées. Après cette étape l'algorithme utilise le *bootstrap* qui est une méthode servant à créer plusieurs ensembles de données d'entraînement issues des données d'entraînement de base.

Ces ensembles sont créés de façon aléatoire ce qui fait qu'un ensemble peut se retrouver plusieurs fois dans un échantillon *bootstrap* ou pas du tout. Ainsi un arbre de décision est créé sur chaque échantillon bootstrap. Cependant un sous-ensemble de variables indépendantes sera utilisé lors de la création d'un arbre de décision.

Lorsque tous les arbres sont créés les prédictions de chaque arbre seront combinées afin de fournir une prédiction finale. Ainsi dans notre cas puisqu'il s'agit d'une classification, un vote majoritaire sera fait qui signifie qu'une classe prédite est celle qui a reçu plus de vote.

En revanche dans le cas d'une régression la prédiction sera faite à partir de la moyenne de prédictions des arbres. Par ailleurs, dans certains travaux nous trouvons que l'algorithme de forêt aléatoire est défini comme une démarche de diviser pour mieux régner (Jemmali, 2021).

L'un des avantages d'un RF est qu'il peut être utilisé sur des valeurs aberrantes car la prédiction obtenue se base sur un ensemble d'arbres plutôt que sur un seul arbre. Également le RF ne fait quasiment pas de surapprentissage par rapport à un seul arbre car chaque arbre est un sous-ensemble de variable indépendante (Tibco, s.d.).

La figure ci-dessous montre le principe de vote majoritaire associée à l'algorithme de RF

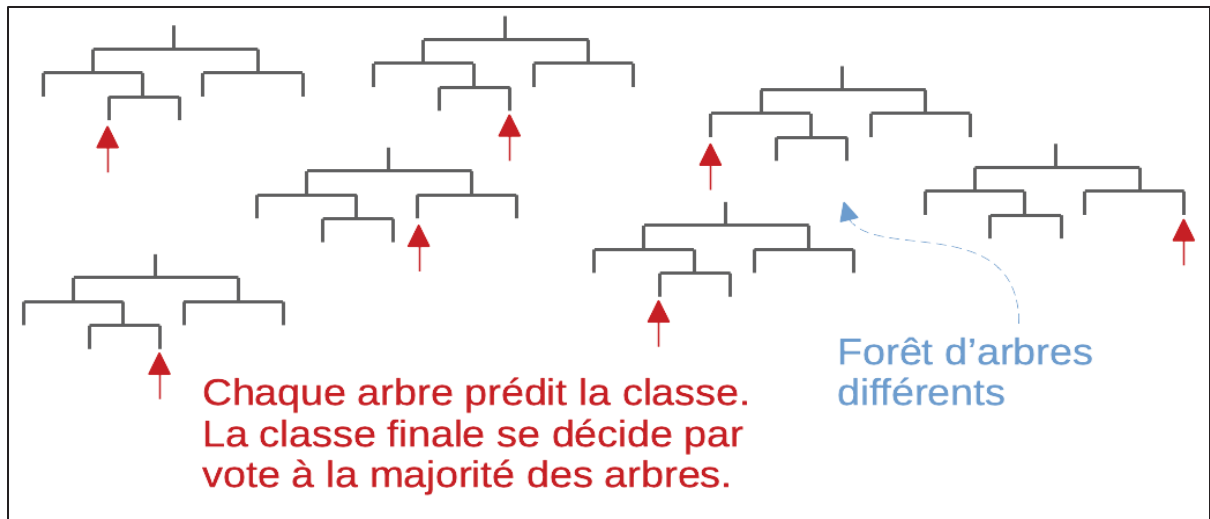


Figure 4.7 Architecture d'une forêt aléatoire tirée de Engels (s.d)

4.4.3 AdaBoost

L'algorithme AdaBoost plus connu sous le nom de Adaptive Boosting est considéré comme un modèle ensembliste car il associe des modèles d'apprentissage faible pour avoir un modèle fort regroupant les qualités de chacun de ces modèles faibles. Contrairement au RF qui utilise une approche de « diviser pour mieux régner », l'AdaBoost quant à lui propose une approche plutôt basée sur « l'union fait la force » (kwamimayeden, 2022b).

La mise en place d'un algorithme d'adaBoost nécessite cinq étapes qui peuvent être décrite ci-dessous.

À l'étape 1, un poids identique w_i est affecté à chaque individu (i) ou observation (i), où la somme des poids individuels est égale à 1.

$$\sum_{1}^{n} w_i = 1 \quad (4.2)$$

Où,

n = nombre totale d'individu ou d'observation.

Étant donné que les poids sont identiques donc w_i est donné par la formule suivante :

$$w_i = \frac{1}{n} \quad (4.3)$$

À l'étape 2, un premier arbre de décision est créé. Cet arbre va fournir des prédictions soit bonnes, soit mauvaises ou soit bonne et mauvaise. Dans le cas d'une bonne classification, l'arbre fournit des prédictions bonnes sur certains individus. En revanche, pour une mauvaise classification les prédictions seront fausses. Si (j) désigne un individu mal classé et (k) un individu bien classé alors l'algorithme calcule l'erreur totale ε_t du premier arbre en sommant le poids des individus mal classés. Si (m) est le nombre total d'individu mal classé et (o) le nombre total d'individu bien classé, alors l'erreur total sur le premier arbre est égale à la somme des poids des individus mal classés ainsi elle est donnée par la formule suivante :

$$\varepsilon_t = \sum_1^m w_j \quad (4.4)$$

Après le calcul de l'erreur total, l'algorithme va calculer la performance (P) de ce premier arbre de décision par la formule suivante :

$$P = \frac{1}{2} * \log \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (4.5)$$

Nous remarquons d'après la formule que plus l'erreur total augmente plus la performance (P) diminue et la prédiction faite par cet arbre sera mauvaise (classe mal prédite).

Au niveau de l'étape 3, l'algorithme effectue une mise à jour des poids de chaque observation ou individu. Ainsi des arbres seront créés successivement de telle sorte qu'un nouvel arbre qui se crée corrigera les erreurs de l'arbre précédent. Concrètement cela signifie que pour corriger les erreurs il faut augmenter le poids des individus mal classés ainsi le nouvel arbre se focalisera en priorité sur les erreurs de l'arbre précédent. Pour le nouvel arbre il va recalculer les poids des individus mal classés (j) dans l'arbre précédent en utilisant la formule suivante :

$$w_{j-actuel} = w_j * e^{(P)} \quad (4.6)$$

Étant donné que le poids des individu mal classés dans l'ancienne arbre sera augmenté alors le poids des individu k bien prédit dans l'ancien arbre sera diminué ainsi dans le nouvel arbre le poids des individus bien classé issu de l'ancien sera donné par la formule suivante :

$$w_{k-actuel} = w_k * e^{(-P)} \quad (4.7)$$

Cependant la particularité est que dans le nouvel arbre construit la somme des nouveaux poids ne sera pas égal à 1. Pour éviter cela, une normalisation sera faite par rapport à la somme totale des nouveaux poids calculés pour le nouvel arbre.

Donc pour tout individu (j) du nouvel arbre, son poids normalisé est donné par la formule suivante :

$$w_{j-actuel-normalisé} = \frac{w_{j-actuel}}{\sum_1^m w_{j-actuel} + \sum_1^o w_{k-actuel}} \quad \forall j \in m, \forall k \in O \quad (4.8)$$

Pour tout individu (k) du nouvel arbre, son poids normalisé est donné par la formule suivante :

$$w_{k-actuel-normalisé} = \frac{w_{k-actuel}}{\sum_1^m w_{j-actuel} + \sum_1^o w_{k-actuel}} \quad \forall j \in m, \forall k \in O \quad (4.9)$$

Dans l'étape 4, l'algorithme va créer un deuxième arbre, par ailleurs les individus pour ce nouvel arbre seront choisis aléatoirement à partir du premier arbre. Cependant la probabilité de sélectionner un individu j ou k du précédent arbre est égale au poids de l'individu j ou k. Ainsi nous notons que plus le poids de l'individu augmente et plus sa chance de faire partie du deuxième arbre augmente. En effet c'est le principe du coefficient de *gini* pondéré qui peut être consulté à la référence suivante (kwamimayeden, 2022a).

Cependant le processus de l'étape 2 jusqu'à l'étape 4 est itératif et est conditionné par le nombre d'arbre qu'on désire créée.

Si le nombre d'arbre est atteint alors l'algorithme passe à l'étape 5.

Dans l'étape 5, les arbres fournissent la prédiction des valeurs de sortie. Pour cela chaque arbre dispose d'un poids de décision traduisant sa performance ainsi les performances des arbres qui retournent une valeur de sortie spécifique seront additionnées.

De ce fait, la décision sera prise en faveur du groupe ayant la somme des performances la plus élevée. Autrement dit la prédiction de la variable dépendante est donnée par la prédiction obtenue par le groupe d'arbre qui aurait une performance totale plus élevée.

Il existe de nombreuses variantes d'algorithme d'AdaBoost, mais nous pouvons fournir une version ci-dessous en l'adaptant aux contextes de l'article écrit par kwamimayeden(2022b).

Algorithme 4.1 Principe AdaBoost adapté de kwamimayeden (2022b)

AdaBoost - adapté de kwamimayeden (2022b)

Entrée : Ensemble donnée entraînement (X_{train}, y_{train})

Nombre maximal d'itération fixé à T

Sortie : Fournir la prédiction finale de la variable dépendante (y) en combinant la performance des arbres $h(X)$ pondérés par leur performance

- 1 **Initialisation des individus :** Attribuer un poids égal pour chaque individus $w_i = 1/n$
- 2 **Pour** chaque itération $t = 1, 2, 3, \dots, T$
- 3 Entrainer un premier arbre de décision h_t sur les individus pondérés
- 4 Calculer l'erreur totale ϵ_t de l'arbre de décision h_t en additionnant les poids des individus mal classé
- 5 Calculer la performance P de l'arbre de décision h_t à l'aide de la formule $P = 0.5 * \log((1 - \epsilon_t) / \epsilon_t)$
- 6 Mettre à jour les poids des individus (i)
- 7 **Si** individu (j) mal classifié par h_t :
- 8 $w_{(j-actuel)} = w_j * \exp(P)$
- 9 **Si** individu (k) bien classifié par h_t :
- 10 $w_{(k-actuel)} = w_k * \exp(-P)$
- 11 Normaliser les poids des individus en les divisant par la somme totale
- 12 **Si** individu (j) mal classifié par h_t :
- 13 $w_{(j-actuel)} = w_{(j-actuel-normalisé)}$
- 14 **Si** individu (k) bien classifié par h_t :
- 15 $w_{(k-actuel)} = w_{(k-actuel-normalisé)}$
- 16 **Fin**

4.4.4 Algorithme k plus proche voisin

Cet algorithme appelé encore KNN, il est considéré comme étant un algorithme utilisable pour la classification et pour la régression. Son fonctionnement est relativement simple et se base sur la prédiction des échantillons les plus proches dans l'espace de caractéristiques afin de fournir la prédiction finale d'une classe d'un échantillon.

Sa mise en application repose sur plusieurs étapes dont nous pouvons citer :

- Un calcul de distance euclidienne effectué entre l'échantillon inconnu et tous les autres échantillons présents dans l'ensemble de données d'apprentissage.

La distance euclidienne D entre deux échantillons p et q situé dans un espace à n dimension, où les vecteurs caractéristiques représentant les échantillons varie de 1 à n peut être donné par la formule suivante tiré de Zhang (2016) :

$$D(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad \forall i \in [1, n] \quad (4.10)$$

Cependant d'autres type de distance peuvent être calculée à savoir les mesures de distance de Manhattan et celle de Minkowski (Chauhan, 2021).

La distance de Manhattan est donnée par la formule suivante

$$D(p, q) = \sum_{i=1}^n |p_i - q_i| \quad \forall i \in [1, n] \quad (4.11)$$

La distance de Minkowski est donnée par la formule suivante

$$D(p, q) = \left(\sum_{i=1}^n (|p_i - q_i|)^k \right)^{\frac{1}{k}} \quad \forall i \in [1, n] \quad (4.12)$$

Où,

k est un paramètre réel positif contrôlant la norme donc si $k=1$ alors la distance de Minkowski devient une distance de Manhattan.

Si $k = 2$, alors la distance de Minkowski sera une distance de type euclidienne.

- Une phase de sélection des voisins k les plus proches pour cela, k est un paramètre d'entrée à l'algorithme qu'il faut fixer au préalable.
- Un principe de vote majoritaire est mis en place, cependant dans le cadre d'une classification d'un individu, sa classe est donnée par la classe majoritaire parmi les k voisins les plus proches. Dans le cas d'une régression, la valeur de la variable dépendante équivaut à la moyenne des valeurs des k plus proches voisins.

Ainsi pour un KNN, si la classe majoritaire ou la valeur cible est connue alors le résultat est attribué à l'échantillon comme prédiction. Malgré la simplicité de son fonctionnement, sa mise en application sur de grands ensembles de données peut être coûteux en termes de temps de calcul. De la même façon, la performance de cet algorithme dépend fortement de la valeur de k donc un mauvais k pourrait entraîner des prédictions fausses.

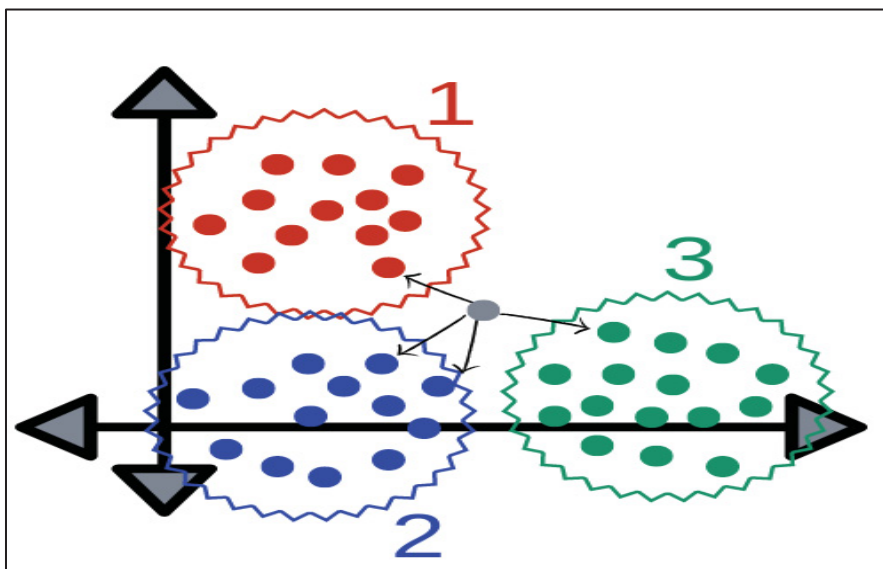


Figure 4.8 Principe des k plus proches voisins

Après avoir défini les algorithmes utilisés dans notre pipeline nous présenterons les résultats dans la prochaine partie.

4.5 Présentation des résultats

Nous avons entraîné le modèle sur les neufs premiers mois de janvier à septembre et tester le modèle sur les données d'octobre à décembre. Afin de comprendre les résultats, nous expliquerons les métriques utilisées ci-dessous.

4.5.1 Précision d'un modèle

La précision sert à évaluer un modèle de classification. Elle se définit comme étant le ratio de prédiction positives correctes parmi toutes les prédictions positives effectuées par le modèle. Ainsi elle s'exprime par la formule suivante :

$$\text{Précision} = \frac{TP}{TP + FP} \quad (4.13)$$

Où

TP = Nombre de vrais positifs,

FP = Nombre de faux positifs

Elle est donnée en pourcentage, plus la précision est proche de 100% et plus le modèle peut être considéré comme performant. Cependant la précision a des limites. Un ensemble de données est dit déséquilibré lorsqu'il y a un écart de taille c'est-à-dire pour un ensemble de données disposant d'une variable dépendante (y) composé de y_a et y_b , alors y_a et y_b sont appelées des classes de ce fait, lorsque la taille de y_a est largement supérieure ou largement inférieure à celle de y_b , alors le jeu de données est dit déséquilibré. Dans de tel situation la précision du modèle sera plus élevée car il prédira systématiquement la classe majoritaire mais ne garantira pas que le modèle est performant. Pour cela d'autres métriques devront être utilisés.

4.5.2 Rappel

Il est encore appelé la sensibilité ou *recall* ou *sensitivity*, il mesure le taux de vrai positif et peut être considéré comme un bon indicateur de performance car il permet de vérifier la capacité d'un modèle de classification à détecter tous les exemples positifs réels dans un ensemble de données. Il est donné par la formule suivante

$$Rappel = \frac{\text{Nombre de vrais positifs}}{\text{Nombre de vrais positifs} + \text{Nombre de faux négatifs}} \quad (4.14)$$

Dans cette formule, le nombre de vrais positifs signifient les exemples positifs réels correctement prédits par le modèle et les faux négatifs signifie les exemples positifs réel incorrectement prédits. Le rappel est aussi un bon indicateur de performance dans les cas où la prédiction d'exemple positif est primordiale et ou les faux négatifs ne sont pas tolérés. Nous pouvons citer l'exemple de notre recherche où nous cherchons à prédire la défaillance de composant. Si une défaillance de composant 2 est prédit comme étant une non-défaillance alors ce composant tombera en panne et engendrera une maintenance corrective. Dans de tel situation le rappel serait un bon indicateur pour évaluer notre modèle. Souvent la précision et le rappel sont combinés pour vérifier la performance d'un modèle.

4.5.3 F-mesure

Elle se nomme également F1-score et est considéré comme un indicateur de performance globale car elle regroupe les métriques de précision et de rappel dans un seul métrique qui est donnée par la formule suivante

$$F - \text{mesure ou } F1 - \text{score} = \frac{2 * (\text{Précision} * \text{Rappel})}{(\text{Précision} + \text{Rappel})} \quad (4.15)$$

La valeur de la F-mesure est comprise entre 0 et 1. Plus cette valeur est proche de 1 et plus la performance du modèle de classification est optimale cependant une valeur nulle signifie une performance faible. La F-mesure est souvent utilisée lorsque les classes sont déséquilibrées car elle permet de vérifier d'une part si le modèle est capable de prédire correctement les vrais

positifs et d'autre part de vérifier sa précision globale. Cependant cette métrique permet de vérifier l'équilibre entre le rappel et la précision.

4.5.4 Exactitude

L'exactitude ou *accuracy* évalue la performance globale d'un modèle en évaluant le ratio de prédictions correctes sur l'ensemble des prédictions effectués peu importe qu'elles soient positives ou négatives. Elle est donnée par la formule suivante

$$Accuracy = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}} \quad (4.16)$$

4.5.5 Moyenne macroscopique (Macro avg)

La moyenne macroscopique est une mesure de performance par classe qui calcule la moyenne des métriques comme la précision, le rappel, la F-mesure pour chaque classe. La macro-avg affecte un poids égal à chaque classe sans tenir compte de sa taille ainsi chaque classe pourra contribuer de manière égale à la mesure agrégé.

4.5.6 Moyenne pondérée (Weighted avg)

En ce qui concerne le *weighted avg*, elle est connue sous l'appellation de moyenne pondérée et se définit comme une méthode qui agrège les métriques de performance par classe. Son fonctionnement se résume à l'attribution de poids différents à chaque classe en fonction de leur importance relative. Par ailleurs, le poids peut être fixé en fonction de la fréquence de chaque classe dans l'ensemble de données. La moyenne pondérée se calcule en deux phases dont la première est de calculer la métrique pour chaque classe et la deuxième phase est de trouver la moyenne pondérée des métriques en prenant en compte la fréquence de classe qui sera le poids alloué (Tremblay, 2021)

Elle est donnée par la formule suivante :

$$\text{Moyenne pondérée (weighted avg)} = \frac{\sum_{i=1}^N (\text{Poids}_i * \text{métrique}_i)}{\sum_{i=1}^N \text{Poids}_i} \quad (4.17)$$

Où N désigne le nombre de classe.

La moyenne pondérée peut être obtenue également par la moyenne de tous les F1-scores (Leung, 2022).

4.5.7 Matrice de confusion

La matrice de confusion est une représentation sous forme de tableau qui permet de visualiser les performances d'un modèle de classification pour cela les prédictions sont comparés aux valeurs réelles. Le nombre de classe à prédire détermine la taille de la matrice c'est-à-dire pour une classification binaire « oui » ou « non » (2 classes), nous aurons une matrice de confusion de taille (2*2) ou 2 lignes et 2 colonnes. Chaque ligne représente les instances¹ dans une classe réelle et chaque colonne matérialise les instances prédites par le modèle.

Les éléments que nous trouvons dans la matrice de confusion sont énumérés ci-dessous :

- Vrai Positif ou *True Positive* (TP) signifie que le modèle a prédit correctement les individus positifs
- Vrai Négatives ou *True Negative* (TN) signifie que le modèle a prédit correctement les individus négatifs
- Faux Positifs ou *False Positive* (FP) signifie que le modèle a prédit des individus négatifs comme positifs on l'appelle l'erreur de type I
- Faux Négatifs ou *False Negative* (FN) signifie que le modèle a prédit des individus positifs comme négatif on l'appelle l'erreur de type II

Cette matrice est également appelé dans la littérature tableau de contingence (Data Science Team, 2021).

Dans certains travaux le taux d'erreur est déterminé pour savoir dans quelle proportion un modèle de classification produit des erreurs de prédictions (Rapid Miner, 2023). Elle est calculée par la formule suivante :

¹ Signifie une observation individuelle dans un ensemble de donnée

$$\text{Taux d'erreur} = \frac{FP + FN}{FP + FN + TP + TN} \quad (4.18)$$

Dans notre recherche nous avons utilisé la bibliothèque de *scikit-learn* en python pour générer les matrices de confusion.

La figure ci-dessous illustre le principe d'une matrice de confusion pour deux classes.

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Figure 4.9 Principe de la matrice de confusion tirée de Rapid Miner (2023)

La matrice de confusion nous a permis dans notre recherche de voir pour chaque classifieur de notre modèle les prédictions de défaillance, nous les présentons ci-dessous.

4.5.8 Résultats pour la régression logistique

Les données d'entraînements sont composées de X_{train} et y_{train} . Cela signifie que X_{train} est une matrice dont les lignes représente les observations effectuées sur les neuf premiers mois

et les colonnes sont les variables indépendantes provenant des données prétraitées de notre pipeline. y_{train} est la variable failure ou variable dépendante composée uniquement des observations effectuées sur les neufs premiers mois.

Les données de test sont X_{test} et y_{test} où X_{test} est la matrice dont les lignes représente les observations effectuées sur les trois derniers mois et les colonnes sont les variables indépendantes provenant des données prétraitées de notre pipeline. y_{test} est la variable failure composée uniquement des observations effectuées sur les trois derniers mois.

Cependant dans notre y_{test} nous trouvons que le nombre de défaillance du composant 2 est de 32, celui du composant 4 est de 8 et enfin le nombre d'observation sans défaillance (fonctionnement normal) est de 2177. Le but de notre modèle est de prédire pour chaque observation la classe associée (défaillance composant2, ou composant 4 ou de pas de défaillance) et ceci dans les prochaines 24h.

La notion des ensembles de données est résumée ci-dessous

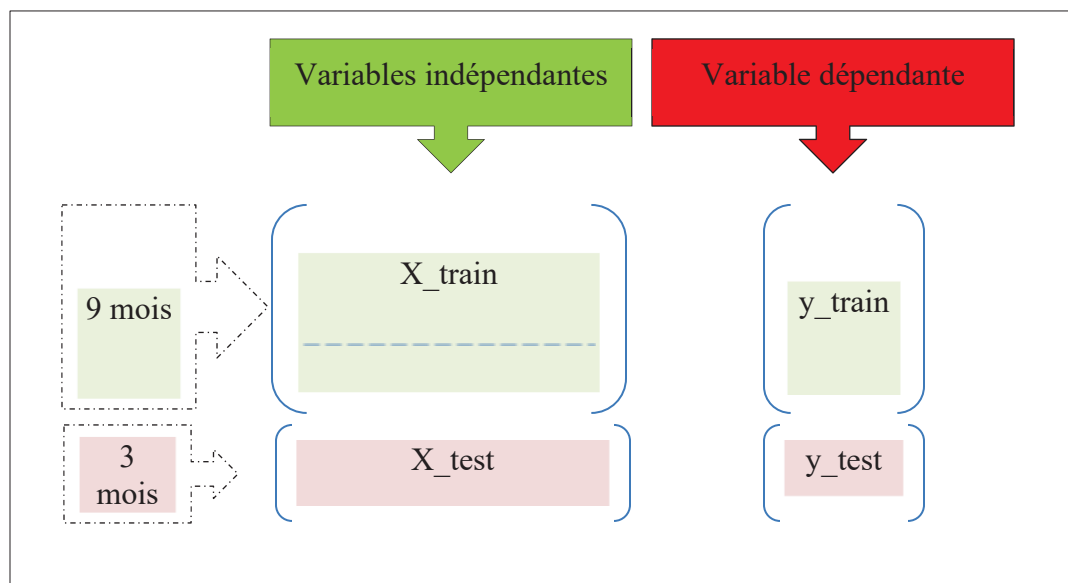


Figure 4.10 Ensemble de formation et de test

L'algorithme de régression logistique nous montre les résultats suivants :

La défaillance du composant 2 a été correctement prédit 32 fois alors que les erreurs de prédictions ont eu lieu 6 fois ce qui fait qu'il y'a 32 prédictions vrais et 6 prédictions fausses donc $TP_{comp2} = 32$, $FP_{comp2} = 6$, $FN_{comp2} = 0$ et $TN_{comp2} = 0$.

Au niveau du composant 4, le modèle l'a bien classifié alors $TP_{comp4} = 8$, $FP_{comp4} = 0$, $TN_{comp4} = 8$ et $FN_{comp4} = 0$. Cependant en fonctionnement normal (none) ou aucune panne, le modèle utilisant la régression logistique a bien classifié 2171 observations de la classe none ce qui fait que $TP_{none} = 2171$, $FP_{none} = 0$, $FN_{none} = 6$ et $TN_{none} = 2171$.

Ainsi sa matrice de confusion est donnée ci-dessous

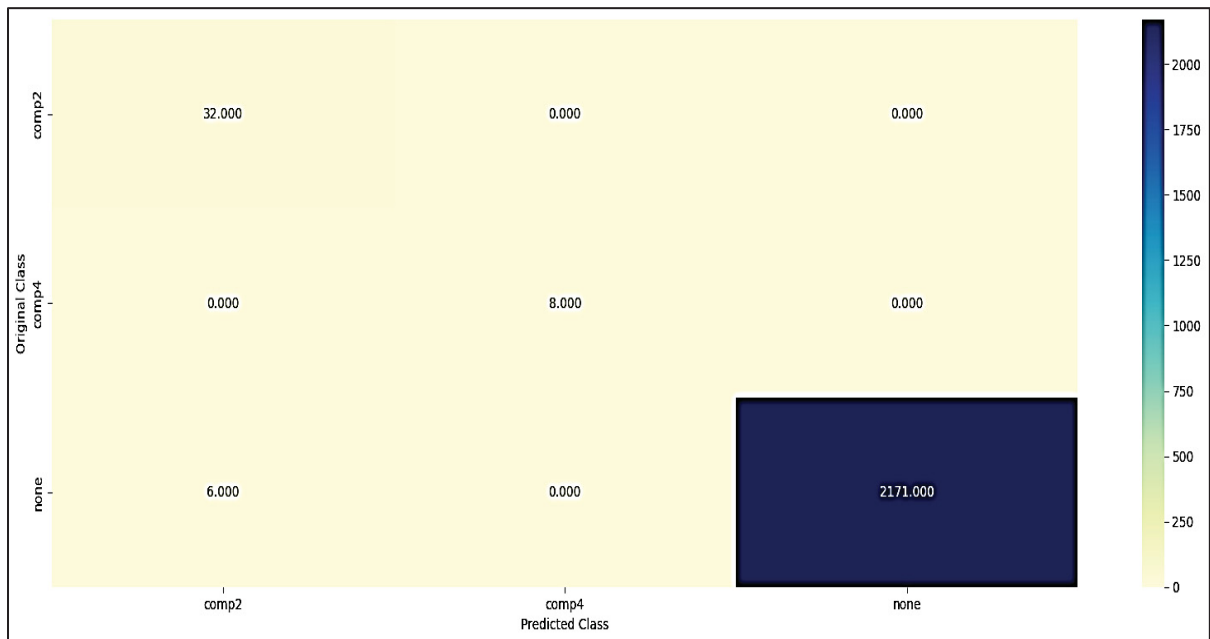


Figure 4.11 Matrice de confusion par la régression logistique

La précision et le rappel ont été calculée par la manière suivante

- La précision sur composant 2 sera donnée par $32 / 32 + 6 = 0.84$
- La précision sur composant 4 sera donnée par $8 / 8 + 0 = 1$
- La précision sur none sera donnée par $2178 / 2178 + 0 = 1$
- Le rappel sur composant 2 sera donné par $32 / 32 + 0 = 1$
- Le rappel sur composant 4 sera donnée par $8 / 8 + 0 = 1$
- Le rappel sur none sera donné par $2178 / 2178 + 0 = 1$

L'ensemble de ces métriques a été résumé dans un rapport de classification.

Tableau 4.11 Rapport de classification par la régression logistique

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	84	100	91	32
Comp4	100	100	100	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	95	100	97	2217
Moyenne pondérée	100	100	100	2217

Le rapport de classification obtenu par la régression logistique montre que ce modèle a un bon score de rappel mais la précision n'est pas parfaite au niveau du composant 2.

4.5.9 Résultats pour la forêt aléatoire

Le modèle du forêt aléatoire a effectué des prédictions justes pour chaque classe ce qui fait que la défaillance du composant 2 a été correctement prédit 32 fois donc $TP_{comp2} = 32$, $FP_{comp2} = 0$, $FN_{comp2} = 0$ et $TN_{comp2} = 0$. Au niveau du composant 4, le modèle l'a bien classifié alors $TP_{comp4} = 8$, $FP_{comp4} = 0$, $TN_{comp4} = 8$ et $FN_{comp4} = 0$.

Ainsi sa matrice de confusion est donnée ci-dessous

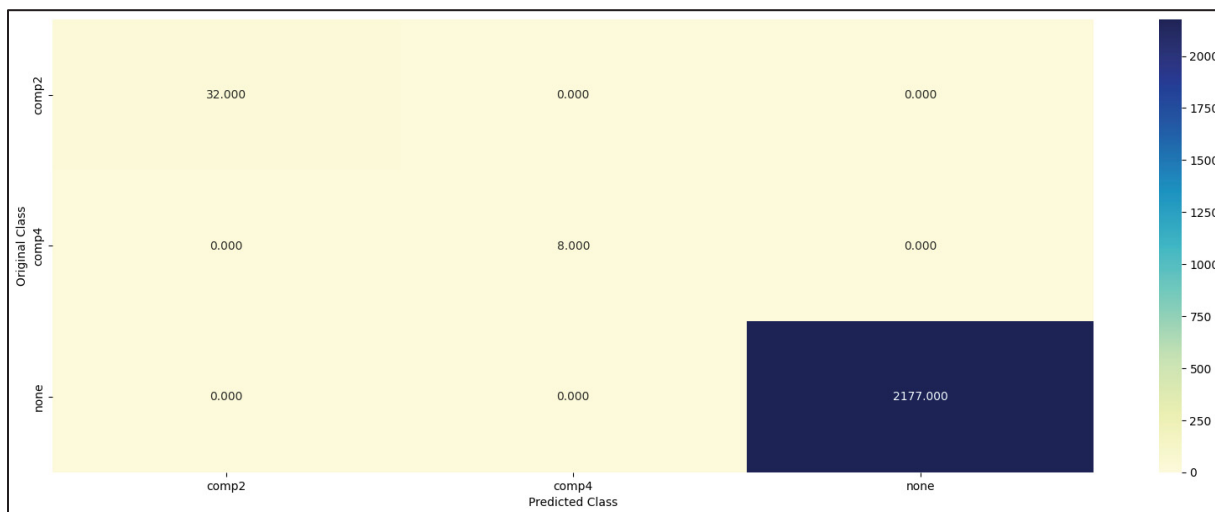


Figure 4.12 Matrice de confusion par la forêt aléatoire

La précision et le rappel ont été calculée par la manière suivante

- La précision sur composant 2 sera donnée par $32 / 32 = 1$
- La précision sur composant 4 sera donnée par $8 / 8 = 1$
- La précision sur none sera donnée par $2178 / 2178 = 1$
- Le rappel sur composant 2 sera donné par $32 / 32 = 1$
- Le rappel sur composant 4 sera donnée par $8 / 8 = 1$
- Le rappel sur none sera donné par $2178 / 2178 = 1$

L'ensemble de ces métriques ont été résumé dans un rapport de classification.

Tableau 4.12 Rapport de classification pour la forêt aléatoire

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	100	100	100	32
Comp4	100	100	100	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	100	100	100	2217
Moyenne pondérée	100	100	100	2217

Le rapport de classification obtenu par le foret aléatoire montre des résultats satisfaisants.

4.5.10 Résultats pour l'AdaBoost

Le modèle d'AdaBoost a prédit correctement 5 défaillances du composant 2 cependant une défaillance du composant 2 a été prédit comme non défaillant donc nous pourrions dire que $TP_comp2 = 5$, $FP_comp2 = 1$, $TN_comp2 = 5$ et $FN_comp2 = 27$.

Au niveau du composant 4, toutes ses défaillances ont été bien classifiées ce qui fait que $TP_comp4 = 2$, $TN_comp4 = 2$, $FP_comp4 = 0$, $FN_comp4 = 0$.

La non-défaillance de composant a été prédit correctement sur 2176 observations, en revanche ce modèle a prédit à 6 reprises une non-défaillance alors qu'en réalité le composant 4 tombera en panne. De la même façon le modèle a prédit faussement un état de non-défaillance sur 27

observations alors que le composant 2 tombera en panne donc nous aurons $TP_none = 2176$, $TN_none = 2176$.

Ainsi sa matrice de confusion est donnée ci-dessous

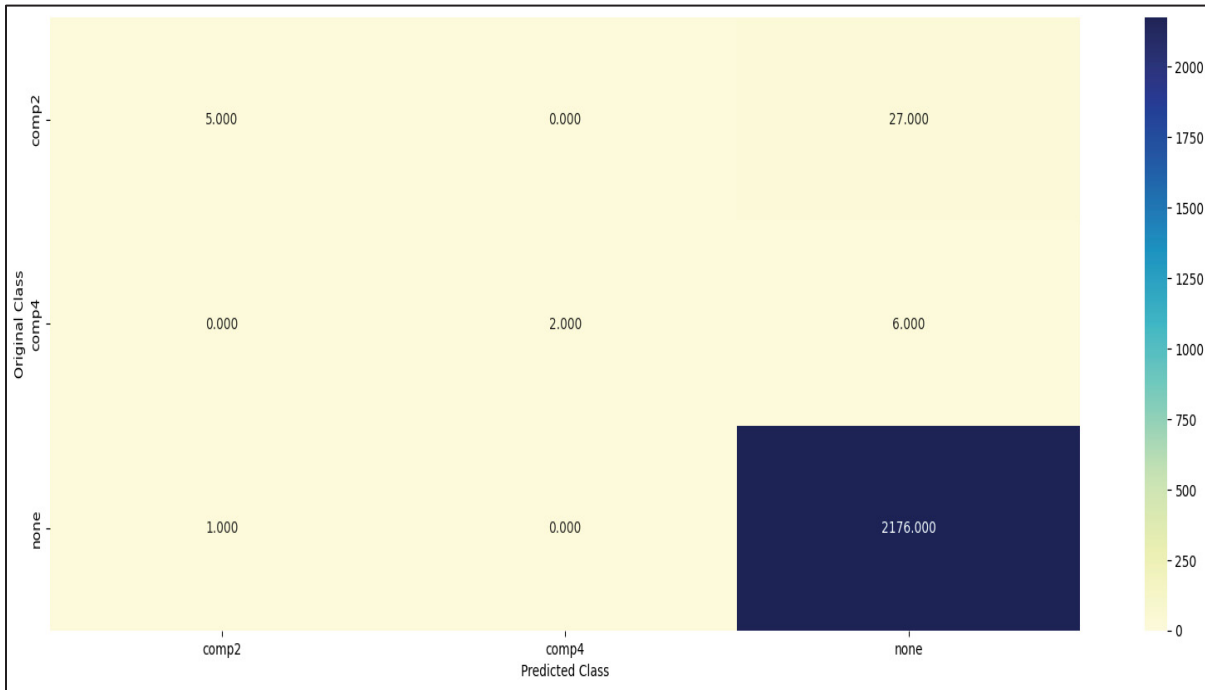


Figure 4.13 Matrice de confusion initial pour AdaBoost

La précision et le rappel ont été calculée par la manière suivante

- La précision sur composant 2 sera donnée par $5 / 5+1 = 0.83$
- La précision sur composant 4 sera donnée par $2 / 2+0 = 1$
- La précision sur none sera donnée par $2176 / 2176+6 = 0.99$
- Le rappel sur composant 2 sera donné par $5 / 5+27 = 0.16$
- Le rappel sur composant 4 sera donnée par $2 / 2+ 6 = 0.25$
- Le rappel sur none sera donné par $2176 / 2176+1 = 1$

L'ensemble de ces métriques ont été résumé dans un rapport de classification.

Tableau 4.13 Rapport de classification pour AdaBoost

	Précision (%)	Rappel (%)	F1-score (%)	Support (%)
Comp2	83	16	26	32
Comp4	100	25	40	8
none	99	100	99	2177
Exactitude			98	2277
Moyenne macroscopique	94	47	55	2217
Moyenne pondérée	98	98	98	2217

4.5.11 Résultats pour algorithme des k plus proche voisin

Le modèle des KNN a prédit correctement toutes les défaillances de composant 2 et 4. Ainsi nous aurons pour le composant 2, $TP_comp2 = 1$, $TN_comp2 = 1$, $FP_comp2=0$ et $FN_comp2 = 0$.

Pour le composant 4, $TP_comp4 = 1$, $TN_comp4 = 1$, $FP_comp4=0$ et $FN-comp4 = 0$.

En fonctionnement normal nous aurons $TP_none = 1$, $TN_none = 1$, $FP_none=0$ et $FN_none = 0$.

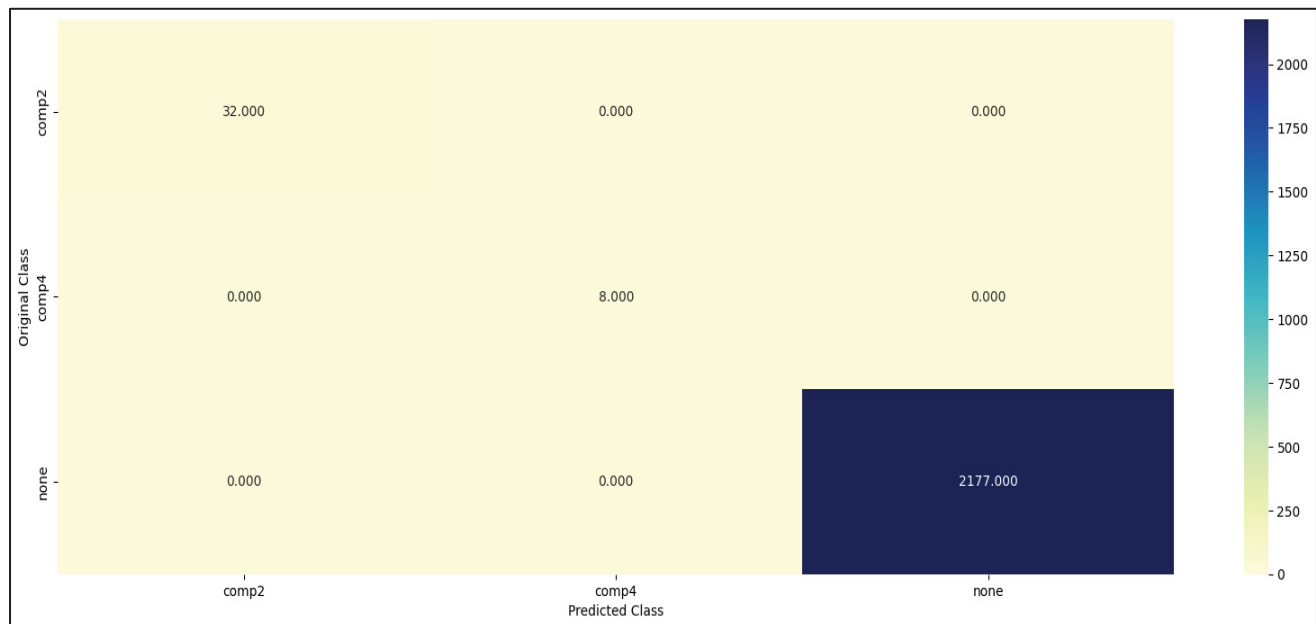


Figure 4.14 Matrice de confusion initiale pour KNN

La précision et le rappel pour chaque classe sera de 100 % comme sur le tableau ci-dessous.

Tableau 4.14 Rapport de classification KNN

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	100	100	100	32
Comp4	100	100	100	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	100	100	100	2217
Moyenne pondérée	100	100	100	2217

L'analyse du modèle avec les quatre types de classifieur ou d'algorithmes ont montré que la régression logistique, forêt aléatoire et le KNN semblent être les meilleurs modèles car la précision et le rappel obtenue par ces algorithmes sont les meilleurs (100%) ou presque (84%). Cependant une validation des résultats sera faite afin de confirmer le modèle qui sera retenu dans nos travaux.

4.6 Validation du modèle

Dans cette partie, nous allons expliquer la notion de surapprentissage ou d'*overfitting* car il permet de vérifier que les performances sur les données d'entraînement sont égales ou approximativement égales à la performance obtenue sur les données de test.

4.6.1 Vérification surapprentissage

Autrement dit le surapprentissage est étudié pour vérifier si le modèle est capable de se généraliser ou de fournir de bonnes prédictions sur des données qu'elle n'a jamais vu auparavant. Or nous savons bien que la mise en place d'un modèle prédictif est de pouvoir faire des prédictions sur des situations nouvelles.

Ainsi nous comparerons les performances des données d'entraînement et de test et nous les résumerons dans le tableau ci-dessous.

Tableau 4.15 Vérification surapprentissage

	Régression Logistique	Foret Aléatoire	AdaBoost	KNN
Données d'entraînement (X_train, y_train)	1	1	0,99	1
Données de test (X_test, y_test)	0,997	1	0,985	1

La vérification du surapprentissage permet de constater que les performances des données d'entraînement et de test pour chaque algorithme sont soit égale ou soit sensiblement égale donc il n'y a pas de surapprentissage.

A présent nous étudierons la courbe ROC qui signifie (*Receiver Operating Characteristic*) qui est très utilisée pour vérifier les performances d'un modèle de classification.

4.6.2 Courbe ROC des algorithmes

La courbe ROC est une visualisation graphique qui permet d'évaluer un modèle de classification. Son principe est de comparer le taux de vrai positifs (taux des TP) et le taux de faux positifs (taux des FP) par rapport aux données réelles. Cependant chaque point correspond à une paire (taux des TP, taux des FP) défini sur un certain seuil de classification. Autrement dit un seuil correspond à une probabilité définie pour une classe spécifique qui à partir duquel un individu ou exemple sera prédit comme appartenant à cette classe. Par exemple si nous supposons que pour un individu puisse être prédit à la classe alpha, l'algorithme fixe une probabilité de 0,60 comme seuil donc si l'algorithme estime que la probabilité de l'individu est 0.65 qui est supérieur à 0,60 alors l'individu sera prédit comme appartenant à la classe alpha. De ce fait, une courbe de ROC idéale se situerait dans le coin gauche du graphique et correspond à un taux de TP élevé et un taux de FP faible, en dessous de la ligne du diagonale, le modèle est considéré comme non performant.

Ainsi la courbe ROC va utiliser différents seuils pour évaluer la performance d'un modèle prédictif.

La courbe ROC obtenue par la régression logistique est défini ci-dessous

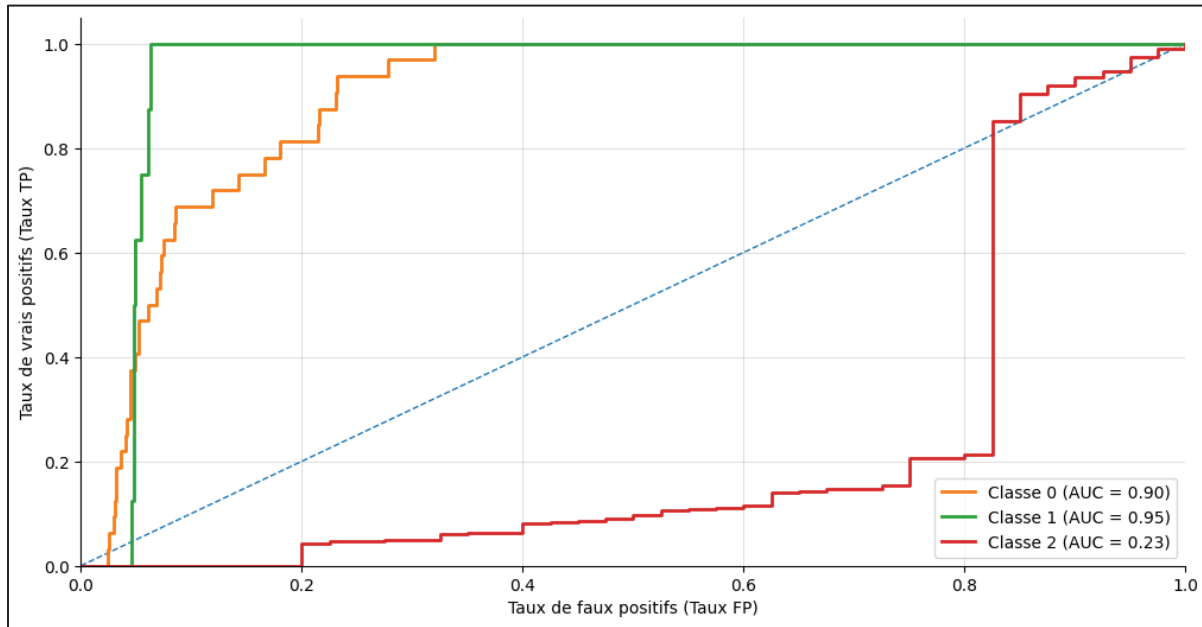


Figure 4.15 Courbe ROC initiale par Régression Logistique

La courbe ROC permet de voir les classes 0 et 1 correspondant respectivement aux classes none et composant 2 présente des AUC (*Area Under the Curve*) ou (aire sous la courbe) très élevée donc nous dirons bien que la précision ne soit pas élevée, nous réglerons les hyperparamètres pour voir si ce modèle pourra être retenu.

La courbe ROC de la forêt aléatoire est donnée ci-dessous

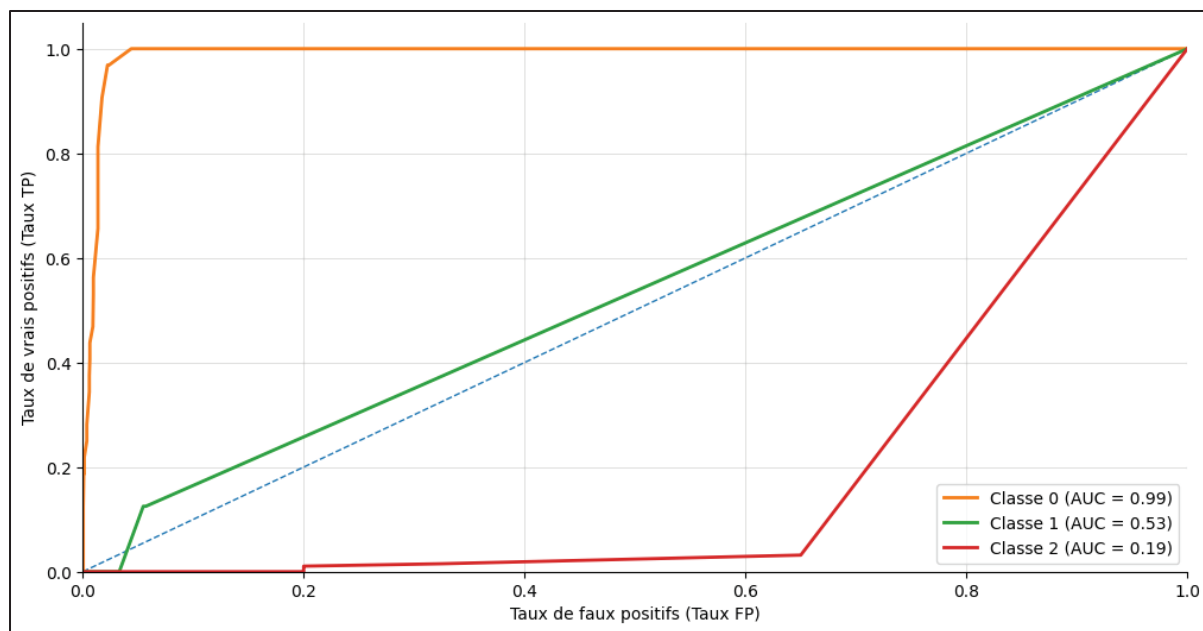


Figure 4.16 Courbe ROC initiale par la Forêt aléatoire

Nous remarquons que la performance de prédiction sur les classes 0 et 1 sont bien mais peuvent être améliorés.

La courbe de ROC pour l'adaBoost est donnée ci-dessous et permet de vérifier que l'AUC de la classe 2 est bonne et est proche de 1, donc l'adaBoost peut prédire correctement le composant 4. Cependant le mode normal et le composant 2 ont des AUC pas excellent et même l'AUC pour le composant 2 est nulle.

Cela prouve que l'AdaBoost ne sera pas utilisé dans notre modèle prédictif.

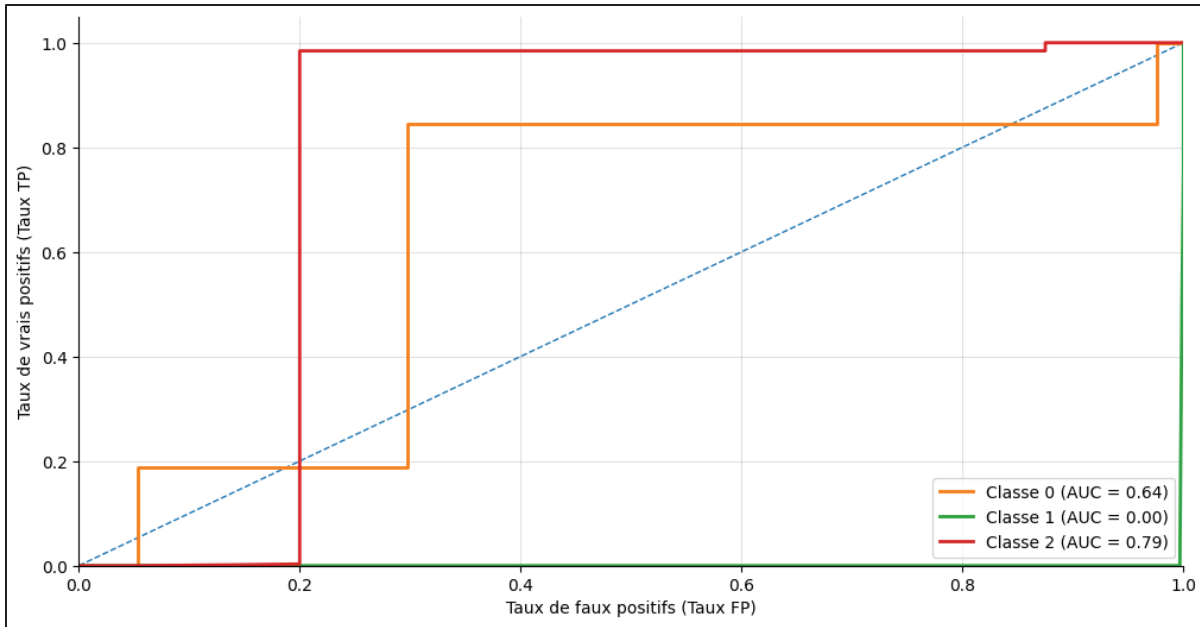


Figure 4.17 Courbe ROC pour AdaBoost

La courbe ROC pour le KNN est donnée ci-dessous et permet de constater que le KNN présente des AUC proche de la ligne médiane donc le modèle avec le KNN a des performances ressemblant à un modèle aléatoire. Donc nous ne sélectionnerons pas le KNN comme modèle.

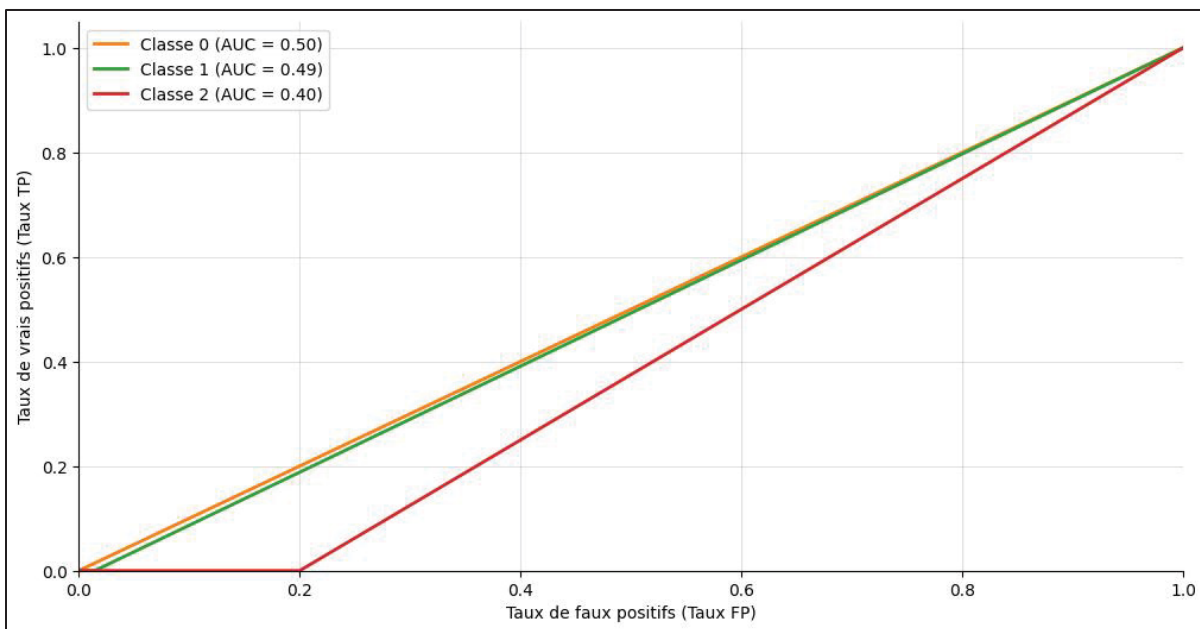


Figure 4.18 Courbe ROC pour KNN

Les algorithmes que nous retiendrions pour notre modèle seront la régression logistique et le forêt aléatoire. Cependant, nous essayeront d'ajuster les hyperparamètres pour vérifier si les performances peuvent augmenter

4.7 Optimisation des hyperparamètres

Les hyperparamètres sont des paramètres qui peuvent influencer la performance du modèle. Ainsi la figure ci-dessous montre le principe utilisé pour effectuer le réglage des hyperparamètres.

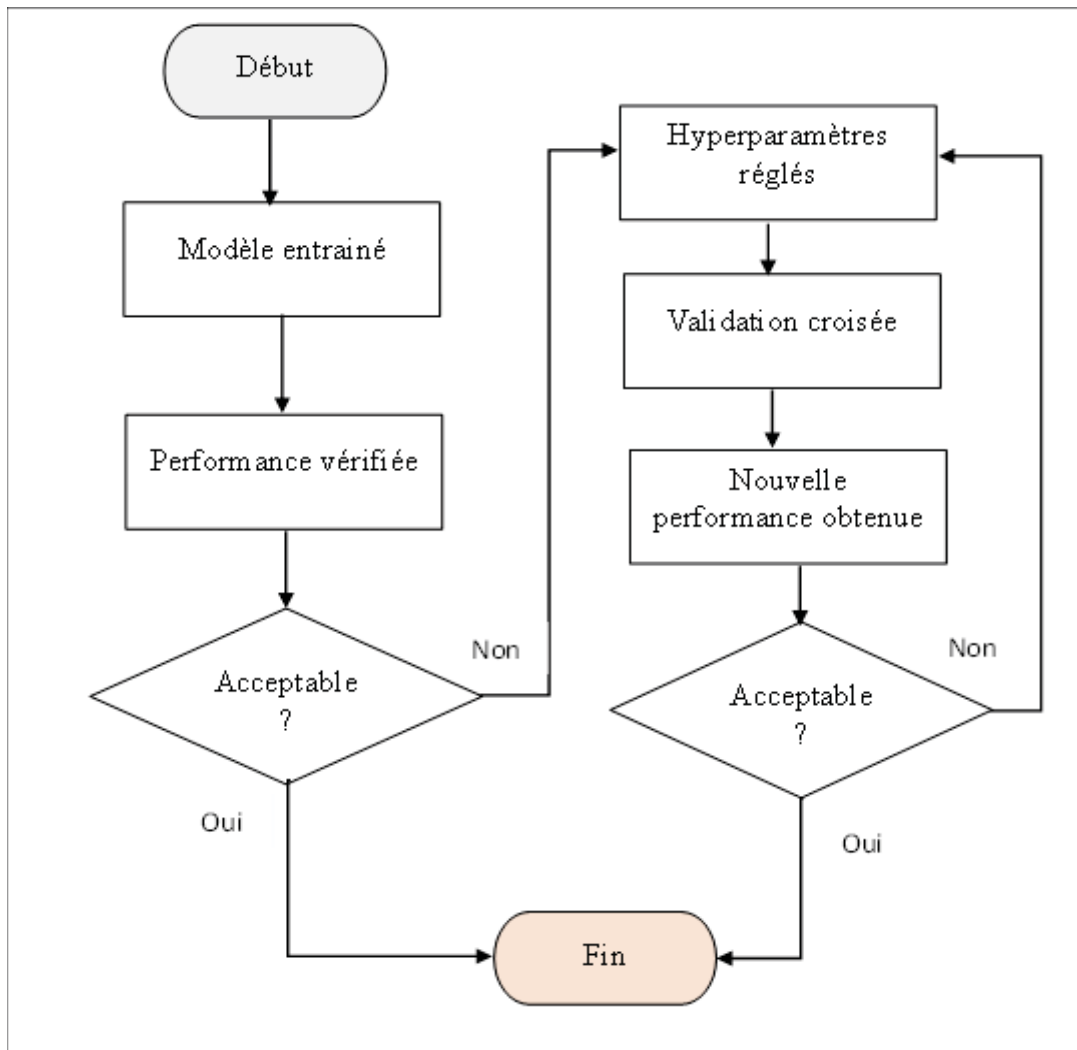


Figure 4.19 Principe du réglage des hyperparamètres

4.7.1 Hyperparamètre pour la Régression Logistique

Au niveau de la régression logistique nous avons effectués les réglages suivants :

- **Régularisation** sur L1
- **Inverse de la force de régularisation** ou **C** réglé sur les valeurs suivants 1, 3, 2 et 5
- **Optimisation de la fonction de cout et ajustement des poids du modèle ou (Solver)** réglé sur liblinear
- Le **nombre maximum d'itération** ou **max-iter** sur les valeurs de 100 et 500

La régularisation est utilisée comme hyperparamètre pour éviter le surapprentissage ainsi s'elle est réglée sur L1 ou (Lasso) va favoriser la sélection des caractéristiques importantes. En ce qui concerne le paramètre C nous pouvons dire que plus sa valeur est petite et plus la régularisation sera forte ainsi le C est déterminé par validation croisée fixé à 10 dans notre cas. De ce fait une validation croisée permet de diviser un ensemble de donné en n plis ou *folds*, ce qui fait que le modèle va s'entraîner sur une partie n plis et l'autre partie restante sera utilisée pour la validation et le test.

Pour le cas du solver, il agit sur la vitesse de convergence de l'algorithme ainsi un solveur réglé sur liblinear sera plus efficace car la taille de notre ensemble de donnée est raisonnable ni trop petite ni trop grande.

Cependant le nombre maximum d'itération permet de définir le nombre de fois, où l'algorithme fera des itérations pour trouver la valeur optimale ou la meilleure solution. Ainsi les meilleurs hyperparamètres pour la régression logistique en fonction de nos valeurs d'entrées sont :

- C=1
- Max itération = 100
- Pénalité = L1
- Solver = liblinear

Ainsi le rapport de classification est donné ci-dessous

Tableau 4.16 Rapport de classification par réglage hyperparamètre de la régression logistique

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	91	100	96	32
Comp4	57	100	73	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	83	100	89	2217
Moyenne pondérée	100	100	100	2217

Nous remarquons que la précision du composant 2 a légèrement augmenté par rapport à avant mais celle du composant 4 s'est dégradé ainsi nous dirons que le réglage des hyperparamètres n'a pas amélioré les performances pour la régression logistique. De ce fait la courbe de ROC ci-dessous le confirme

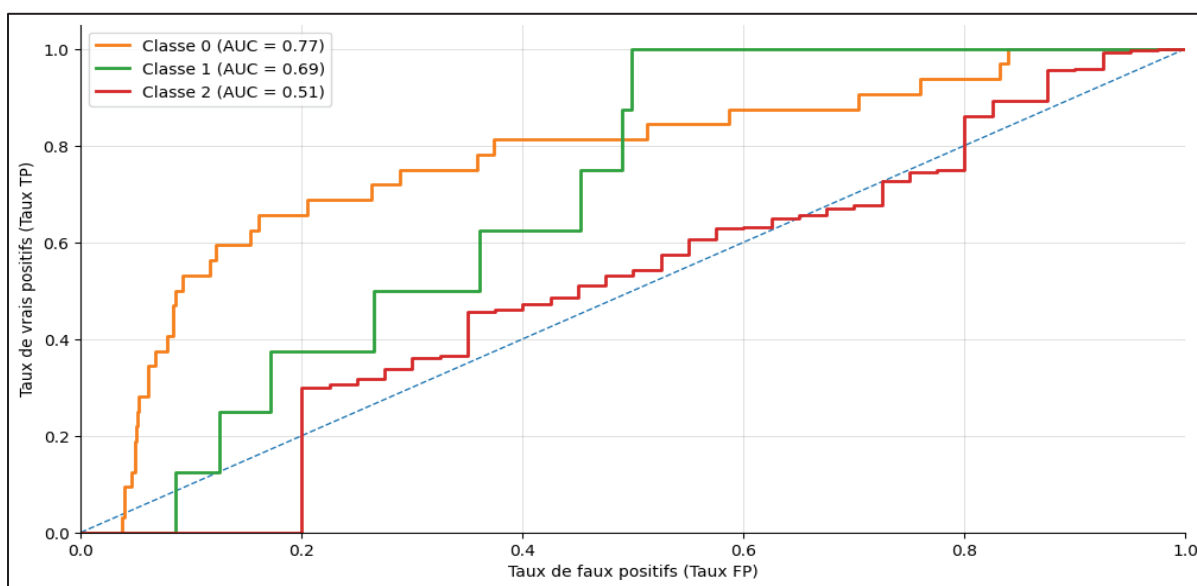


Figure 4.20 Courbe ROC par réglage des hyperparamètres de la régression logistique

La courbe ROC nous montre que les AUC ont fortement diminués à part l'AUC pour le composant 4 qui a légèrement augmenté. Le réglage d'hyperparamètres permet d'améliorer la performance mais dans certains cas les performances du modèle vont se dégrader.

Dans notre cas la performance obtenue est pire ainsi la régression logistique ne sera pas utilisée dans notre modèle final.

4.7.2 Réglage hyperparamètre pour la Forêt Aléatoire

La forêt aléatoire a fourni de bons taux de rappel et de précision mais sa courbe de ROC avait montré un AUC faible pour le composant 4. Ainsi nous régleront les hyperparamètres afin de vérifier si la performance sera meilleure. De ce fait, nous fixerons l'hyperparamètre `n_estimator` à 1000 et en utilisant une validation croisée égale à 10.

L'hyperparamètre `n_estimator` est utilisé dans les forêts aléatoires pour spécifier le nombre d'arbre à utiliser ainsi si le nombre d'arbre utilisé est élevé alors le modèle sera complexe et pourra mieux établir la relation entre les variables dans les données d'entraînement. D'où le nombre fixé à 1000.

Le rapport de classification est fourni ci-dessous

Tableau 4.17 Rapport de classification pour la Forêt aléatoire avec 1000 arbres

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	100	100	100	32
Comp4	100	25	40	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	100	75	80	2217
Moyenne pondérée	100	100	100	2217

Nous remarquons que le taux de rappel du composant 4 a baissé donc le réglage de cet hyperparamètre n'améliore pas la performance ainsi la courbe ROC ci-dessous montre que les performances ont peu augmenté sur les classes 1 et 2.

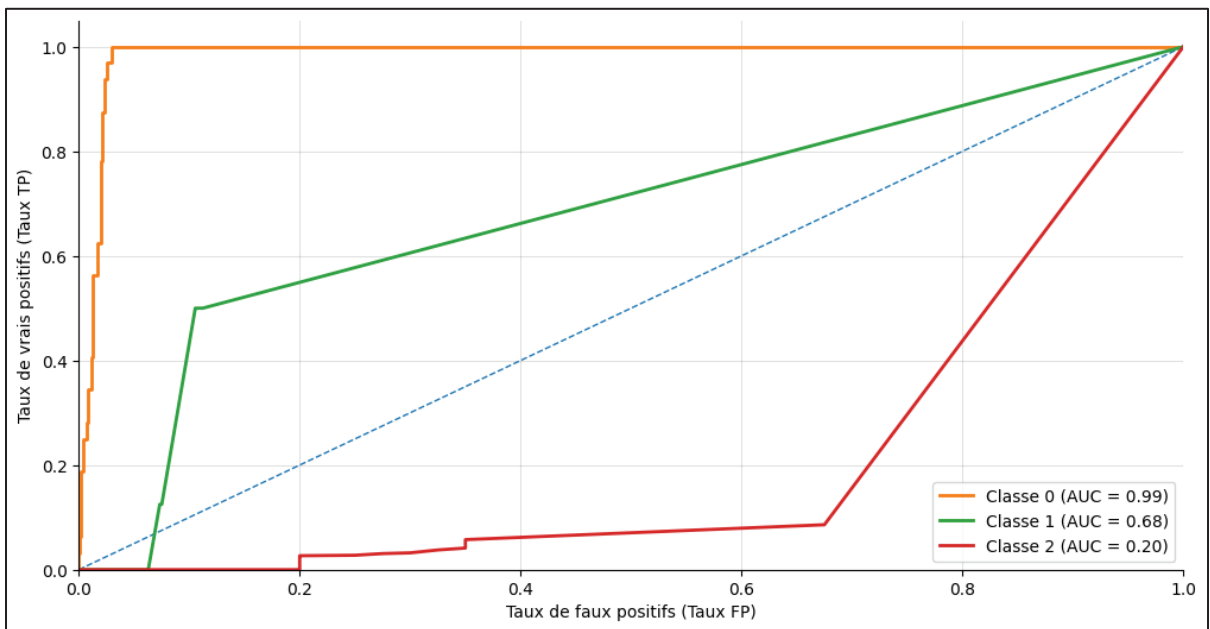


Figure 4.21 Courbe ROC pour forêt aléatoire à 1000 arbres

Le réglage des hyperparamètres a peu amélioré les AUC donc nous retenons la forêt aléatoire sans réglage des hyperparamètres comme modèle de base de notre recherche.

Étant donné que nous avons beaucoup de variables indépendantes, nous chercherons à les diminuer afin d'avoir un modèle de forêt aléatoire qui sera entraîné sur des données plus petites.

4.7.3 Sélection des variables les plus importantes

Nous utilisons les variables les plus importantes afin de réduire la dimension de nos données. Les variables les plus importantes sont résumés sur les figures ci-dessous.

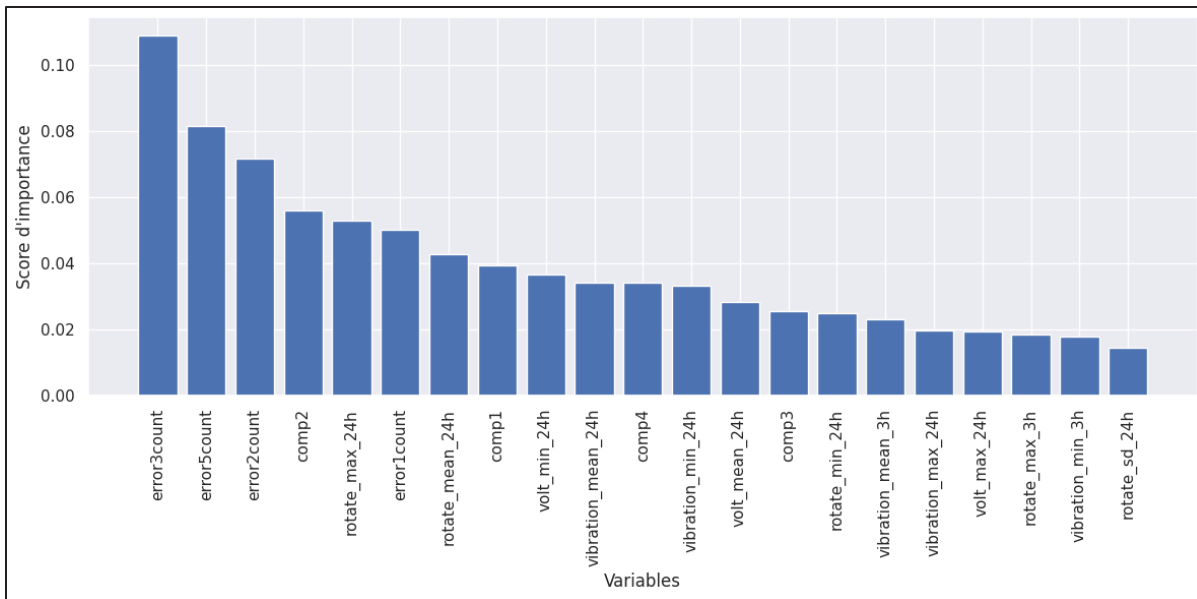


Figure 4.22 Première partie des variables les plus importantes

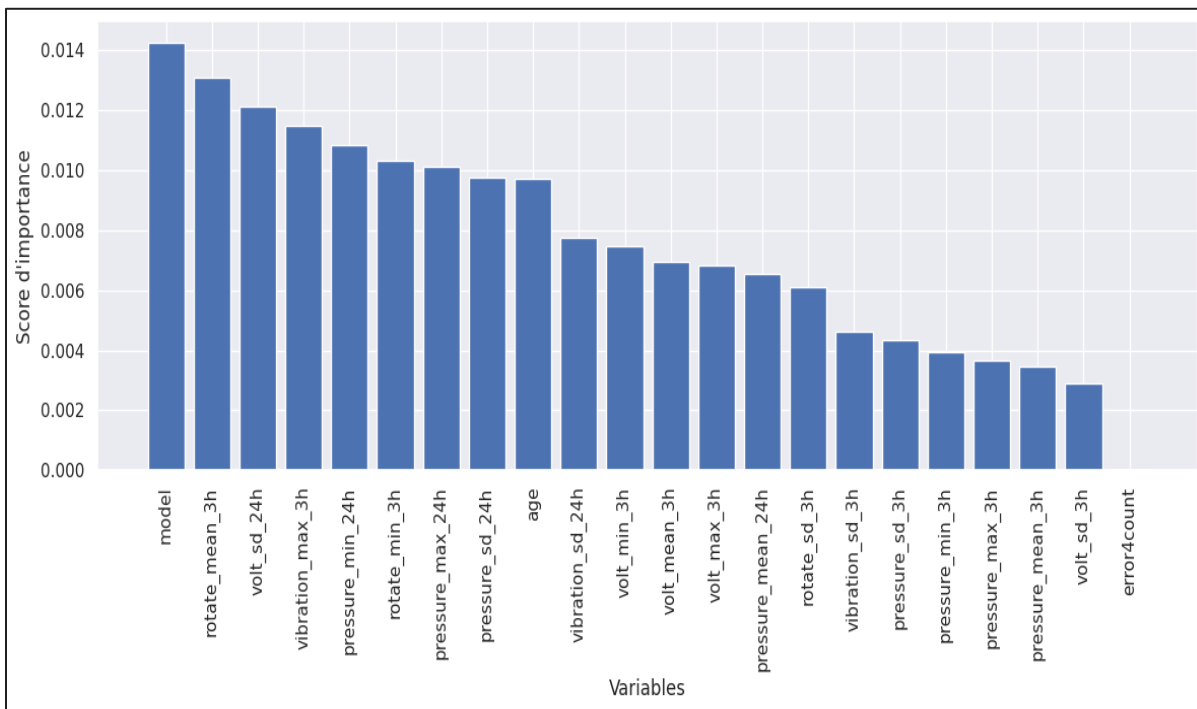


Figure 4.23 Deuxième partie des variables les plus importantes

Afin de déterminer le nombre de variables importantes, nous nous basons sur le score d'importance ainsi nous pourrions fixer une limite au-dessus duquel les variables seront sélectionnées. De ce fait dans notre travail nous prenons un seuil à 0,010.

Ainsi le nombre de variables d'entraînement diminuera et nous aurons un nouvel ensemble d'entraînement composé de :

- X_train_s1 de taille (6523, 28)
- y_train_s1 de taille (6523,)
- X_test_s1 de taille (2217, 28)
- y_test_s1 de taille (2217,)

Le pipeline composé de la forêt aléatoire va s'entraîner sur les données X_train_s1 et y_train_s1 ainsi le rapport de classification est fourni ci-dessous

Tableau 4.18 Rapport de classification obtenu avec forêt aléatoire et seuil S1

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	100	100	100	32
Comp4	100	25	40	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	100	75	80	2217
Moyenne pondérée	100	100	100	2217

Nous constatons que les performances ne sont pas optimales surtout au niveau du rappel pour le composant 4. Ainsi, nous réglerons les hyperparamètres suivants :

- Nombre d'arbres ou n_estimators à 100, 300 et 600
- Profondeur maximale des arbres ou max_depth sur 0 (none), 5 et 10

Les meilleurs hyperparamètres trouvés sont max_depth = 10 et n_estimators = 600

Ainsi le rapport de classification est fourni dans le tableau 4.19.

Tableau 4.19 Rapport de classification avec nombre arbre, profondeur et S1

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	100	100	100	32
Comp4	100	88	93	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	100	96	98	2217
Moyenne pondérée	100	100	100	2217

Le taux de rappel a nettement augmenté passant de 25% à 88% ainsi nous allons continuer à régler les hyperparamètres mais cette fois ci, nous choisissons `n_estimators` à 600 et 1000 et `max_depth` à 10. La RF à trouver les meilleurs hyperparamètres pour `n_estimators` à 1000 et `max_depth` à 10 (une valeur unique dans notre cas).

De ce fait les résultats sont représentés ci-dessous

Tableau 4.20 Rapport de classification pour forêt aléatoire et deuxième réglage sous S1

	Précision (%)	Rappel (%)	F1-score (%)	support
Comp2	100	100	100	32
Comp4	100	100	100	8
none	100	100	100	2177
Exactitude			100	2277
Moyenne macroscopique	100	100	100	2217
Moyenne pondérée	100	100	100	2217

Les résultats obtenus sont satisfaisants donc nous dirons que le réglage des hyperparamètres augmente les performances d'où notre modèle de prédiction sera un pipeline composé des données prétraités et un algorithme de forêt aléatoire réglé sur 1000 Arbres de décision et présentant une profondeur maximale de 10. Ainsi, nous aurons besoin de 28 Variables pour faire la prédiction.

Les 28 variables retenues sont :

- error3count, error5count, error2count, comp2, rotate_max-24h, error1count.
- rotate_mean_24h, comp1, volt_min-24h, vibration_mean_24h,
- comp4, vibration_min_24h, volt_mean_24h, comp3, rotate_min_24h
- vibration_mean_3h, vibration_max_24h, volt_max_24h, rotate_max_3h
- vibration_min_3h, rotate_sd_24h, model, rotate_mean_3h, volt_sd_24h
- vibration_max_3h, pressure_min_24h, rotate_min_3h, pressure_max_24h

4.8 Prédiction et Déploiements

Notre modèle a pour but de prédire la défaillance de composant dans les prochaines 24h. Ainsi nous déterminons pour chaque classe la probabilité prédite par notre modèle de prédiction. Autrement dit si le modèle prédit une défaillance de composant 2 dans un avenir proche, alors la probabilité associée sera déterminée.

Dans les données de test, sur les trois derniers mois, nous avons sélectionnés 15 observations, où nous déterminerons les probabilités des classes. Ainsi les probabilités de non-défaillance ou de bon fonctionnement sont données ci-dessous.

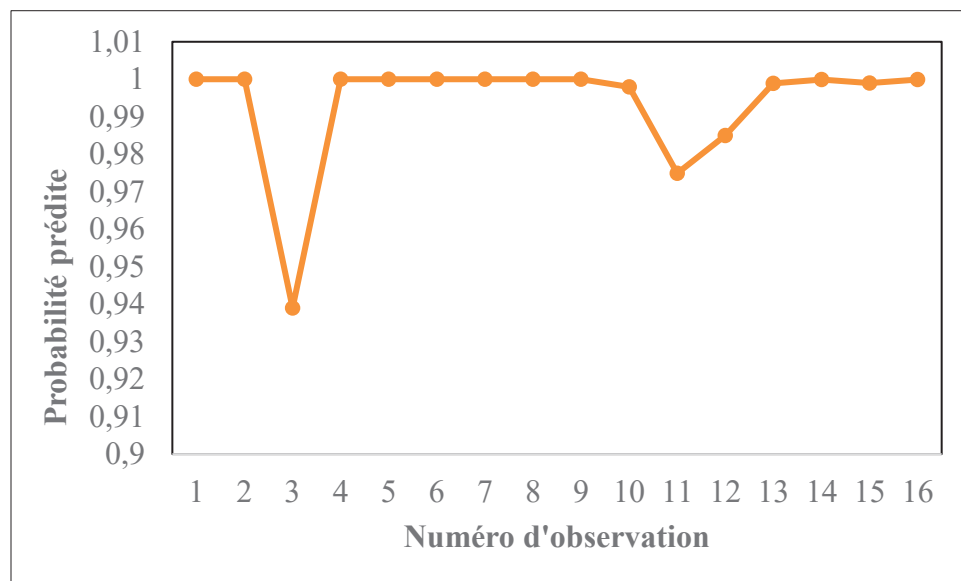


Figure 4.24 Probabilité défaillance de la classe none

La probabilité de défaillance de composant 2 est donnée ci-dessous

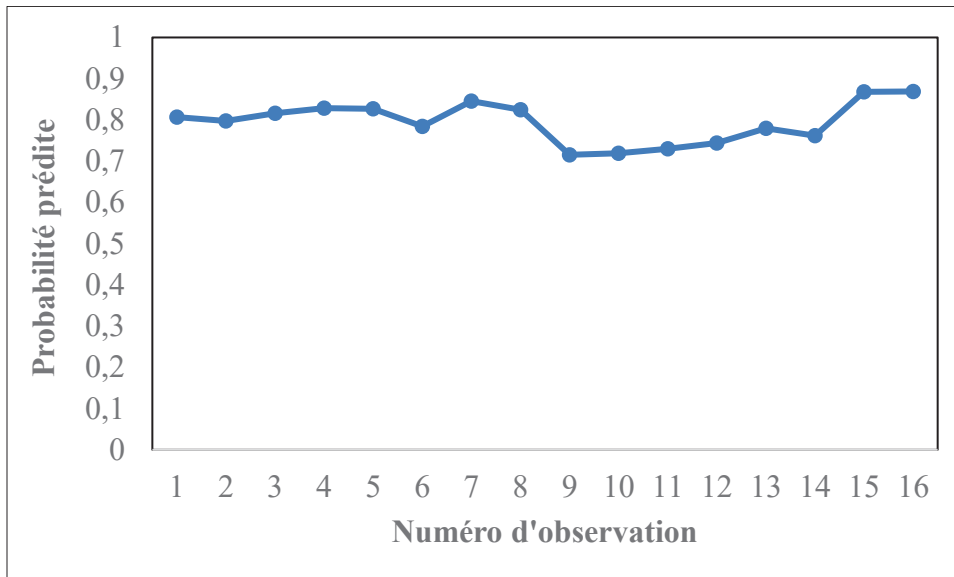


Figure 4.25 Probabilité de défaillance du composant 2

La probabilité de défaillance du composant 4 est donnée ci-dessous

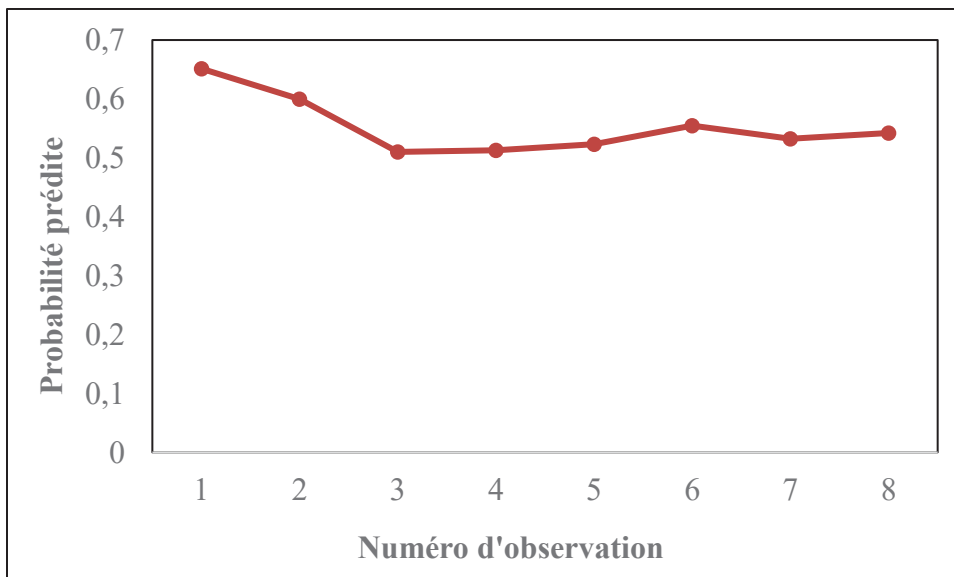


Figure 4.26 Probabilité de défaillance du composant 4

La distribution des probabilités permet de montrer que la distribution de probabilité de non-défaillance est égale à 1 ou proche de 1 signifie que notre modèle prédit correctement les états normaux. Le déploiement consiste à utiliser et à exploiter le modèle ainsi nous cherchons la probabilité de défaillance des composants 1, 2, 4 et none dans la deuxième quinzaine du mois de mois de novembre ainsi les périodes étudiés seront :

- 2015-11-16 à 06h00 correspondant à la période 0
- 2015-11-17 à 06h00 correspondant à la période 1
- 2015-11-18 à 06h00 correspondant à la période 2
- 2015-11-19 à 06h00 correspondant à la période 3
- 2015-11-20 à 06h00 correspondant à la période 4

Pour cela nous utilisons la machine 1 pour faire les prédictions et les résultats sont présentés ci-dessous

Tableau 4.21 Probabilité de défaillance sur la machine 1

Période	Proba compl	Proba comp2	Proba comp4	Proba none
0	2,139037e -07	0,001	2,008839e-07	0,999
1	2,139037e -07	0,000	2,008839e-07	1,000
2	2,139037e -07	0,000	2,008839e-07	1,000
3	2,139037e -07	0,001	2,008839e-07	0,999
4	2,139037e -07	0,000	2,008839e-07	1,000

4.9 Conclusion

Ce chapitre nous a permis de mettre en place une stratégie de maintenance intelligente de type prédictive. Le modèle prédictif ainsi mis en place permettra de déterminer l'état d'un équipement manufacturier dans un avenir proche c'est-à-dire dans les prochaines 24h. Pour cela, nous avons démontré toutes les étapes nécessaires pour la mise en place d'un tel modèle. Par ailleurs nous avons utilisé un pipeline composé de quatre algorithmes dont le forêt aléatoire a montré les meilleures performances. Le réglage des hyperparamètres a montré que les performances peuvent être augmentés d'une part et d'autre part, il peut les dégrader. Le déploiement de notre modèle nous permet de déterminer les probabilités qui seront utilisés dans la résolution de notre dernier sous-problème.

CHAPITRE 5

IMPACT DE L'ÉCONOMIE CIRCULAIRE EN MILIEU INDUSTRIEL

5.1 Introduction

Dans ce chapitre, nous exploitons un modèle d'optimisation qui nous permettra de voir l'impact de l'économie circulaire. Pour cela la fiabilité des composants sera un moyen pour voir l'évolution du profit dans un système manufacturier utilisant une maintenance intelligente associé à un élément d'économie circulaire. Ainsi nous prenons comme hypothèse que la fiabilité d'un équipement peut augmenter les gains. De ce fait nous expliquerons la notion de fiabilité ainsi que ses différents aspects, ensuite un modèle d'optimisation sera mis en place qui maximisera le profil et enfin une discussion sera établie afin de voir l'impact de la disposition de composants.

5.2 Notion de Fiabilité

La fiabilité se définit comme étant la capacité d'un système ou d'un équipement à remplir une fonction spécifique dans des situations particulières associé à un intervalle de temps donné (Bellaouar & Beleulmi, 2014). En d'autres termes la fiabilité fait allusion à l'aptitude d'un équipement à fonctionner de manière constante sans panne. Cependant la relation entre la fiabilité et la défaillance sera traitée ci-dessous.

5.2.1 Relation entre fiabilité et défaillance

Il faut noter que la fiabilité d'un équipement, dépend de plusieurs facteurs comme :

- la conception
- la fréquence de maintenance
- la fréquence d'utilisation
- l'âge de l'équipement
- l'environnement et les facteurs externes.

Ainsi plus un équipement est considéré comme fiable et plus le taux de défaillance diminue de ce fait la performance globale sera augmentée. Le taux de défaillance peut être mesuré en prenant en compte le nombre de panne, ou d'arrêt ou de dysfonctionnement sur une période bien définie. Nous supposons que l'évènement A désigne un état de bon fonctionnement à un instant précis (t) et que l'évènement B représente un état de défaillance ou de non-fonctionnement à l'instant ($t + \delta t$) et sachant que

$$p(A) = p(T > t) \text{ et } p(B) = p(T \leq t + \delta t) \quad (5.1)$$

Alors

$$p(A \cap B) = p(t < T < t + \delta t)$$

$$p(A \cap B) = F(t + \delta t) - F(t)$$

$$p(A \cap B) = (1 - R(t + \delta t)) - (1 - R(t))$$

$$p(A \cap B) = R(t) - R(t + \delta t)$$

d'où :

$$p(B/A) = \frac{R(t) - R(t + \delta t)}{R(t)} \quad (5.2)$$

La fonction de défaillance est notée $F(t)$ et peut s'exprimer par la forme suivante

$$F(t) = P(T \leq t)$$

L'évolution de la défaillance suit une tendance particulière appelée courbe en baignoire (openclassrooms, 2023).

En fonction du cycle de vie, le taux de défaillance peut varier. De ce fait nous observons trois phases que sont :

- Une phase dite de Période de jeunesse où le taux de défaillance décroît
- Une phase d'utilisation de l'équipement correspondant à la période de maturité où la période de durée de vie utile alors le taux de défaillance λ est constant dans ce période
- Une phase de vieillesse (usure) où λ aura tendance à augmenter.

Ainsi la courbe est décrite ci-dessous

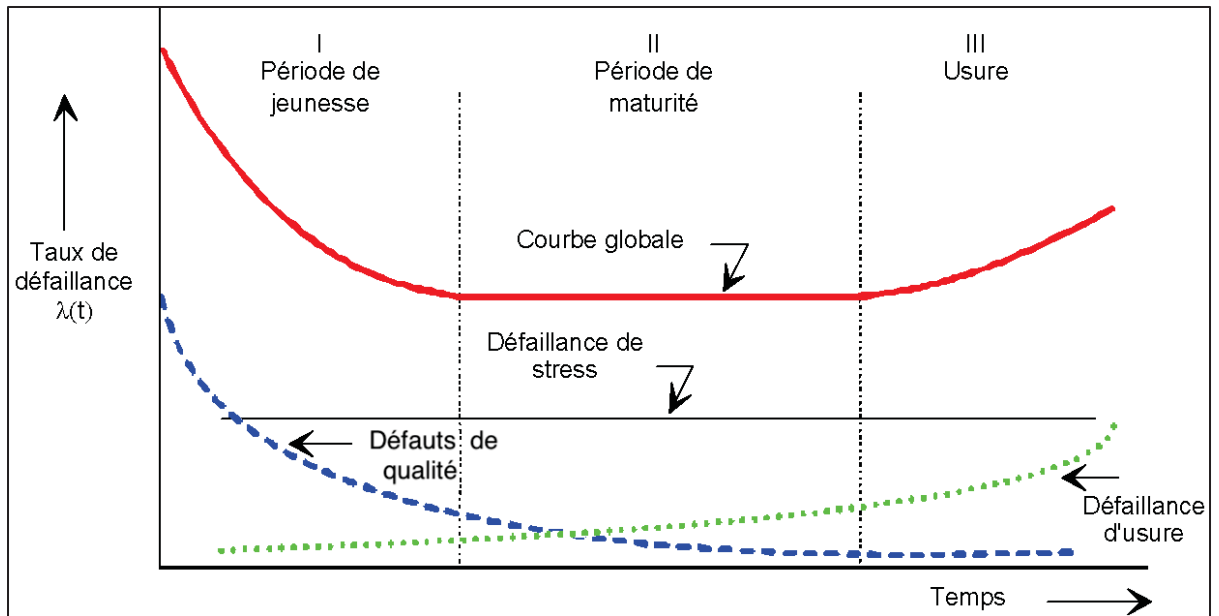


Figure 5.1 Évolution taux de défaillance en fonction du temps tirée de openclassrooms (2023)

La courbe en baignoire peut varier en fonction du type d'équipement (Bellaouar & Beleulmi, 2014). Pour cela les auteurs disent qu'en fonction de l'équipement nous aurons une courbe proche de la figure ci-dessous.

La courbe A représente une distribution de défaillance pour un équipement mécanique tandis que le B est celle d'équipement électromécanique et enfin le C signifie le taux de défaillance d'un équipement électronique.

En ce qui concerne un équipement électronique, sa période de jeunesse est marquée par une diminution rapide du taux de défaillance car les défauts observés sont dus à la conception de l'équipement qui n'est pas parfaite. Cette tendance disparaît progressivement parce que des tests sont effectués sur les équipements produits avant leur commercialisation. Sur la phase 2 le taux de défaillance est constant. De ce fait l'apparition de panne sera aléatoire et sera également liée à des facteurs externes. Au niveau de la phase 3 le taux de défaillance augmente en fonction de l'âge ou à l'usure des pièces qui composent l'équipement.

Pour un équipement mécanique sa période de vie utile est minimisée voire nulle ainsi sa phase 1 sera marquée par une décroissance du taux de défaillance en raison du rodage et dans sa

phase 2, le taux de panne ou de défaillance augmente à cause de la fatigue et de l'usure de l'équipement.

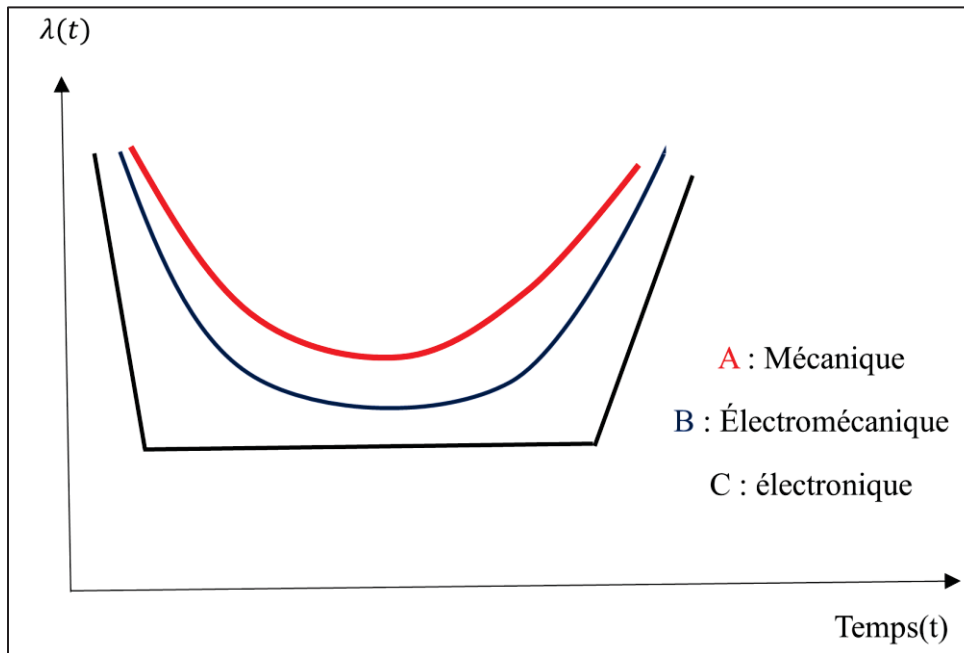


Figure 5.2 Évolution du taux de défaillance en fonction du type d'équipement adaptée de Bellaour (2014)

5.2.2 Fiabilité composant en série

Un équipement est dit en série lorsqu'il est composé de sous éléments placés les uns après les autres de sorte que la sortie d'un composant devient l'entrée du suivant. Dans une telle situation la fiabilité de l'équipement dépendra de la fiabilité individuelle de sous éléments cependant la fiabilité globale sera inférieure à la fiabilité d'un sous équipement pris individuellement. Étant donné que la fiabilité R est égale à $1 - P$, où P désigne la probabilité de défaillance, alors nous exprimons la fiabilité de (n) sous équipement non identique placés en série par la formule suivante :

$$R_{\text{série-non-identique}} = \prod_{i=1}^n R_i, \forall i \in [1, n] \quad (5.3)$$

Où R_i désigne la fiabilité d'un sous équipement individuel. Ainsi en utilisant les probabilités de défaillance la fiabilité sera donnée par la formule suivante

$$R_{\text{série-non-identique}} = \prod_{i=1}^n 1 - P_i, \forall i \in [1, n] \quad (5.4)$$

Où

P_i désigne la probabilité de défaillance d'un sous équipement.

Dans le cas de (n) sous équipement identiques placé en série, alors la fiabilité globale $R_{\text{série}}$ sera donnée par la formule suivante

$$R_{\text{série-identique}} = R_i^n, \forall i \in [1, n] \quad (5.5)$$

En utilisant la probabilité P_i alors la fiabilité globale pour (n) sous équipement identique placés en série sera exprimée par la formule suivante

$$R_{\text{série-identique}} = (1 - P_i)^n, \forall i \in [1, n] \quad (5.6)$$

La figure ci-dessous montre le principe de sous équipement en série

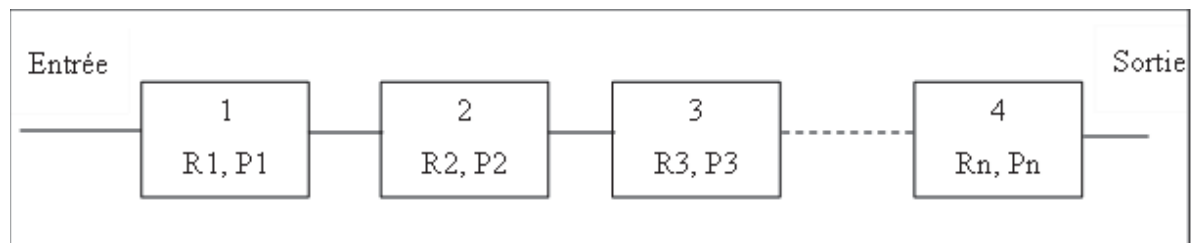


Figure 5.3 Disposition en série

5.2.3 Fiabilité composant en parallèle

Un équipement composé de sous -éléments est placé en parallèle signifie que les sous éléments sont placés les uns après les autres de sorte que la sortie de tous les composants du sous élément 1 devient l'entrée de tous les composants du sous élément 2, ainsi de suite pour (n) sous composants.

Autrement dit si (n) sous-éléments sont placés en parallèle alors, chaque sous élément aura la même entrée et la même sortie. Dans un tel situation La fiabilité de l'équipement dépendra de

la fiabilité individuelle de chaque sous éléments qui la compose. Cependant la fiabilité globale sera supérieure à la fiabilité d'un sous équipement pris individuellement.

Étant donné que la fiabilité R est égale à $1 - P$, où P désigne la probabilité de défaillance, alors nous exprimons la fiabilité de (n) sous-éléments identiques en parallèle selon la formule suivante :

$$R_{\text{parallèle-identique}} = 1 - (1 - R_i)^n, \forall i \in [1, n] \quad (5.7)$$

Où n représente le nombre de sous équipement.

En utilisant la probabilité de défaillance d'un sous équipement P_i alors la formule $R_{\text{parallèle-identique}}$ peut être exprimée de la façon suivante.

$$R_{\text{parallèle-identique}} = 1 - (1 - (1 - P_i))^n, \forall i \in [1, n] \quad (5.8)$$

Dans le cas de (n) sous équipements placés en parallèle et non identique alors

$$R_{\text{parallèle-non-identique}} = 1 - \prod_{i=1}^n (1 - R_i), \forall i \in n \quad (5.9)$$

En utilisant la probabilité de défaillance d'un sous équipement P_i alors la formule $R_{\text{parallèle-non-identique}}$ peut être exprimée de la façon suivante.

$$R_{\text{parallèle-non-identique}} = 1 - \prod_{i=1}^n (1 - (1 - P_i)), \forall i \in n \quad (5.10)$$

La figure ci-dessous montre le principe de sous équipement en parallèle

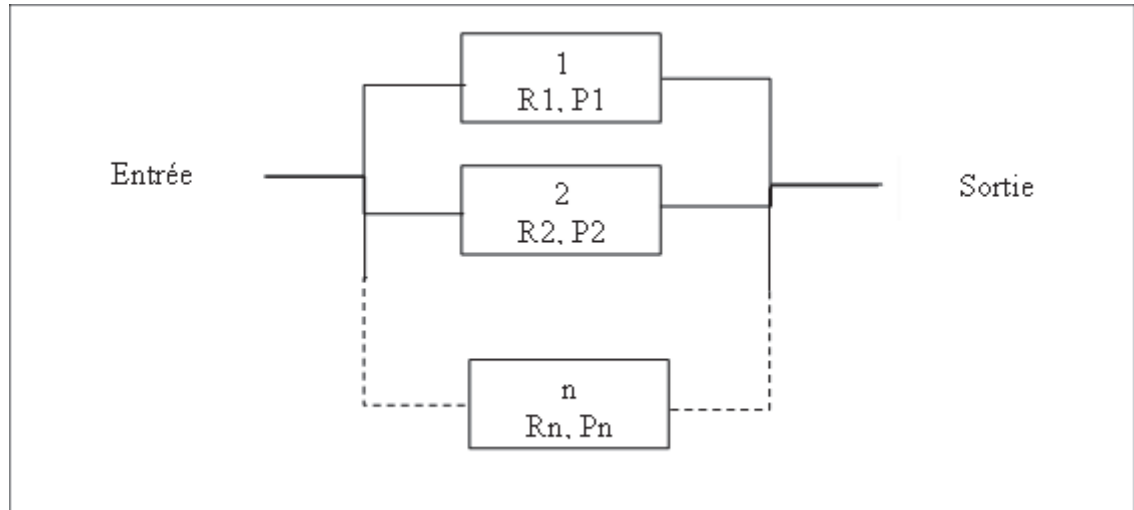


Figure 5.4 Disposition en parallèle

5.3 Fiabilité appliquée à la maintenance prédictive

La maintenance prédictive permet de prédire la défaillance de composant donc nous exploiterons les résultats obtenus par le tableau 4.21. Ainsi pour chaque défaillance de composant nous calculerons la fiabilité associée. De ce fait, nous utiliserons le composant 1, 2 et 4 et nous les disposerons dans différents schémas. Afin d'alléger le texte nous notons C1 signifie comp1, C2 signifie comp2 et C4 signifie comp4. Les dispositions en série et en parallèle seront formées à partir de ces composants cités ci-dessus.

5.3.1 Disposition du sous cas 1

Le sous cas 1, est composé de 2 composants 1, 2 composants 2 et 2 composants 4 tous placés en parallèle. La figure ci-dessous représente le sous cas 1

Le sous cas 1, peut être composé de 3 blocs en parallèle eux-mêmes que sont :

- Bloc α égal à C1 en parallèle avec C1
- Bloc β égal à C2 en parallèle avec C2
- Bloc γ égal à C4 en parallèle avec C4

La fiabilité du Bloc α est égal à :

$$R_{\alpha} = 1 - (1 - R_{c1})^2 = 1 - (1 - (1 - P_{c1}))^2 \quad (5.11)$$

La fiabilité du Bloc β est égale à :

$$R_{\beta} = 1 - (1 - R_{c2})^2 = 1 - (1 - (1 - P_{c2}))^2 \quad (5.12)$$

La fiabilité du Bloc γ est égale à :

$$R_{\gamma} = 1 - (1 - R_{c4})^2 = 1 - (1 - (1 - P_{c4}))^2 \quad (5.13)$$

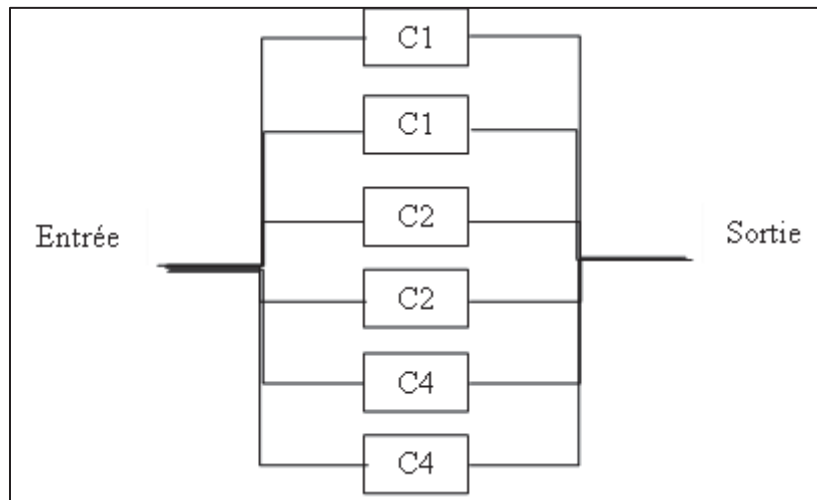


Figure 5.5 Dispositif sous cas 1

Les 3 blocs sont en parallèle donc la fiabilité globale sera donnée par α , β et γ en parallèle γ donc nous aurons :

$$R_{sous\ cas\ 1} = 1 - (1 - R_{\alpha}) * (1 - R_{\beta}) * (1 - R_{\gamma}) \quad (5.14)$$

Nous appliquons les résultats du tableau 4.21 afin de déterminer la fiabilité du sous cas 1

Tableau 5.1 Fiabilité pour le sous cas 1

Période	R_α	R_β	R_γ	$R_{\text{sous cas 1}}$
0	1	0,999999	1	1
1	1	1	1	1
2	1	1	1	1
3	1	0,999999	1	1
4	1	1	1	1

5.3.2 Disposition du sous cas 2

Le sous cas 2, est composé de 2 composants 1, 2 composants 2 et 2 composants 4 tous placés en parallèle. La figure ci-dessous représente le sous cas 2

Le sous cas 2, peut être composé de 3 blocs en parallèle eux-mêmes que sont :

- Bloc $\alpha 2$ égal à C1 en série avec C1
- Bloc $\beta 2$ égal à C2 en série avec C2
- Bloc $\gamma 2$ égal à C4 en série avec C4

La fiabilité du Bloc $\alpha 2$ est égal à :

$$R_{\alpha 2} = (R_{c1})^2 = (1 - P_{c1})^2 \quad (5.15)$$

La fiabilité du Bloc $\beta 2$ est égale à :

$$R_{\beta 2} = (R_{c2})^2 = (1 - P_{c2})^2 \quad (5.16)$$

La fiabilité du Bloc $\gamma 2$ est égale à :

$$R_{\gamma 2} = (R_{c4})^2 = (1 - P_{c4})^2 \quad (5.17)$$

Les 3 blocs sont en parallèle donc la fiabilité globale sera donnée par $\alpha 2$, $\beta 2$ et $\gamma 2$ en parallèle donc nous aurons :

$$R_{\text{sous cas 2}} = 1 - (1 - R_{\alpha 2}) * (1 - R_{\beta 2}) * (1 - R_{\gamma 2}) \quad (5.18)$$

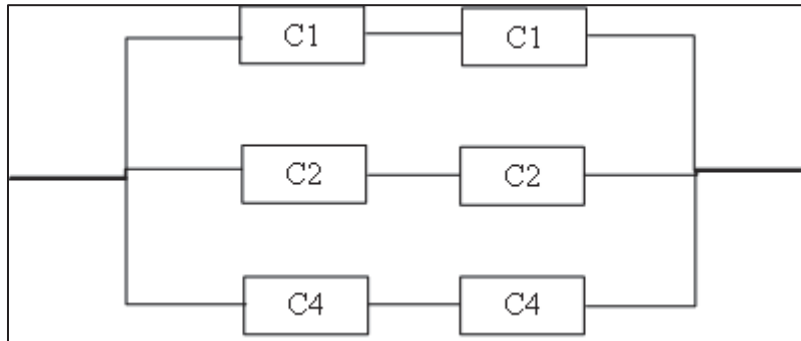


Figure 5.6 Dispositif sous cas 2

Nous appliquons les résultats du tableau 4.21 afin de déterminer la fiabilité du sous cas 2

Tableau 5.2 Fiabilité pour le sous cas 2

Période	$R_{\alpha 2}$	$R_{\beta 2}$	$R_{\gamma 2}$	$R_{\text{sous cas 2}}$
0	0,99999957	0,998001	0,9999996	1
1	0,99999957	1	0,9999996	1
2	0,99999957	1	0,9999996	1
3	0,99999957	0,998001	0,9999996	1
4	0,99999957	1	0,9999996	1

5.3.3 Disposition du sous cas 3

Dans cette partie, nous étudierons 3 versions qui sont une version A, B et C. Nous les présentons ci-dessous

- **Sous cas 3 version A**

Le sous cas 3 version A, est composé de trois paires de composants placées en parallèle. Toutes les paires sont assemblées en série. Les paires 1, 2 et 3 sont constitués respectivement de C1 en parallèle avec C2, de C2 en parallèle avec C1 et deux C4 en parallèle.

Ainsi ces blocs seront décrits ci-dessous :

- Bloc (*paire* – 1A) égal à C1 en parallèle avec C2
- Bloc (*paire* – 2A) égal à C2 en parallèle avec C1
- Bloc (*paire* – 3A) égal à C4 en parallèle avec C4

Dans la suite, de cette partie nous nommons les paires par 1A, 2A et 3A

La fiabilité du Bloc 1A est égal à :

$$R_{1A} = (1 - (1 - R_{C1}) * (1 - R_{C2})) \quad (5.19)$$

$$R_{1A} = (1 - (1 - (1 - P_{C1}) * (1 - (1 - P_{C2})))) \quad (5.20)$$

Étant donné que 1A et 2A sont constitués des mêmes composants donc la fiabilité du paire 2A est égale à celle de 1A

$$R_{1A} = R_{2A} \quad (5.21)$$

La fiabilité du paire 3 est égal à :

$$R_{3A} = 1 - (1 - R_{C4})^2 \quad (5.22)$$

Autrement

$$R_{3A} = 1 - (1 - (1 - P_{C4})^2) \quad (5.23)$$

La fiabilité globale est donnée par les 3 paires en série

$$R_{\text{sous cas 3-A}} = R_{1A} * R_{2A} * R_{3A} \quad (5.24)$$

D'où

$$R_{\text{sous cas 3-A}} = R_{1A}^2 * R_{3A} \quad (5.25)$$

La figure sous cas 3 A est donnée ci-dessous

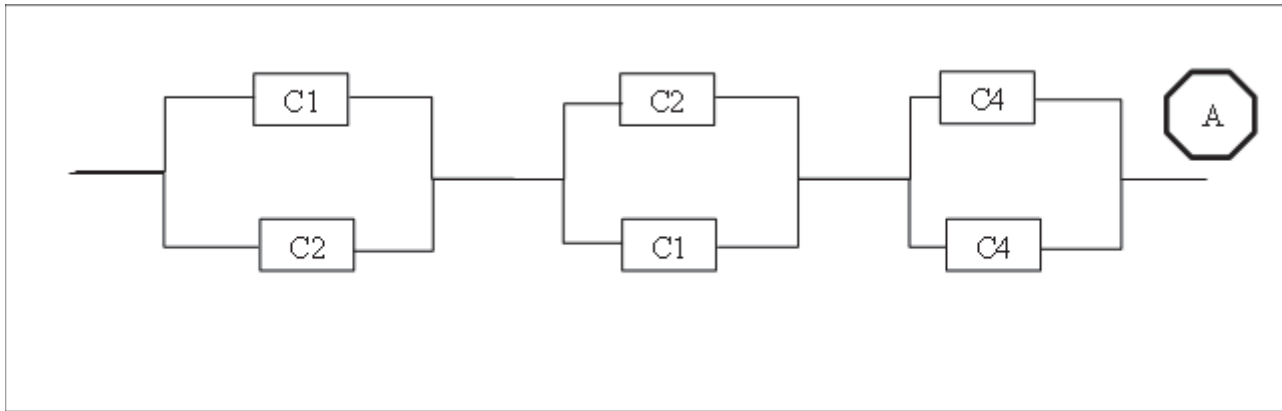


Figure 5.7 Dispositif Sous cas 3 - A

Nous appliquons les résultats du tableau 4.21 afin de déterminer la fiabilité du sous cas 3- A

Tableau 5.3 Fiabilité sous cas 3-A

Période	R_{1A}	R_{2A}	R_{3A}	$R_{\text{sous cas 3-A}}$
0	1	1	1	1
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

- **Sous cas 3 version B**

Le sous cas 3 version B, est composé de trois paires de composants placées en parallèle. Toutes les paires sont assemblées en série. Les paires 1, 2 et 3 sont constitués respectivement de C1 en parallèle avec C4, de C4 en parallèle avec C1 et deux C2 en parallèle.

Ainsi ces blocs seront décrits ci-dessous :

- Bloc (*paire – 1B*) égal à C1 en parallèle avec C4
- Bloc (*paire – 2B*) égal à C4 en parallèle avec C1
- Bloc (*paire – 3B*) égal à C2 en parallèle avec C2

Dans la suite, de cette partie nous nommons les paires par 1B, 2B et 3B

La fiabilité du Bloc 1B est égal à :

$$R_{1B} = (1 - (1 - R_{C1}) * (1 - R_{C4})) \quad (5.26)$$

$$R_{1B} = (1 - (1 - (1 - P_{C1}) * (1 - (1 - P_{C4})))) \quad (5.27)$$

Étant donné que 1B et 2B sont constitués des mêmes composants donc la fiabilité de la paire 2B est égale à celle de 1B

$$R_{1B} = R_{2B} \quad (5.28)$$

La fiabilité du paire 3 est égal à :

$$R_{3B} = 1 - (1 - R_{C2})^2 \quad (5.29)$$

Autrement

$$R_{3B} = 1 - (1 - (1 - P_{C2})^2) \quad (5.30)$$

La fiabilité globale est donnée par les 3 paires en série

$$R_{\text{sous cas 3-B}} = R_{1B} * R_{2B} * R_{3B} \quad (5.31)$$

D'où

$$R_{\text{sous cas 3-B}} = R_{1B}^2 * R_{3B} \quad (5.32)$$

La figure sous cas 3 B est donnée ci-dessous

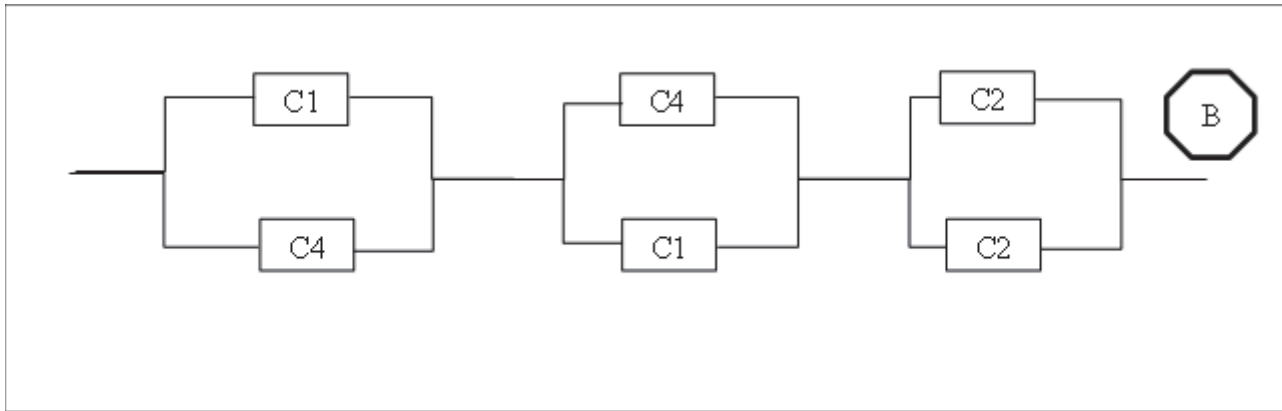


Figure 5.8 Dispositif Sous cas 3 B

Nous appliquons les résultats du tableau 4.21 afin de déterminer la fiabilité du sous cas 3- B

Tableau 5.4 Fiabilité sous cas 3-B

Période	R_{1B}	R_{2B}	R_{3B}	$R_{\text{sous cas 3-B}}$
0	1	1	0,999999	0,999999
1	1	1	1	1
2	1	1	1	1
3	1	1	0,999999	0,999999
4	1	1	1	1

- **Sous cas 3 version C**

Le sous cas 3 version C, est composé de trois paires de composants placées en parallèle. Toutes les paires sont assemblées en série. Les paires 1, 2 et 3 sont constitués respectivement de C2 en parallèle avec C4, de C4 en parallèle avec C2 et deux C1 en parallèle.

Ainsi ces blocs seront décrits ci-dessous :

- Bloc (*paire – 1C*) égal à C2 en parallèle avec C4
- Bloc (*paire – 2C*) égal à C4 en parallèle avec C2
- Bloc (*paire – 3C*) égal à C1 en parallèle avec C1

Dans la suite, de cette partie nous nommons les paires par 1C, 2C et 3C

La fiabilité du Bloc 1C est égal à :

$$R_{1C} = (1 - (1 - R_{C2}) * (1 - R_{C4})) \quad (5.33)$$

$$R_{1C} = (1 - (1 - (1 - P_{C2}) * (1 - (1 - P_{C4})))) \quad (5.34)$$

Étant donné que 1C et 2C sont constitués des mêmes composants donc la fiabilité de la paire 2C est égale à celle de 1C

$$R_{1C} = R_{2C} \quad (5.35)$$

La fiabilité de la paire 3 est égal à :

$$R_{3C} = 1 - (1 - R_{C1})^2 \quad (5.36)$$

Autrement

$$R_{3C} = 1 - (1 - (1 - P_{C1})^2) \quad (5.37)$$

La fiabilité globale est donnée par les 3 paires en série

$$R_{\text{sous cas 3-c}} = R_{1C} * R_{2C} * R_{3C} \quad (5.38)$$

D'où

$$R_{\text{sous cas 3-c}} = R_{1C}^2 * R_{3C} \quad (5.39)$$

La figure sous cas 3 C est donnée ci-dessous :

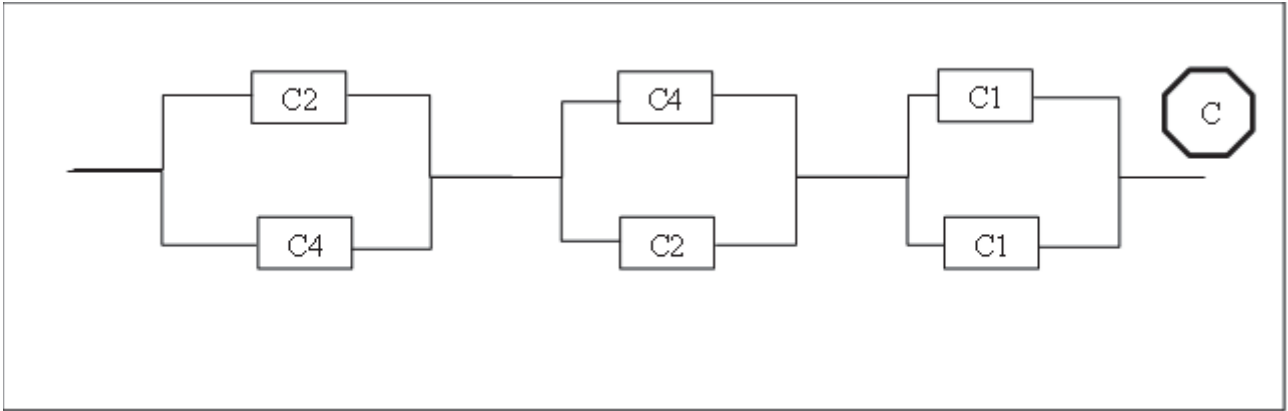


Figure 5.9 Dispositif sous cas 3-C

Nous appliquons les résultats du tableau 4.21 afin de déterminer la fiabilité du sous cas 3- C

Tableau 5.5 Fiabilité sous cas 3-C

Période	R_{1C}	R_{2C}	R_{3C}	$R_{\text{sous cas 3-C}}$
0	1	1	1	1
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

- **Sous cas 3 version A'**

Nous appliquons le même principe que pour tous les dispositifs ainsi la fiabilité pour chaque paire est exprimée ci-dessous :

$$R_{1A'} = (1 - (1 - (1 - P_{C1}) * (1 - (1 - P_{C2})))) \quad (5.40)$$

$$R_{2A'} = (1 - (1 - (1 - P_{C4}) * (1 - (1 - P_{C2})))) \quad (5.41)$$

$$R_{3A'} = (1 - (1 - (1 - P_{C1}) * (1 - (1 - P_{C4})))) \quad (5.42)$$

$$R_{\text{sous cas 3-A'}} = R_{1A'} * R_{2A'} * R_{3A'} \quad (5.43)$$

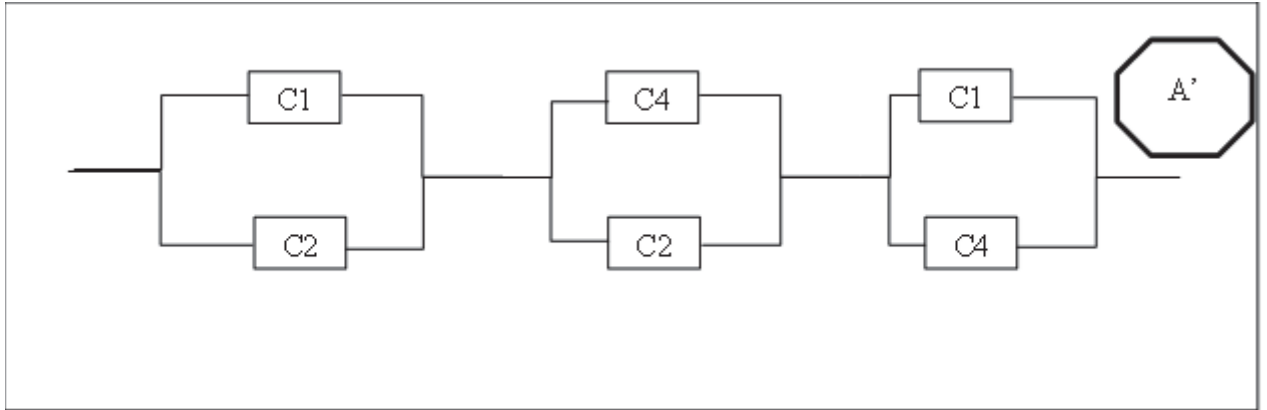


Figure 5.10 Disposition sous cas 3-A'

Tableau 5.6 Fiabilité sous cas 3-A'

Période	$R_{1A'}$	$R_{2A'}$	$R_{3A'}$	$R_{\text{sous cas 3-A'}}$
0	1	1	1	1
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

- **Sous cas 3 version B'**

Les fiabilités pour chaque paire est exprimée ci-dessous :

$$R_{1B'} = (1 - (1 - (1 - P_{C1}) * (1 - (1 - P_{C4})))) \quad (5.44)$$

$$R_{2B'} = (1 - (1 - (1 - P_{C4}) * (1 - (1 - P_{C2})))) \quad (5.45)$$

$$R_{3B'} = (1 - (1 - (1 - P_{C1}) * (1 - (1 - P_{C2})))) \quad (5.46)$$

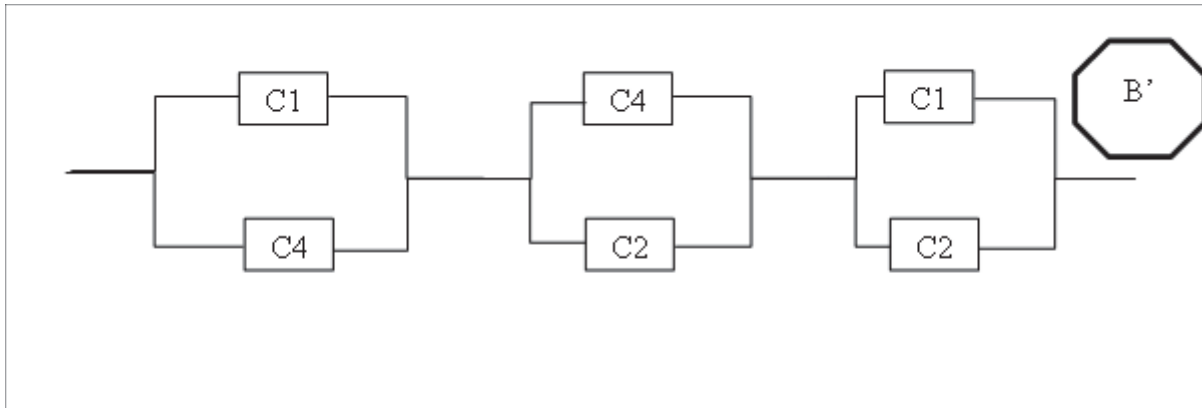


Figure 5.11 Dispositif sous cas 3- B'

Tableau 5.7 Fiabilité sous cas 3-B'

Période	$R_{1B'}$	$R_{2B'}$	$R_{3B'}$	$R_{\text{sous cas 3-B'}}$
0	1	1	1	1
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

La plupart des sous cas étudiés présentent une fiabilité de 1 car la période étudiée correspond à un fonctionnement normal donc les composants sont supposés ne pas tomber en panne. Dans la prochaine partie nous vérifierons l'impact de ces fiabilités dans un modèle d'optimisation

- **Sous cas 3 version C'**

Dans ce cas de figure, nous remarquons des équivalences entre les paires. Ainsi :

- $1C = 2A'$
- $2C' = 1B'$
- $3C' = 1A'$

Cette version est représentée ci-dessous

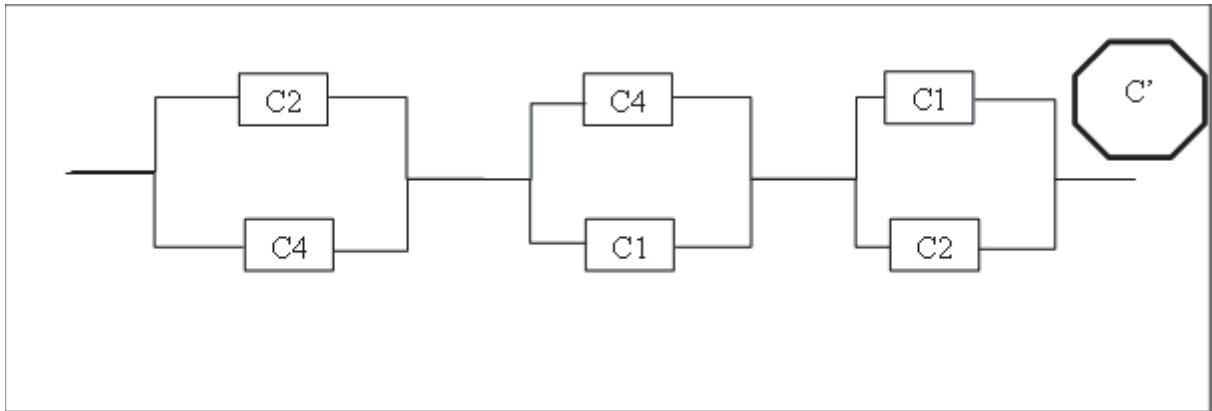


Figure 5.12 Disposition sous cas C'

Ainsi la fiabilité sera calculée ci-dessous.

Tableau 5.8 Fiabilité Sous cas 3-C'

Période	$R_{1C'}$	$R_{2C'}$	$R_{3C'}$	$R_{\text{sous cas 3-C'}}$
0	1	1	1	1
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

Nous remarquons que les fiabilités obtenues sont égales à 1 car la période étudiée correspond à un fonctionnement normal donc les composants ne sont pas en panne ainsi la fiabilité globale sera maximisée. Nous exploiterons ces fiabilités prédites dans un modèle d'optimisation qui sera traitée dans la partie suivante.

5.4 Modèle d'optimisation

Dans cette partie, notre modèle d'optimisation sera adapté aux travaux réalisés sur une planification de maintenance (Liao & Huang, 2014). Le volet économie circulaire a été intégré. Le modèle est appliqué dans un système manufacturier constitué de quatre types de machine

différentes nommées M1', M2', M3' et M4'. Le système manufacturier est représenté dans la figure ci-dessous.

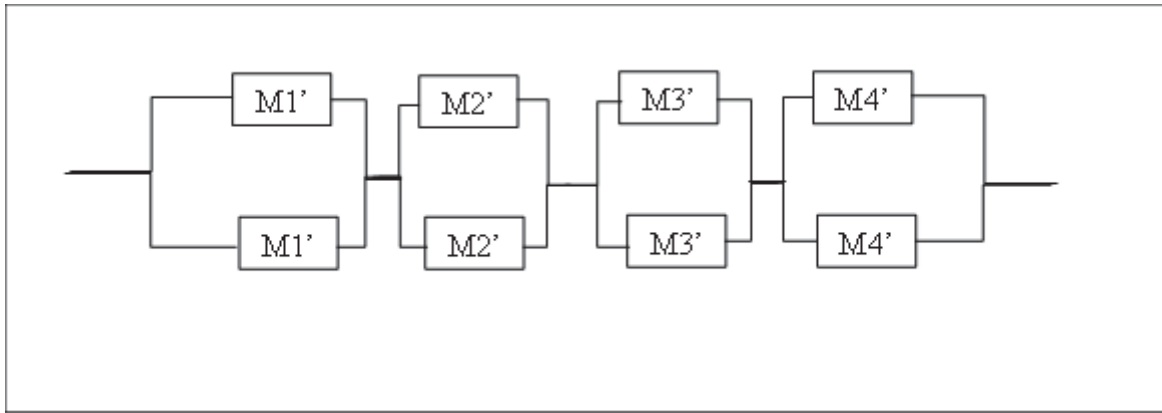


Figure 5.13 Système manufacturier

La fiabilité globale du système manufacturier est donnée par la formule suivante :

$$R_{systeme} = [1 - (1 - R_{M1'})^2] * [1 - (1 - R_{M2'})^2] * [1 - (1 - R_{M3'})^2] * [1 - (1 - R_{M4'})^2] \quad (5.47)$$

Nous supposons que les machines sont identiques alors la formule ci-dessus devient :

$$R_{systeme} = [1 - (1 - R_{M1'})^2]^4 \quad (5.48)$$

Dans ce système manufacturier, un produit commence par un traitement dans la machine M1' et parcourt successivement toutes les machines jusqu'à M4'. Une seule machine par paire est maintenue cela signifie par exemple que sur une même période nous ne pouvons pas faire la maintenance des 2 machines M1', car la production ne doit pas être interrompue par la maintenance. S'il n'y a pas de maintenance prévue sur une paire alors la production est répartie sur les deux machines concernées. Nous décrivons les objectifs de la production ci-dessous.

5.4.1 Objectifs de la production

Notre système manufacturier traite quatre types de produits que sont A, B, C et D. Chaque produit passe selon un temps spécifique dans une machine.

Nous l'appelons le temps de traitement du produit (i) sur la machine (j). Le nombre maximal pour chaque type de machine est de 2. Une planification opérationnelle a permis de mettre en place un plan de maintenance préventive initiale pour les machines. La production suit la demande du marché. Cette demande est satisfaite par la vente du produit(i) à la période(p). La production doit tenir compte du stockage car des frais sont associés aux stocks des produits lors de la dernière période c'est à dire chaque unité de produit stockées dans la dernière période coute 0,5 dollars. Cependant, à la dernière période, le stockage est limité à 50 pour chaque produit alors qu'en début de production le stock de tous les produits est nul. Pour des raisons économiques, le nombre maximal de produits stocké ne doit pas dépasser 100. Également des couts de maintenance sont appliqués par période mais aussi par type de machine.

Ainsi les couts de maintenance de la première période jusqu'à la dernière sont respectivement fixés à 110, 100, 70, 80 et 90 dollars.

Chaque unité de produit vendu rapporte un certain profil.

Le recyclage est effectué sur les machines lors de la maintenance ainsi le cout de recyclage varie d'une machine à une autre. L'ensemble de ces informations sont fournis par les tableaux ci-dessous.

Tableau 5.9 Temps de traitement du produit sur une machine

MACHINE\PRODUIT	A	B	C	D	Nombre machine
M1'	0,08	0,12	0,07	0,1	2
M2'	0,1	0,1	0,08	0,06	2
M3'	0,06	0,05	0,13	0,08	2
M4'	0,1	0,03	0,1	0,12	2
BENEFICE unitaire	5	6	7	8	

Les ventes sont données dans le tableau 5.10

Tableau 5.10 Ventes prévues

PERIODE\PRODUIT	A	B	C	D
0	600	800	300	200
1	600	500	500	100
2	500	300	0	0
3	300	300	100	0
4	100	200	500	300

Le planning de maintenance initial est donné ci-dessous

Tableau 5.11 Planning de maintenance préventive initial

PERIODE\MACHINE	M1'	M2'	M3'	M4'
0	1	0	1	1
1	0	1	0	0
2	1	0	1	1
3	0	1	0	0
4	1	0	1	1

L'objectif principal sera de maximiser les profils pour cela nous décrivons, les différentes étapes du modèle d'optimisation.

5.4.2 Ensembles et indices

Nous avons trois ensembles que sont :

- PRODUIT d'indice i concernent les quatre produits A, B, C et D donc $PRODUIT = \{A, B, C, D\}$, alors $i \in PRODUIT$
- PERIODE d'indice p concernent les cinq périodes 0, 1, 2, 3, 4 donc $PERIODE = \{0, 1, 2, 3, 4\}$, alors $p \in PERIODE$
- MACHINE d'indice j concernent les quatre machines M1', M2', M3' et M4' donc $MACHINE = \{M1', M2', M3', M4'\}$, alors $j \in MACHINE$.

5.4.3 Paramètres

Les paramètres sont les éléments connus et qui nous servent à bâtir le modèle d'optimisation, nous aurons donc :

- Le profil unitaire en dollar de chaque produit $i \in PRODUIT$ noté w_i
- Le temps de traitement d'un produit $i \in PRODUIT$ sur une machine $j \in MACHINE$ noté t_{ji}
- Le nombre maximal de machine $j \in MACHINE$ pour chaque paire notée n_j
- La vente prévue de chaque produit $i \in PRODUIT$ pour la période $p \in PERIODE$, elle est notée s_{pi}
- La quantité de produits $i \in PRODUIT$ vendue pour la période $p \in PERIODE$, elle est notée y_{pi}
- La maintenance d'une machine $j \in MACHINE$ au cours de la période $p \in PERIODE$ est notée z_{pj_init}
- Le cout de recyclage pour chaque machine $j \in MACHINE$ noté $c_{recyclage}$
- Le cout de maintenance pour chaque période $p \in PERIODE$ noté $c_{maint_periode}$
- La fiabilité du système pour chaque période $p \in PERIODE$ notée $Fiabilite_{periode_p}$

5.4.4 Variable de décision

Les variables de décision seront donc la production, le stock et le planning de maintenance optimal que sont :

- X_{pi} : Quantité de produits $i \in PRODUIT$ fabriquée à la période $p \in PERIODE$
- Z_{pj} : variable binaire représentant la maintenance de la machine $j \in MACHINE$ à la période $p \in PERIODE$ alors $Z_{pj} = \begin{cases} 1 & \text{si la maintenance est planifiée} \\ 0 & \text{sinon} \end{cases}$
- I_{pi} : Inventaire du produit $i \in PRODUIT$ à la période $p \in PERIODE$

5.4.5 Fonction-objectif

Le modèle d'optimisation permet de maximiser le bénéfice pour cela les gains seront soustraits aux couts afin d'avoir le profil optimal tout en prenant en compte la fiabilité et les contraintes d'économie circulaire.

Les bénéfices sont donnés par la formule suivante :

$$\text{Bénéfice} = \text{Fiabilité} * (\text{Gains} - \text{Couts})$$

Notons PERIODE = P, PRODUIT = I et MACHINE = J

Ainsi la formule est la suivante :

$$\begin{aligned} \text{Benefice}_{init} &= \sum_{p \in P} \sum_{i \in I} w_i * X_{pi} - \sum_{i \in I} 0,5 * I_{pi}(5, i) & (5.49) \\ &- \sum_{p \in P} \sum_{j \in J} c_{recyclage} * Z_{pj} \\ &- \sum_{p \in P} \sum_{j \in J} c_{maint_periode} * Z_{pj} \end{aligned}$$

$$\text{MaxZ} = \sum_{p \in P} \text{Fiabilite}_{periode_p} * \text{Benefice}_{init} \quad (5.50)$$

5.4.6 Contraintes

- Contraintes d'inventaires

Il permet de dire que le niveau d'inventaire ou le stock à la période $p \in P$ est égale au niveau d'inventaire ou de stock de la période précédente $p-1$ auquel nous rajoutons les quantités produites X_{pi} et nous soustrayons les quantités vendues y_{pi} . Elle s'exprime par la formule suivante :

$$I_{pi}(p - 1) + X_{pi} = y_{pi} + I_{pi}; \quad \forall p \in P; \quad \forall i \in I \quad (5.51)$$

- Contraintes initiales d'inventaires

Cette contrainte permet de spécifier que le niveau d'inventaire pour chaque produit à la période 0 est égal à 0 ainsi la formule est la suivante

$$X_{pi}(1, i) = y_{pi}(1, i) + I_{pi}(1, i); \quad \forall p \in P; \quad \forall i \in I \quad (5.52)$$

- Contraintes finales d'inventaires

Le niveau d'inventaire pour chaque produit à la dernière période c'est-à-dire à la période 4 doit être égal à 50. Elle sera représentée par la formule suivante

$$I_{pi}(5, i) = 50 ; \forall i \in I \quad (5.53)$$

- Contraintes sur la limite d'inventaire

L'inventaire d'un produit à une période ne doit pas dépasser la limite qui est fixée à 100.

$$I_{pi} \leq 100 ; \forall i \in I ; \forall p \in P \quad (5.54)$$

- Contraintes de production

Elle permet de spécifier que l'utilisation d'une machine ne doit pas dépasser le temps maximal de travail, étant donné que chaque période dure 24h et que l'utilisation des machines dépendent du nombre de machine disponible auquel nous prenons en compte la fiabilité donc la contrainte de production sera donnée par la formule suivante

$$\sum_{j \in J} t_{ji} * X_{pi} + 24 * Z_{pj} * Fiabilite_{periode_p} \leq 24 * n_j ; \forall i \in I ; \forall p \in P ; \forall j \in J \quad (5.55)$$

- Contraintes de vente

Cette contrainte prend en compte la vente prévue et se traduit par la quantité de produits vendus doit être inférieur ou égale à la quantité de vente prévu donc elle s'exprime par la formule suivante

$$y_{pi} \leq s_{pi} ; \forall i \in I ; \forall p \in P \quad (5.56)$$

- Contraintes de maintenance préventive

La maintenance optimale doit tenir compte du plan de maintenance initiale. Ainsi les contraintes de maintenance seront données par la formule (5.57)

$$\sum_{j \in J} Z_{pj} \leq m_p; \forall p \in P \quad (5.57)$$

Avec

$$m_p = \begin{cases} 3 & \text{si } p \in \{0, 2, 4\} \\ 1 & \text{si } p \in \{1, 3\} \end{cases}$$

$$\sum_{p \in P} Z_{pj} \leq m_j; \forall j \in J \quad (5.58)$$

Avec

$$m_j = \begin{cases} 3 & \text{si } j \in \{M1', M3', M4'\} \\ 2 & \text{si } j \in \{M2'\} \end{cases}$$

- Contraintes d'intégrité et de non-négativité des variables de décision

$$-X_{pi} \leq 0; \forall p \in P; \forall i \in I \quad (5.59)$$

$$-I_{pi} \leq 0; \forall p \in P; \forall i \in I \quad (5.60)$$

$$Z_{pj} \in \{0,1\}; \forall p \in P; \forall j \in J \quad (5.61)$$

5.5 Expérimentation

Le modèle d'optimisation est résolu sous LINGO 19.0 en utilisant un total de 239 contraintes dont les 80 sont non linéaires. Ainsi notre problème est traité comme un problème d'optimisation mathématique dont le but est de maximiser la fonction-objectif. Notre modèle est de type quadratique car nous avons des contraintes linéaires et non linéaires.

Pour cela plusieurs scénarios sont testés.

- Scénario 1

Dans ce scénario, la machine M1' est équivalent au dispositif du sous cas 1 alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous cas 1})^2]^4 \quad (5.62)$$

Ainsi la résolution a permis d'avoir un gain de 188830 \$

Le plan de production optimal est donné ci-dessous

Tableau 5.12 Plan de production optimale pour scénario 1

PERIODE\PRODUIT	A	B	C	D
0	700	900	400	300
1	501	401	400	1
2	500	300	50	3
3	299	300	50	1
4	150	249	550	345

Nous remarquons que le modèle fonctionne bien car les quantités fabriquées respectent les contraintes et qu'il n'a pas de surproduction inutile. L'évolution de l'inventaire est donnée ci-dessous

Tableau 5.13 Inventaire optimale pour scénario 1

PERIODE\PRODUIT	A	B	C	D
0	100	100	100	100
1	1	1	0	1
2	1	1	50	4
3	0	1	0	5
4	50	50	50	50

L'inventaire a respecté les contraintes donc nous proposons de stocker au maximum tous les produits à la période 0, ensuite en fonction des machines arrêtées lors d'une maintenance le plan de production pourra s'ajuster en conséquence. Le plan de maintenance optimale est donné ci-dessous

Tableau 5.14 Plan de maintenance optimale pour scénario 1

PERIODE\MACHINE	M1'	M2'	M3'	M4'
0	0	1	1	1
1	0	0	1	0
2	1	0	1	1
3	1	0	0	0
4	1	1	0	1

En ce qui concerne la maintenance, nous proposons le maximum de maintenance toutes les 2 périodes afin de bien répartir la disponibilité des machines et de favoriser une bonne gestion

des équipes de maintenance car si nous faisons 3 maintenances sur 3 périodes consécutives alors un déséquilibre lié à l'ordonnancement pourrait surgir.

- Scénario 2

Dans ce scénario, la machine M1' est équivalent au dispositif du sous cas 2 alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous\ cas\ 2})^2]^4 \quad (5.63)$$

Étant donné que la fiabilité du sous cas 1 est le même que celle du sous cas 2, alors nous aurons le bénéfice obtenu et les mêmes plans optimaux de production, d'inventaire et de maintenance

- Scénario 3

Dans ce scénario, la machine M1' est équivalent au dispositif du sous cas 3-A alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous\ cas\ 3-A})^2]^4 \quad (5.64)$$

Étant donné que la fiabilité du sous cas 3-A pour chaque période est de 1, alors la fiabilité du système pour chaque période sera égale à 1. Notre modèle a trouvé la même solution optimale du scénario 1.

- Scénario 4

Dans ce scénario, la machine M1' est équivalent au dispositif du sous cas 3-B alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous\ cas\ 3-B})^2]^4 \quad (5.65)$$

Étant donné que la fiabilité du sous cas 3-B pour les périodes 0 et 3 est de 0,999999 et de 1 pour les autres périodes, alors la fiabilité du système pour chaque période sera égale à 1. Notre modèle a trouvé la même solution optimale du scénario 1.

- Scénario 5

Dans ce scénario, la machine M1' est équivalent au dispositif du sous cas 3-C alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous\ cas\ 3-C})^2]^4 \quad (5.66)$$

Étant donné que la fiabilité du sous cas 3-C pour toutes les périodes est 1, alors la fiabilité du système pour chaque période sera égale à 1. Notre modèle a trouvé la même solution optimale du scénario 1.

- Scénario 6

Dans ce scénario, la machine M1' est équivalent au dispositif du sous cas 3-A' alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous\ cas\ 3-A'})^2]^4 \quad (5.67)$$

Étant donné que la fiabilité du sous cas 3-A' pour toutes les périodes est 1, alors la fiabilité du système pour chaque période sera égale à 1. Notre modèle a trouvé la même solution optimale du scénario 1.

- Scénario 7 et 8

Dans ces scénarii, la machine M1' est équivalent au dispositif du sous cas 3-B' et 3-C' alors la fiabilité du système du scénario 7 sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous\ cas\ 3-B'})^2]^4 \quad (5.68)$$

La fiabilité du système du scénario 8 sera donnée par la formule suivante :

$$R_{systeme} = Fiabilite_{periode_p} = [1 - (1 - R_{sous\ cas\ 3-C'})^2]^4 \quad (5.69)$$

Étant donné que la fiabilité du sous cas 3-B' et 3-C'; Pour toutes les périodes sont égales à 1, alors les fiabilités du système pour chaque période seront égales à 1. Ainsi pour tous ces scénarii, notre modèle a trouvé la même solution optimale que celle du scénario 1.

L'analyse de tous les scénarii peut être visualisé dans la figure ci-dessous

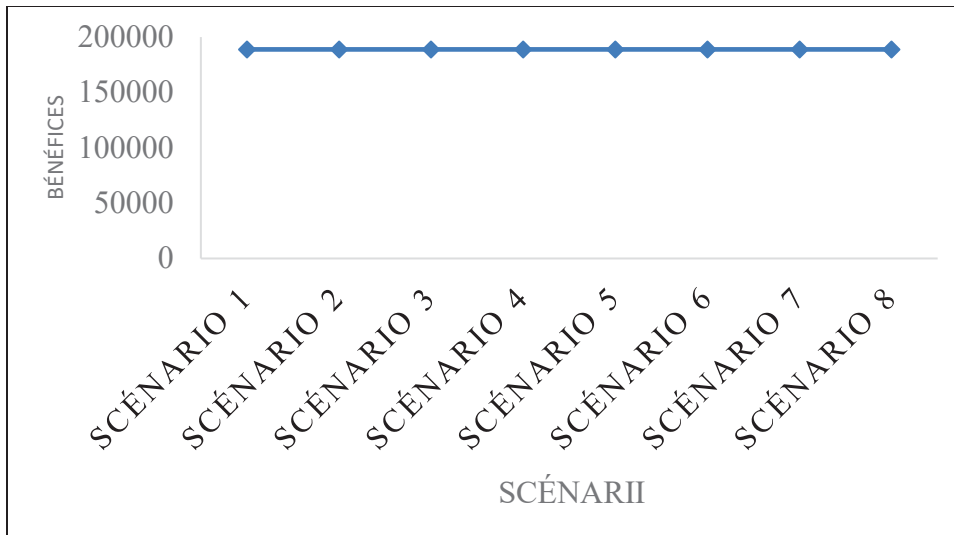


Figure 5.14 Évolution du bénéfice en fonction des scénarii

Nous remarquons qu'importe le scénario le bénéfice ne varie pas et vaut 1.888.300 \$. L'explication est dû au fait que la période étudiée concerne une non-défaillance de composants c'est à dire durant cette période les composants ne tombent pas en panne du coup leur probabilité de défaillance est faible alors leur fiabilité va augmenter et tendre vers 1. Ainsi nous remarquons dans ce cas de figure que la disposition des composants n'influence pas le profil. Dans la dernière partie de ce mémoire nous étudierons un nombre réduit de composant disposé uniquement en série ou parallèle.

5.6 Discussion

Nous prenons un nouveau dispositif constitué uniquement de 3 composants C1, C2 et C4. Chaque nouveau dispositif sera traité ci-dessous

- Dispositif en série

Dans ce dispositif C1, C2 et C4 sont placés en série donc la fiabilité de ce dispositif est donnée par la formule suivante :

$$R_{\text{dispositif-série}} = R_{c1} * R_{c2} * R_{c3} = (1 - P_{c1}) * (1 - P_{c2}) * (1 - P_{c3}) \quad (5.70)$$

Nous supposons que la machine M1' est équivalent au dispositif en série alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = \left[1 - (1 - R_{dispositif-serie})^2\right]^4 \quad (5.71)$$

Les fiabilités sont données par le tableau ci-dessous

Tableau 5.15 Fiabilité dispositif en série

Période	R_{C1}	R_{C2}	R_{C4}	$R_{Dispositif-serie}$	$R_{systeme}$
0	1	0,999	1	0,999	0,999996
1	1	1	1	1	1
2	1	1	1	1	1
3	1	0,999	1	0,999	0,999996
4	1	1	1	1	1

Ainsi la résolution a permis d'avoir un gain de 188829,7 \$

Le plan de maintenance a été modifié comme ci-dessous

Tableau 5.16 Plan de maintenance optimal dispositif en série

PERIODE\MACHINE	M1'	M2'	M3'	M4'
0	1	0	1	1
1	1	0	0	0
2	0	1	1	1
3	0	0	1	0
4	1	1	0	1

Le plan de maintenance a changé mais le nombre de maintenance maximal est toujours maintenu toutes les 2 périodes. Par ailleurs, le plan de production et d'inventaire est également modifié.

- Dispositif en parallèle

Dans ce dispositif, les composants C1, C2 et C4 sont placés en parallèle donc la fiabilité de ce dispositif est donnée par la formule suivante :

$$\begin{aligned} R_{dispositif-parallèle} &= 1 - (1 - R_{C1}) * (1 - R_{C2}) * (1 - R_{C4}) \quad (5.72) \\ &= 1 - (1 - (1 - P_{C1})) * (1 - (1 - P_{C2})) * (1 \\ &\quad - (1 - P_{C4})) \end{aligned}$$

Nous supposons que la machine M1' est équivalent au dispositif en parallèle alors la fiabilité du système sera exprimée de la manière suivante :

$$R_{systeme} = Fiabilite_{periode_p} = \left[1 - (1 - R_{dispositif-parallèle})^2\right]^4 \quad (5.73)$$

Les fiabilités sont données par le tableau ci-dessous

Tableau 5.17 Fiabilité dispositif en parallèle

Période	R_{C1}	R_{C2}	R_{C4}	$R_{Dispositif-parallèle}$	$R_{systeme}$
0	1	0,999	1	1	1
1	1	1	1	1	1
2	1	1	1	1	1
3	1	0,999	1	1	1
4	1	1	1	1	1

Le modèle d'optimisation a trouvé un bénéfice de 188830\$. Cependant le plan de maintenance optimal est le même que celui trouver dans le scénario 1

L'exploitation des résultats affirme que la fiabilité dépend fortement de la disposition des composants ainsi le dispositif en parallèle a donné un meilleur bénéfice par rapport à celui en série. Ainsi nous dirons que pour maximiser les gains dans une entreprise manufacturière il faut privilégier une disposition en parallèle.

5.7 conclusion

Dans ce chapitre nous avons mis en place un modèle d'optimisation de type quadratique qui maximise les gains en tenant compte de contraintes économique, sociales et environnementales. De ce fait l'économie circulaire par l'intermédiaire du recyclage au niveau de la maintenance a été traitée. Ainsi notre sous problème a été résolu en considérant plusieurs configurations.

Cependant la fiabilité obtenue grâce aux différentes configurations a montré que le profil dépend de l'état de l'équipement et de ses composants. Par ailleurs, une disposition en parallèle est plus performante qu'une disposition en série.

CONCLUSION

Dans ce mémoire nous avons traité une problématique liée à la stratégie de maintenance intelligente en milieu industriel, appliqué à l'économie circulaire. Ainsi tout au long de ce travail, nous avons montré comment les objectifs économiques et environnementales peuvent être satisfaites. Pour cela plusieurs pistes ont été explorés en utilisant les sous problèmes. Nous avons prouvé qu'une stratégie de maintenance intelligente permet de relever de nombreux défi comme l'amélioration de la disponibilité des ressources techniques, la réduction des couts de maintenances, l'ordonnancement efficiente des équipes de maintenance. Ainsi chaque chapitre a été mis en place pour traiter une partie de notre problématique.

Dans le premier chapitre, nous avons décrit l'utilité de notre recherche en nous basant sur le contexte manufacturier soumis à de nombreux défis. De ce fait nous avons présenté notre méthodologie suivant trois axes que sont l'utilisation de données collectées sur des machines, la mise en place d'une stratégie de maintenance intelligente de type prédictive et l'impact de la maintenance intelligente en utilisant l'économie circulaire.

Dans le deuxième chapitre les travaux antérieurs ont été présentés. Cette partie appelée revue de la littérature nous a permis de mieux comprendre l'évolution de la maintenance. De ce fait plusieurs notions indispensables pour la compréhension de notre sujet ont été abordés en l'occurrence les termes de défaillances, de dégradation et de durée de vie utile.

Dans le troisième chapitre l'analyse exploratoire des données a été traité. Pour cela les statistiques descriptives nous ont permis de visualiser les données, de faire la comparaison des données. Par ailleurs, des mesures statistiques ont été appliqués afin de trouver des différences dans les données.

Dans le quatrième chapitre, une stratégie de maintenance prédictive est développée en utilisant un pipeline regroupant des données prétraitées associé à un module de prédiction. Plusieurs classifieurs ont été utilisés dont le plus performant est le classifieur de foret aléatoire dont son

rôle ici est de prédire la défaillance de composant dans un avenir proche. Nous avons obtenu dans nos travaux plusieurs classes représentant les défaillances du composant 1, 2, 4 et l'état de fonctionnement normale. Ainsi notre modèle prédictif a permis de déterminer les probabilités de défaillance de chaque classe afin de trouver la fiabilité associée.

Dans le cinquième chapitre nous avons mis en place un modèle de programmation quadratique en nombres entiers mixtes qui permet de maximiser les profils en tenant en compte des contraintes de production, de stockage, de maintenance et d'économie circulaire via le recyclage. Ce modèle d'optimisation a été appliqué dans un contexte manufacturier disposant d'équipements connectés soit en série ou parallèle ou en série et parallèle. Les résultats obtenus nous ont permis de confirmer l'hypothèse selon laquelle le profil varie avec la disposition des équipements.

Dans les travaux futurs, les valeurs extrêmes du jeu de données de télémétrie pourraient être traité et remplacé par d'autres valeurs afin de vérifier les performances du modèle de prédiction. L'utilisation de ARIMA pourrait être pertinent dans la construction du modèle prédictif. Les métaheuristiques pourraient être intégré dans le modèle d'optimisation afin de déterminer un plan de maintenance plus détaillée.

ANNEXE I

SYNTHESE REVUE DE LA LITTÉRATURE

Travaux	Thèmes	Démarches	Méthodes
(Burggräf et al., 2022)	Approche holistique de maintenance intelligente	Analyse du cycle de vie.	Maximiser durée cycle de vie
(Schöggl et al., 2023)	Économie circulaire	Analyse descriptive hiérarchiques de type grappe et non paramétriques	Test de Kruskal-Wallis et la corrélation de rang de spearman
(Chen Xianda et al., 2021).	Économie circulaire	Réduire les émissions et la neutralité des carbones	Modèle d'optimisation multi-objectif
(Fumagalli et al., 2016)	Levier pour la mise en place d'une Maintenance intelligente	Détection des pannes	Surveillance des données
(Lever & King, 1993)	Levier pour la mise en place d'une Maintenance intelligente	Classification des évènements de la machine	Fusion de plusieurs sources de données
(Liu et al., 2018)	Levier pour la mise en place d'une Maintenance intelligente	Gestion des données	Oléoduc et la gazoduc clouding ou SCADA

Travaux	Thèmes	Démarches	Méthodes
(Sozen et al., 2022)	Levier pour la mise en place d'une Maintenance intelligente	Utilisation d'un PIG (Pipeline Inspection Gauge)	Probabilités de défaillances par simulation Monte-Carlo
(Mrazovac et al., 2021)	Maintenance intelligente	Apprentissage automatisée	Détection de sons anormaux
(Fordal et al., 2023)	Maintenance intelligente	Apprentissage automatisée	Détection de défaillance par réseau de neurone artificiel ANN
(Calzavara et al., 2021)	Maintenance intelligente	Apprentissage automatisée	Prédiction état de santé d'une machine par DBSCAN
		Par segmentation	
(Sharma & Tiwari, 2022)	Maintenance intelligente	Apprentissage automatisée	Surveillance état de santé avec des classifieurs SVM, KNN, RF, MLP
(Mohan et al., 2021)	Maintenance intelligente	Apprentissage automatisée	Prédiction des défauts en utilisant ARIMA
(Antao et al., 2018)	Planification de la maintenance intelligente	Module Maintenance intelligente	Maintenance prédictive.
		Module d'optimisation	Algorithme génétique
		Module de Simulation	Simulation données réels
(Gordon et al., 2020)	Planification de la maintenance intelligente	Module Maintenance intelligente	Maintenance prescriptive
		Module d'optimisation	Optimisation par GAMS

Travaux	Thèmes	Démarches	Méthodes
(Beldar et al., 2022)	Planification de la maintenance intelligente	Application sur plusieurs machines en parallèle et non identique	Algorithmes méta-heuristiques et MILP
(Sakthivelmurugan et al., 2022)	Économie circulaire en contexte manufacturier	Réutilisation des produits en fin de vie	Recyclage
(Lei et al., 2022)	Économie circulaire en contexte manufacturier	Comparaison de plusieurs approches d'économie circulaire	Techniques d'IA en économie circulaire
(Rajput & Singh, 2022)	Économie circulaire en contexte manufacturier	Économie circulaire en logistique inverse	Modèle d'optimisation MILP
(Wang et al., 2016)	Économie circulaire en contexte manufacturier	Économie circulaire et Fiabilité	Programmation non-linéaire

ANNEXE II

ANALYSE EXPLORATOIRE DES DONNÉES

Accès aux bloc notes

```
# Accès à google drive
from google.colab import drive
drive.mount('/content/drive')
```

Bibliothèques

```
#Importation des librairies
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
```

Jeux de données

```
# Chargement des données se trouvant dans google drive en utilisant la
librairie Pandas

telemetry =
pd.read_csv('/content/drive/MyDrive/Modelisation_memoire/PdM_telemetry.csv
')
errors =
pd.read_csv('/content/drive/MyDrive/Modelisation_memoire/PdM_errors.csv')

maint =
pd.read_csv('/content/drive/MyDrive/Modelisation_memoire/PdM_maint.csv')

failures =
pd.read_csv('/content/drive/MyDrive/Modelisation_memoire/PdM_failures.csv'
)
machines =
pd.read_csv('/content/drive/MyDrive/Modelisation_memoire/PdM_machines.csv'
)
```


ANNEXE III

MODÈLE PRÉDICTIVE

Librairies utilisées pour la modélisation

```
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.metrics import classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, recall_score,
accuracy_score, precision_score
from sklearn.model_selection import GridSearchCV
```

Algorithmes utilisés

```
# Création d'une liste d'algorithmes de classification
seed=42
classifiers = [LogisticRegression(random_state=seed),
                RandomForestClassifier(random_state=seed),
                AdaBoostClassifier(random_state=seed),
                KNeighborsClassifier(n_neighbors=5)
                ]
```

Matrices de confusion

```
def plot_confusion_matrix(model, test_y, predict_y):
    C = confusion_matrix(test_y, predict_y)
    A = ((C.T) / (C.sum(axis=1))).T
    B = (C / C.sum(axis=0))

    labels = ['comp2', 'comp4', 'none']
    # representing A in heatmap format
    print("\n", "-"*20, "\033[0;30;1mMatrice de confusion %s\033[0m"
%model, "-"*20, "\n")
    plt.figure(figsize=(20,7))
    sns.heatmap(C, annot=True, cmap="YlGnBu", fmt=".3f",
xticklabels=labels, yticklabels=labels)
    plt.xlabel('Predicted Class')
    plt.ylabel('Original Class')
    #plt.savefig(f'/content/drive/MyDrive/Modelisation_memoire/matrice de
confusion de {model}.png')
    plt.show()
```


ANNEXE IV

MODÈLE D'OPTIMISATION CODE LINGO

! Ensembles et Indices;

SETS:

PRODUIT: w_i;

PERIODE: m_p, c_maint_periode, Fiabilite_periode_p;

MACHINE: n_j, m_j, c_recyclage;

PRODUCTION(PERIODE, PRODUIT): X_pi, y_pi, I_pi, s_pi;

MAINTENANCE(PERIODE, MACHINE): Z_pj, Z_pj_init;

PRODUCTIVITE(MACHINE, PRODUIT): t_ji;

ENDSETS

! Déclaration des Ensembles;

DATA:

PRODUIT = @OLE();

PERIODE = @OLE();

MACHINE = @OLE();

! Déclaration des paramètres;

w_i = @OLE();

m_p = @OLE();

n_j = @OLE();

m_j = @OLE();

s_pi = @OLE();

y_pi = @OLE();

Z_pj_init = @OLE();

c_recyclage = @OLE();

c_maint_periode = @OLE();

Fiabilite_periode_p = @OLE();

! Déclaration des variables de décision;

@OLE() = X_pi;

@OLE() = Z_pj;

@OLE() = I_pi;

ENDDATA

! Fonction-objectif;

```
Benefice_init = (@SUM(PERIODE(p): @SUM(PRODUIT(i): w_i(i) * X_pi(p,i)))) -
  (@SUM(PRODUIT(i): 0.5 * (I_pi(5,i))))-
  (@SUM(MAINTENANCE(p,j): c_recyclage(j)*Z_pj(p,j))) -
  (@SUM(MAINTENANCE(p,j): c_maint_periode(p)*Z_pj(p,j)));
```

```
MAX = (@SUM(PERIODE(p): Fiabilite_periode_p(p)*Benefice_init));
```

! Contrainte de production;

```
@FOR(PRODUIT(i): @FOR(PERIODE(p): @FOR(MACHINE(j): @SUM(MACHINE(j):
(t_ji(j, i) * X_pi(p,i)) + (24 * Z_pj(p,j)*Fiabilite_periode_p(p))) <= (24*(n_j(j))))));
```

! contraintes inventaires;

```
@FOR(PRODUCTION(p,i):
X_pi(1, i) = y_pi(1,i)+I_pi(1,i);
@FOR(PRODUCTION(p,i)| p #NE# 1:
I_pi(p-1,i)+X_pi(p,i)= y_pi(p,i)+I_pi(p,i);
@FOR(PRODUIT(i): I_pi(5,i)=50);
@FOR(PRODUCTION(p,i): I_pi(p,i) <= 100);
```

! contraintes de maintenance préventive;

```
@FOR(MACHINE(j): @SUM(PERIODE(p): Z_pj(p,1)) = 3);
@FOR(MACHINE(j): @SUM(PERIODE(p): Z_pj(p,2)) = 2);
@FOR(MACHINE(j): @SUM(PERIODE(p): Z_pj(p,3)) = 3);
@FOR(MACHINE(j): @SUM(PERIODE(p): Z_pj(p,4)) = 3);
```

```
@FOR(PERIODE(p): @SUM(MACHINE(j): Z_pj(1,j)) <= 3);
@FOR(PERIODE(p): @SUM(MACHINE(j): Z_pj(2,j)) <= 1);
@FOR(PERIODE(p): @SUM(MACHINE(j): Z_pj(3,j)) <= 3);
@FOR(PERIODE(p): @SUM(MACHINE(j): Z_pj(4,j)) <= 1);
@FOR(PERIODE(p): @SUM(MACHINE(j): Z_pj(5,j)) <= 3);
```

! Contraintes de ventes;

```
@FOR(PRODUCTION(p,i): y_pi(p,i) <= s_pi(p,i));
```

! Contraintes d'intégrité et de non-négativité des variables de décision;

```
@FOR(PRODUCTION(p,i): -X_pi(p,i)<=0);
@FOR(PRODUCTION(p,i): -I_pi(p,i)<=0);
@FOR(PRODUCTION(p,i): @GIN(X_pi(p,i)));
@FOR(PRODUCTION(p,i): @GIN(I_pi(p,i)));
@FOR(MAINTENANCE(p,j): @BIN(Z_pj(p,j)));
```

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- ADEME Expertises. (2022, 24 mars). Économie de la fonctionnalité – Ademe. *Agence de la transition écologique*. Repéré à <https://expertises.ademe.fr/economie-circulaire/economie-fonctionnalite>
- Aït-El-Hadj, S. (2013). Éco-conception et méthodologie de conception, quelle convergence ? *Marché et organisations*, 17(1), 31-50. <https://doi.org/10.3917/maorg.017.0031>
- Antao, L., Reis, J., & Goncalves, G. (2018). Continuous Maintenance System for Optimal Scheduling Based on Real-Time Machine Monitoring. Dans *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)* (pp. 410-417). Turin : IEEE. <https://doi.org/10.1109/ETFA.2018.8502612>
- Beldar, P., Moghtader, M., Giret, A., & Ansariipoor, A. H. (2022). Non-identical parallel machines batch processing problem with release dates, due dates and variable maintenance activity to minimize total tardiness. *Computers & Industrial Engineering*, 168, 108135. <https://doi.org/10.1016/j.cie.2022.108135>
- Bellaouar, P. A., & Beleulmi, M. A. S. (2014). FIABILITE MAINTENABILITE DISPONIBILITE. *Faculté des Sciences de la Technologie Département Génie des Transports*, (Université Constantine 1), 106.
- Benda, M. (1993). AI adventures worth writing home about. Dans *Proceedings of International Joint Conference on Artificial Intelligence, 28 Aug.-3 Sept. 1993* (Vol. vol.1, pp. 747-50). San Mateo, CA, USA : Morgan Kaufmann Publishers.
- Bouabdellah, B. I., & Aissani, N. (2022). Smart Glasses and Augmented Reality for Maintenance 4.0.
- Burggräf, P., Wagner, J., Steinberg, F., Heinbach, B., Wigger, M., & Saßmannshausen, T. (2022). Life Cycle Assessment for Adaptive Remanufacturing: incorporating ecological considerations into the planning of maintenance activities – a case study in the German heavy machinery industry. *Procedia CIRP*, 105, 320-325. <https://doi.org/10.1016/j.procir.2022.02.053>

- Calzavara, G., Oliosi, E., & Ferrari, G. (2021). A Time-aware Data Clustering Approach to Predictive Maintenance of a Pharmaceutical Industrial Plant. Dans *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 13-16 April 2021 (pp. 454-8). Piscataway, NJ, USA : IEEE. <https://doi.org/10.1109/ICAIIIC51459.2021.9415206>
- Chauhan, G. (2021, 24 février). k-NN (k-Nearest Neighbors) Starter Guide. *Machine Learning HD*. Repéré à <https://machinelearninghd.com/k-nn-k-nearest-neighbors-starter-guide/>
- Chen Xianda, Gao Chao, Yang Yihao, Qian Jin, Pan Jiahui, & Chen Xingyu. (2021). Research on Electric Vehicle Planning under the Background of Carbon Neutrality. Dans *ICAIIIS 2021: 2021 2nd International Conference on Artificial Intelligence and Information Systems*, 28-30 May 2021 (p. 211 (4 pp.)). New York, NY, USA : ACM. <https://doi.org/10.1145/3469213.3470418>
- Data Science Team. (2021, 19 mars). Ingénierie des caractéristiques. *datascience*. [Site web éducatifs]. Repéré à <https://datascience.eu/fr/apprentissage-automatique/ingenierie-des-caracteristiques/>
- Delpla, V. (2021). Optimisation des chaines d’approvisionnement en boucle fermée avec l’internet des objets dans un contexte d’économie circulaire à l’ère de l’industrie 4.0, (Mémoire de Maitrise, (École de technologie supérieure, Montréal, QC).
- EC. (2019). Valorisation & recyclage. *economiecirculaire.org*. Repéré à <https://www.economiecirculaire.org/static/h/valorisation--recyclage.html>
- El kihel, Y., El kihel, A., & Bouyahrouzi, E. M. (2022). Contribution of Maintenance 4.0 in Sustainable Development with an Industrial Case Study. *Sustainability*, 11090 (26 pp.). <https://doi.org/10.3390/su141711090>
- Fidan Boylu Ouz. (2016). Microsoft Azure Predictive Maintenance. [Jeu de donnée]. Repéré à <https://www.kaggle.com/datasets/arnabbiswas1/microsoft-azure-predictive-maintenance>
- Fordal, J. M., Schjølberg, P., Helgetun, H., Skjermo, T. Ø., Wang, Y., & Wang, C. (2023). Application of sensor data based predictive maintenance and artificial neural networks to enable Industry 4.0. *Advances in Manufacturing*, 11(2), 248-263. <https://doi.org/10.1007/s40436-022-00433-x>

- Fumagalli, L., Macchi, M., Colace, C., Rondi, M., & Alfieri, A. (2016). A Smart Maintenance tool for a safe Electric Arc Furnace. *IFAC-PapersOnLine*, 49(31), 19-24. <https://doi.org/10.1016/j.ifacol.2016.12.155>
- Gordon, C. A. K., Burnak, B., Onel, M., & Pistikopoulos, E. N. (2020). Data-Driven Prescriptive Maintenance: Failure Prediction Using Ensemble Support Vector Classification for Optimal Process and Maintenance Scheduling. *Industrial & Engineering Chemistry Research*, 59(44), 19607-19622. <https://doi.org/10.1021/acs.iecr.0c03241>
- Jemli, M. (2023). Utilisation des dispositifs MEMS pour la maintenance intelligente des machines industrielles, (Mémoire de Maitrise, (École de technologie supérieure, Montréal, QC).
- Jemmali, M. (2022). Maintenance prédictive pour une turbine de puissance en utilisant des algorithmes d'apprentissage machine, (Mémoire de Maitrise, (École de technologie supérieure, Montréal, QC).
- Jolliffe, I. T., & Cadima, J. (2016, 13 avril). Analyse en composantes principales : bilan et développements récents. <https://doi.org/10.1098/rsta.2015.0202>
- Krajnc, A. (2023). Régression logistique, algorithme de classification | Jedha. [Site web éducatifs]. Repéré à <https://www.jedha.co/formation-ia/algorithme-regression-logistique>
- kwamimayeden. (2022a, 27 avril). L'arbre de décision (Decision Tree) en Machine Learning. *Kwami Mayeden*. Repéré à <https://kwamimayeden.com/larbre-de-decision-en-machine-learning/>
- kwamimayeden. (2022b, 26 juin). L'algorithme AdaBoost en Machine Learning. *Kwami Mayeden*. Repéré à <https://kwamimayeden.com/lalgorithme-adaboost-en-machine-learning/>
- Lei, Z., Cai, S., Cui, L., Wu, L., & Liu, Y. (2022). How do different Industry 4.0 technologies support certain Circular Economy practices? *Industrial Management & Data Systems*, 123(4), 1220-1251. <https://doi.org/10.1108/IMDS-05-2022-0270>
- Lelong, L. (2023). La collecte de données, méthodes et outils. *Jedha Bootcamp*. [Site web éducatifs]. Repéré à <https://www.jedha.co/formation-analyse-donnee/collecte-donnees>

- Lelong, L. (s.d.). Arima, modèle de prévision des séries temporelles | Jedha. Repéré à <https://www.jedha.co/formation-ia/algorithmes-arima>
- Leung, K. (2022, 13 septembre). Micro, Macro & Weighted Averages of F1 Score, Clearly Explained. *Medium*. Repéré à <https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f>
- Lever, P. J. A., & King, R. H. (1993). Intelligent sensor data analysis using AI based techniques. Dans *Symposium on Emerging Computer Techniques for the Minerals Industry, February 1, 1993 - February 1, 1993* (pp. 239-248). Tuscaloosa, AL, USA : Publ by Soc for Mining, Metallurgy & Exploration Inc.
- Liao, Y. H., & Huang, B. (2014). Optimization and Analysis of Production Plan of Mechanical Products. *Advanced Materials Research*, 926-930, 3862-3865. <https://doi.org/10.4028/www.scientific.net/AMR.926-930.3862>
- Liu, M., Shimin, J., & Mancang, Y. (2018). Design replica consistency maintenance policy for the oil and gas pipeline clouding SCADA multiple data centers storage system. Dans *2nd International Conference on Intelligent and Interactive Systems and Applications, IISA 2017, June 17, 2017 - June 18, 2017* (Vol. 686, pp. 715-721). Beijing, China : Springer Verlag. https://doi.org/10.1007/978-3-319-69096-4_101
- McDonald, M., Normandin, D., & Sauvé, S. (2016). *L'économie circulaire: Une transition incontournable*. (S.l.) : Presses de l'Université de Montréal. <https://doi.org/10.4000/books.pum.4151>
- Menet-Nédélec, F., Lampert, L., & Riou, P. (2011). Réseau Hydrologique Littoral Normand (RHLN) – Suivi 2008.
- Minitab. (2021). Qu'est-ce qu'un intervalle de confiance? [mtbconcept]. Repéré à <https://support.minitab.com/fr-fr/minitab/20/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/what-is-a-confidence-interval/>
- Mohan, R., Roselyn, P., Uthra, A., Devaraj, D., & Umachandran, K. (2021). Intelligent machine learning based total productive maintenance approach for achieving zero downtime in industrial machinery. *Computers & Industrial Engineering*, 157, 633-54. <https://doi.org/10.1016/j.cie.2021.107267>

- Mrazovac, B., Ilian, V., & Hulea, M. (2021). Beyond Industry 4.0: Leveraging AI-powered Anomalous Sound Detection for Smart Maintenance. Dans *2021 Zooming Innovation in Consumer Technologies Conference (ZINC), 26-27 May 2021* (pp. 43-7). Piscataway, NJ, USA : IEEE. <https://doi.org/10.1109/ZINC52049.2021.9499309>
- OpenClassrooms. (2023). Fusionnez des données avec Pandas. *OpenClassrooms*. [Site web éducatifs]. Repéré à <https://openclassrooms.com/fr/courses/7771531-decouvrez-les-librairies-python-pour-la-data-science/7857932-fusionnez-des-donnees-avec-pandas>
- openclassrooms. (2023, 27 avril). Calculez le taux de défaillance et la période de vie. *OpenClassrooms*. Repéré à <https://openclassrooms.com/fr/courses/5263401-initiez-vous-a-la-surete-de-fonctionnement/6548551-calculez-le-taux-de-defaillance-et-la-periode-de-vie>
- Parent, S.-É. (s.d.). *12 Les séries temporelles | Analyse et modélisation d'agroécosystèmes*. (S.l.) : (s.n.). Repéré à <https://essicolo.github.io/ecologie-mathematique-R/chapitre-temps.html>
- Rajput, S., & Singh, S. P. (2022). Industry 4.0 model for integrated circular economy-reverse logistics network. *International Journal of Logistics Research and Applications*, 25(4-5), 837-877. <https://doi.org/10.1080/13675567.2021.1926950>
- Rapid Miner. (2023). Confusion Matrix. *RapidMiner*. Repéré à <https://rapidminer.com/glossary/confusion-matrix/>
- RECYC-QUÉBEC. (2023). L'économie circulaire, une priorité [Site web d'entreprise]. *RECYC-QUÉBEC*. Repéré à <https://www.recyc-quebec.gouv.qc.ca/entreprises-organismes/mieux-gerer/economie-circulaire/>
- Ribeiro, D. R. S., Mendes, L. G., Forcellini, F. A., & Frazzon, E. M. (2022). Maintenance 4.0: A Literature Review and SWOT Analysis. Dans M. Freitag, A. Kinra, H. Kotzab, & N. Megow (Éds), *Dynamics in Logistics* (pp. 409-422). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-031-05359-7_33
- Ruel, A. (2021, 22 juin). Qu'est-ce que la réalité augmentée et comment fonctionne-t-elle ? *Visao*. Repéré à <https://visao.ca/fr/technologie-realite-augmentee/>
- Saint-Esprit, P.-E. (2023, 13 juin). Quels sont les enjeux de l'économie circulaire et collaborative ? *Manutan*. Repéré à <https://www.manutan.com/blog/fr/economie-circulaire/quels-sont-les-enjeux-de-leconomie-circulaire-et-collaborative>

- Sakthivelmurugan, E., Senthilkumar, G., & Karthick, K. N. (2022). Analysis of the impact of circular economy over linear economy in the paper processing industry. *Materials Today: Proceedings*, 66, 1446-1452. <https://doi.org/10.1016/j.matpr.2022.05.449>
- Schöggel, J.-P., Rusch, M., Stumpf, L., & Baumgartner, R. J. (2023). Implementation of digital technologies for a circular economy and sustainability management in the manufacturing sector. *Sustainable Production and Consumption*, 35, 401-420. <https://doi.org/10.1016/j.spc.2022.11.012>
- Sharma, S., & Tiwari, S. K. (2022). A novel feature extraction method based on weighted multi-scale fluctuation based dispersion entropy and its application to the condition monitoring of rotary machines. *Mechanical Systems and Signal Processing*, 171, 108909. <https://doi.org/10.1016/j.ymsp.2022.108909>
- Sozen, L., Yurdakul, M., & Ic, Y. T. (2022). Risk-based inspection planning for internal surface defected oil pipelines exposed to fatigue. *International Journal of Pressure Vessels and Piping*, 104804. <https://doi.org/10.1016/j.ijpvp.2022.104804>
- Thomas, L. (2020, 8 mai). Longitudinal Study | Definition, Approaches & Examples. *Scribbr*. Repéré à <https://www.scribbr.com/methodology/longitudinal-study/>
- Tibco. (s.d.). What is a Random Forest? *TIBCO Software*. Repéré à <https://www.tibco.com/reference-center/what-is-a-random-forest>
- Tremblay, C. (2021, 17 novembre). Balanced Accuracy Weighted, Accuracy and imbalanced data. *Kobia*. Repéré à <https://kobia.fr/classification-metrics-balanced-accuracy/>
- Voxco. (2023a). La régression logistique - Voxco. Repéré à <https://www.voxco.com/fr/blog/regression-logistique/>
- Voxco. (2023b). Qu'est-ce qu'une étude transversale? - Voxco. Repéré à <https://www.voxco.com/fr/blog/etude-transversale/>
- Wang, Q., Yuan, X., Zuo, J., Zhang, J., Hong, J., & Lin, C. (2016). Optimization of Ecological Industrial Chain design based on reliability theory – a case study. *Journal of Cleaner Production*, 124, 175-182. <https://doi.org/10.1016/j.jclepro.2016.02.096>

- Zhao, J., Tian, L., Cheng, L., Zhang, Y., & Zhu, C. (2022). Review on RUL Prediction Methods for Lithium-ion Battery. Dans *2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)* (pp. 1501-1505). <https://doi.org/10.1109/ICPSAsia55496.2022.9949753>
- Zwingelstein, G. (2019, 10 octobre). Maintenance prédictive intelligente pour l'industrie 4.0. *Techniques de l'Ingénieur*. Repéré à <https://www.techniques-ingenieur.fr/base-documentaire/genie-industriel-th6/mise-en-uvre-de-la-maintenance-42136210/la-maintenance-predictive-intelligente-pour-l-industrie-4-0-mt9572/maintenance-predictive-intelligente-pour-l-industrie-4-0-mt9572niv10003.html>